

Speaker accent recognition through statistical descriptors of Mel-bands spectral energy and neural network model

Abstract

Accent recognition is one of the most important topics in automatic speaker and speaker-independent speech recognition (SI-ASR) systems in recent years. The growth of voice-controlled technologies has become part of our daily life, nevertheless variability in speech makes these spoken language technologies relatively difficult. One of the profound variability is accent. By classifying accent types, different models could be developed to handle SI-ASR. In this paper, we classified three accents in English language recorded from three main ethnicities in Malaysia namely Malay, Chinese and Indian using artificial neural network model. All experiments were performed in speaker-independent and three most accent-sensitive words-independent modes. Mel-bands spectral energy was extracted from eighteen bands taking the statistical values of each speech sample i.e. mean, standard deviation, kurtosis and the ratio of standard deviation to kurtosis to characterize the spectral energy distribution. The system was evaluated using independent test dataset, partial-independent test dataset and training dataset. The best three-class accuracy rate of 99.01% with independent test dataset was obtained. The overall accuracy rate for several trials was averaged to 96.79% with the average learning time at 49 epochs.

Keywords — Accent recognition, Mel-bands, neural network, spectral energy, statistical analysis