

Dehak, N., Dumouchel, P. and P. Kenny. "Modeling Prosodic Features with Joint Factor Analysis for Speaker Verification" *IEEE Transactions on Audio, Speech and Language Processing*, 15 (7) September 2007 : 2095-2103.

Abstract

In this paper, we introduce the use of continuous prosodic features for speaker recognition and we show how they can be modeled using joint factor analysis. Similar features have been successfully used in language identification. These prosodic features are pitch and energy contours spanning a syllable-like unit. They are extracted using a basis consisting of Legendre polynomials. Since the feature vectors are continuous (rather than discrete) they can be modeled using a standard Gaussian Mixture Model (GMM). Furthermore, speaker and session variability effects can be modeled in the same way as in conventional joint factor analysis. We find that the best results are obtained when we use the information about the pitch, energy, and the duration of the unit all together. Testing on the core condition of NIST 2006 speaker recognition evaluation data gives an equal error rate of 16.6% and 14.6%, with prosodic features alone, for all trials and English only trials, respectively. When the prosodic system is fused with a state-of-the-art cepstral joint factor analysis system, we obtain a relative improvement of 8% (all trials) and 12% (English only) compared to the cepstral system alone.

Keywords : Speaker recognition, prosodic features, Legendre polynomial, joint factor analysis.