ABSTRACT

In speech training aids for providing visual feedback of the articulatory efforts, time-varying vocal tract shape during speech production is generally obtained by linear prediction analysis of the speech signal and assuming a constant area at the glottis end as a reference. Errors in the estimated vocal tract shape, caused by variation in the area at the glottis end during speech production can be overcome by using area of the mouth opening as the reference. This area can be estimated by detecting the inner lip contour from the video recording of speaker's face during the utterance. A method for detection of the inner lip contour, based on color transformation and template matching, is presented for reducing the errors caused by presence of teeth and tongue. Face detection by Viola-Jones algorithm, localization using a mouth detection technique, and outer lip contour detection are used to narrow down the search region for the mouth opening. Presence of the teeth is masked by separate color transformations for upper and lower lip segments. For reducing the errors due to visibility of the tongue, which may not have any significant separation from the lips in the color space, a template matching method is employed. It is used separately for the upper and lower lip segments to obtain the mouth opening area. The method has been validated against graphically measured values of the mouth opening and found to be successful in estimating the mouth opening area, and it is not affected by skin hue and presence of teeth. Techniques based on active contours, optical flow and neural networks have also been investigated. The performances of the techniques have been evaluated for vowel sequences and vowel-consonant-vowel utterances.