Jagbandhu, A visual feedback of vocal tract shape for speech training, M. Tech. Thesis, Department of Electrical Engineering, Indian Institute of Technology Bombay, June 2012.

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

Abstract: The absence of auditory feedback in hearing-impaired children severely affects the accuracy of articulation. For providing a visual feedback of articulatory efforts for speech training, the vocal tract shape can be estimated by LPC analysis of the speech signal. Since the signal energy is very low during stop closures segments of vowel-consonant-vowel (VCV) utterances, the estimated area values are unrelated to the place of closure. It has been reported earlier that vocal tract shape during stop closures can be estimated by performing bivariate polynomial interpolation of the estimated area during the transition segments preceding and following the stop closure. The project involves investigations for improving the visual feedback of vocal tract shape.

For validation of the vocal tract shape estimated from speech signal with reference to the X-Ray Microbeam database, an automated method to find the place of maximum constriction from the X-Ray Microbeam images is developed. It involves estimation of axial curve of the vocal tract from the pellet points given in the database and measuring the vocal tract opening as a distance between the palatal and the tongue outlines measured along the normal to the axial curve. For a realistic display of the vocal tract shape, it is shown that cubic B-spline interpolation of the estimated area values results in a smooth vocal tract shape and a distinct place of constriction. In the LPC-based estimation, area ratios on both sides of the section interfaces are converted into areas by assuming a constant normalized area of unity at the glottis end. This assumption fails during VC and CV transition segments as the vocal tract configuration is dynamically varying during these segments. Lip area value estimated from the video images can be used as a reference area for scaling the area ratios. It is shown that use of lip scaled area values for bivariate surface modeling results in consistent estimation of the place of closure for alveolar and bilabial stops. An animation for providing the visual feedback of the vocal tract shape for vowels and VCV utterances for the hearing-impaired is also developed.