Towards deriving compact and meaningful articulatory representations: an analysis of feature extraction techniques

Veena D. Singampalli and Philip J.B. Jackson

In this paper, we present an analysis of linear feature extraction techniques to derive a compact and meaningful representation of the articulatory data. We used 14-channel EMA (ElectroMagnetic Articulograph) data from two speakers from the MOCHA database [2]. As representations, we considered the registered articulator fleshpoint coordinates, transformed PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) features. Critical articulators were identified with registered coordinates, whereas critical modes were identified with PCA and LDA using our critical articulator identification algorithm [1]. The phone distributions in each feature space were modelled as univariate Gaussian distributions and identification of critical dimensions was controlled using a range of threshold values. Kullback Leibler (KL) divergence (evaluation divergence) was employed as a measure of goodness of fit of the models to actual phone models. We found that, of all representations, the LDA space yielded the smallest evaluation divergence between the model and phone pdfs, i.e., the best fit. The PCA representation that accommodated all the correlations amongst articulators gave the next best fit, closely followed by two other PCA representations which were designed to capture the strongest correlation patterns found amongst the articulators. The goodness of fit improved by 22% for male speaker and 46% for female speaker when LDA was used over the best PCA representation, and 77% for male and 66% for female over registered coordinate space at the point where average number of critical dimensions matched those obtained from IPA.

For PCA and LDA, the compactness of the representation was investigated by discarding the least significant modes. Recognition accuracy and evaluation divergence were used to analyse the performance of the resulting representations. No significant change in the recognition performance was found as the dimensionality was reduced from 14 to 8 (from t-test) whereas discarding further modes deteriorated the performance significantly. Similar performance trends were found when goodness of fit was investigated as the feature dimensionality was reduced. Experiments on LDA feature space improved the recognition accuracy by 2% on average over the best PC representation. We also present an articulatory interpretation of the PCA and LDA modes recovered by mapping the modes back to the registered articulatory coordinate space. Future work focuses on trajectory generation in different feature spaces guided by the findings of this study.

1. Veena D Singampalli and Philip JB Jackson. Statistical identification of critical, dependent and redundant articulators. Proc. Interspeech., Antwerpen, Belgium, pages 70-73, 2007.

2. A.A.Wrench. A new resource for production modelling in speech technology. Proc. Inst. of Acoust.,Stratford-upon-Avon, UK, 2001.