Abstract

Creating an accurate Speech Emotion Recognition (SER) system depends on extracting features relevant to that of emotions from speech. In this paper, the features that are extracted from the speech samples include Mel Frequency Cepstral Coefficients (MFCC), energy, pitch, spectral flux, spectral roll-off and spectral stationarity. In order to avoid the curse of dimensionality, statistical parameters, i.e., mean, variance, median, maximum, minimum, and index of dispersion have been applied on the extracted features. For classifying the emotion in an unknown test sample, Support Vector Machines (SVM) has been chosen due to its proven efficiency. Through experimentation on the chosen features, an average classification accuracy of 86.6% has been achieved using one-v/s-all multi-class SVM which is further improved to 100% when reduced to binary form problem. Classifier metrics viz. precision, recall, and F-score values show that the proposed system gives improved accuracy for Emo-DB.
- Finkelstein, S. et al. Investigating the influence of virtual peers as dialect models on
  students' prosodic inventory. INTERSPEECH. (Sep. 2012), 60-67.
- Buzo, A., Gray, A., Gray, R and Markel, J. Speech coding based upon vector
  562-574.
- Burges, C. A tutorial on Support Vector Machines for pattern recognition. Data Mining
- Hsu, C., Chang, C. and Lin, C. A practical guide to Support Vector Machines. 2003,
  1-16.

**Index Terms**

Computer Science

Signal Processing

**Keywords**

Feature extraction  dimensionality reduction  feature classification  Support Vector

Machines

Emotion recognition