

Hierarchical method to classify emotions in speech signals

**B.N.W.M.R.A. Boragolla^{*}, F.F. Farook, H.M.T.S.K. Herath
M.P.B. Ekanayake, G.M.R.I. Godaliyadda and J.V. Wijayakulasooriya**

*Department of Electrical and Electronic Engineering, Faculty of Engineering,
University of Peradeniya, Sri Lanka
^{*}rboragolla@gmail.com*

Automated speech emotion recognition is a crucial element in the computational study of human communication behaviors. This process supports human decision making and helps to design human-machine interfaces that facilitate efficient communication. In the designing phase, feature extraction, selection and classification methodology are major concerns. This study proposes a hierarchical approach for automated speech emotion recognition.

In this study, a publicly available German database (EMODB), which contains speech samples uttered in 7 different emotions, namely Anger, Boredom, Disgust, Fear, Happiness, Neutrality and Sadness was used. The hierarchical classification methodology proposed is based on Fisher Linear Discriminant Analysis (FLDA) and K Nearest Neighbour search (KNN) to classify the emotional states of speech. The hierarchy is mainly composed of three stages; in the first stage, angry speech signals are separated from others. Then, in the second stage sad speech signals are extracted from the non-angry set obtained in the first stage. In the third stage, the rest of the speech signals are divided into two different clusters, one with neutral and boredom speech signals and the other with happy, disgust and fear speech signals in it.

In this scenario the same emotional impact is delivered through various speech patterns and through different people; hence, the degree of freedom is high. Thus, a simple standard classification method cannot be used. This makes the sparseness of the clusters complex. Further, the emotion classification is complicated when compared to the speaker or speech recognition, because some emotion specific hidden features are to be discovered independent of the speaker and the speech.

In many of the previous studies, a direct approach is used to classify emotions with large feature sets. In contrast, a hierarchical method is followed in this study to improve the recognition rates with a minimum number of features to make the system more efficient by saving computational cost. This method improves the recognition rates over the previous studies [1] by 9%, 3% and 8% on sad, neutral and bored emotions. As particular features help to distinguish some specific emotional content of the speech signal explicitly from other emotions, the hierarchical method becomes more effective.