

Artificial Neural Network based Emotions Recognition System for Tamil Speech

Dayani Paranthaman¹, S. Thirukumaran (thiruvks@gmail.com)¹

¹Department of Physical Science, Vavuniya Campus of the University of Jaffna, Sri Lanka.

Abstract

Emotion has become the important part in communication between human and machine, so the detection of emotions has become important part in pattern recognition through Artificial Neural Network (ANN). Human's emotions can be detected based on the physiological measurements, facial expressions and speech. Since human shows different expressions for a particular emotion when they are speaking therefore the emotions can be quantified. The English speech dataset is provided with descriptions of each emotional context available in Emotional Prosody Speech and Transcripts in the Linguistic Data Consortium (LDC).

The main objective of this project describes the ANN based approach for Tamil speech emotions recognition by analyzing four basic emotions sad, angry, happy and neutral using the mid-term features. Tamil speeches are recorded with four emotions by males and females using the software "Cubase". The time duration is set to three seconds with the sampling frequency of 44.1 kHz as it is the logical and default choice for most digital audio material.

For the simulations, these recorded speech samples are categorized into different datasets and 40 samples are included in each dataset. Preprocessing includes sampling, normalization and segmentation and is performed on the speech signals. In the sampling process the analog signals are converted into digital signals then each speech sentence is normalized to ensure that all the sentences are in the same volume range. Next, the signals are separated into frames in the segmentation process. Then, the mid-term features such as speech rate, energy, pitch and Mel Frequency Cepstral Coefficients (MFCC) are extracted from the speech signals. Mean and Variance values are calculated from the extracted features. To create the classifier for the emotions, the above statistical results as an input matrix with their related emotions-target matrix are fed to train, validate and test.

The neural network back propagation algorithm is executed by the classifier to recognize completely new samples of Tamil speech datasets. Each of the datasets consists of different combinations of speech sentences with different emotions. Then, the new speech samples are assigned to identify the recognition rate of the speech emotions using the confusion matrix.

In conclusion, the selected mid-term features of Tamil speech signals classify the four emotions with the overall accuracy of 83.45%. Thus, the mid-term features selected are proven to be the good representations of emotions for Tamil speech signals and correctly recognize the Tamil speech emotions using ANN. The input gathered by a group of experienced drama artists who have the voice with the good emotional feelings would help to increase the accuracy of the dataset.

Keywords: *Artificial Neural Network, Confusion matrix, Mel Frequency Cepstral Coefficients*