Nagesh S. Nayak, Estimation and display of vocal tract Shape for Speech Training Aids, M. Tech. Thesis, Department of Electrical Engineering, Indian Institute of Technology Bombay, June 2011.

Supervisor(s): Prof. V. Rajbabu and Prof. P. C. Pandey

Abstract: For providing visual feedback of articulatory efforts in speech training aids, the vocal tract shape can be estimated by LPC analysis of the speech signal. In this method, the vocal tract is modelled as a concatenation of equal length sections and the ratios of the areas on the two sides of section interfaces are calculated. The ratios are converted into areas by scaling them with respect to the area of a reference section. For a dynamic display of the vocal tract shape during speech utterances with transitional tract configuration, the reference area cannot be consistently obtained from the speech signal. The lip opening area as estimated from a video recording of the speaker's face can be used as a reference area for obtaining the vocal tract shape. The objective of this research is to validate the use of the area of the lip opening for obtaining the vocal tract shape and to devise a method for estimating the lip opening area from the recorded video frames. Use of the area of the lip opening for scaling the area ratios obtained from analysis of speech frames resulted in vocal tract area values which showed a good match with those obtained from direct imaging methods. Two techniques for estimating the area of the lip opening were investigated, based on template matching and active contours. The template matching technique uses a horizontal rectangle and a filled ellipse as templates in a mouth sub-image selected after face detection. It satisfactorily tracked the horizontal and vertical opening of the lips in the video images of speakers with different skin hues, recorded under good lighting conditions. A comparison of the estimated area values with those obtained from manual measurements from the video frames showed that the template matching performed better than the active contours. Active contours were found to be well suited for tracking the outer boundary of the lips. A combination of the two techniques needs to be further explored for obtaining the lip area as a reference for obtaining the vocal tract shape.