K. S. Nataraj, Estimation of vocal tract shape for speech training aids, M. Tech. Thesis, Department of Electrical Engineering, Indian Institute of Technology Bombay, June 2012.

Supervisor(s): Prof. P. C. Pandey

Abstract: Speech-training aids providing a visual feedback of articulatory efforts can be used for improving articulation by the hearing-impaired persons. LPC-based estimation of vocal tract shape works satisfactorily for vowels but fails during stop closure due to very low signal energy and lack of spectral information. The vocal tract shape during the stop closures of vowel-consonant-vowel (VCV) utterances can be estimated by bivariate surface modelling of the vocal tract area function during the vowel-consonant (VC) and consonant-vowel (CV) transition segments. The accuracy of this method depends on the accurate location of the transition segments and accuracy of the estimation of dynamically changing shapes. This thesis presents investigations for improving the estimation of vocal tract shape during the steady state and transition segments as well as during the closure segments. A windowed energy index is calculated as the ratio of the energy of the windowed signal to the frame energy, and it is shown that the shapes in the frames corresponding to the valleys in this index have a reduced variability. Thus the selection of the frames based on this index can be used for improving the consistency of vocal tract shape estimation during VC and CV transition segments of VCV utterances and these shapes can be used to more accurately estimate the place of articulation during stop closures of VCV utterances. Different LPC analysis techniques were investigated for estimation of the vocal tract shape of vowels. As compared to autocorrelation method, covariance and lattice methods showed reduced variability. However, covariance analysis resulted in unstable predictor polynomial for some utterances. A technique for detecting the VC and CV transitions in VCV utterances based on a measure of the rate of change of vocal tract area function is presented. The automatically marked start and end points of transitions showed a good match with the manually marked ones and resulted in a consistent estimation of the place of closure of velar, alveolar, and bilabial stops.