K. S. Nataraj / Supervisor: Prof. P. C. Pandey, "Estimation of place of articulation of fricatives from spectral parameters using artificial neural network", *PhD Thesis*, Department of Electrical Engineering, IIT Bombay, September 2021.

Abstract

Speech-training aids providing visual feedback of the place of articulation, quantified as the distance of maximum constriction from the lips, are useful for improving the consonant articulation of children with hearing impairment. For such feedback, the place of articulation needs to be estimated from the speech signal. The research objective is to develop a speaker-independent method for estimating the place of articulation during fricatives for visual speech training.

The relation between the place of articulation of fricatives and their spectral parameters is investigated using the simultaneously acquired speech signals and articulograms available in the X-ray microbeam database. An automated graphical technique is developed for estimating the place of articulation from the articulograms, with the estimated place of articulation values closely matching those obtained by manual marking. Investigation relating the place of articulation with the spectral parameters is carried out using several earlier reported spectral parameters and a set of proposed spectral parameters. Earlier reported parameters, including spectral moments and spectral peak frequency, and the proposed parameters, including maximum-sum segment centroid, normalized sum of absolute spectral slopes, and four spectral energy parameters, were found to be associated with the place of articulation. An investigation is carried out for ANN-based speaker-independent mapping from spectral parameters of the frication segments to the place of articulation, using a feedforward network with multiple hidden layers and different number of neurons, different training data sizes, different sets of spectral parameters as the input, the place of articulation estimated from the articulograms as the reference, a dataset with 10,112 utterances, and five-fold cross-validation. Networks with two hidden layers were found to be adequate for all input parameter sets. The estimation using the proposed set of parameters resulted in the smallest mean RMS error of 2.55 mm, with scope for improving the estimation by increasing the training data size. The errors for alveolar and palatal fricatives were comparable to the standard deviation of the reference values. The errors for labiodentals were larger than the standard deviation of the reference values but smaller than their distance from the alveolars. The results indicated that the proposed ANN-based speaker-independent estimation could be used for feedback of the place of articulation.

A significant part of the estimation error could be attributed to non-uniqueness in the mapping. A perceptual study on the relative importance of transition segments adjacent to the fricative and the frication showed the place perception to be determined by a combination of the frication and transition segments. Investigation using the spectral parameters computed from the vocalic transition adjacent to frication showed that the ANN-based place estimation could be improved by supplementing the frication information with the vocalic information represented by the transition and vowel parameters.