

## **Analysis of Emotional Speech Recognition Using Artificial Neural Network**

**Archana, A.F.C. and Thirukumaran, S.**

Department of Physical Science, Vavuniya Campus of the University of Jaffna

Email: thiruvks@gmail.com

### **Abstract**

This paper presents an artificial neural network based approach for analyzing the classification of emotional human speech. Speech rate and energy are the most basic features of speech signal but they still have significant differences between emotions such as angry and sad. The feature pitch is frequently used in this work and auto-correlation method is used to detect the pitch in each of the frames. The speech samples used for the simulations are taken from the dataset Emotional Prosody Speech and Transcripts in the Linguistic Data Consortium (LDC). The LDC database has a set of acted emotional speeches voiced by the males and females. The speech samples of only four emotions categories in the LDC database containing both male and female emotional speeches are used for the simulation. In the speech pre-processing phase, the samples of four basic types of emotional speeches sad, angry, happy, and neutral are used. Important features related to different emotion states are extracted to recognize speech emotions from the voice signal then those features are fed into the input end of a classifier and obtain different emotions at the output end. Analog speech signal samples are converted to digital signal to perform the pre-processing. Normalized speech signals are segmented in frames so that the speech signal can maintain its characteristics in short duration. 23 short term audio signal features of 40 samples are selected and extracted from the speech signals to analyze the human emotions. Statistical values such as mean and variance have been derived from the features. These derived data along with their related emotion target are fed to train using artificial neural network and test to make up the classifier. Neural network pattern recognition algorithm has been used to train and test the data and to perform the classification. The confusion matrix is generated to analyze the performance results. The accuracy of the neural network based approach to recognize the emotions improves by applying multiple times of training. The overall correctly classified results for two times trained network is 73.8%, whereas it is 83.8% when increasing the training times to ten. The overall system provides a reliable performance and correctly classifying more than 80% emotions after properly trained.

**Keywords:** *Confusion matrix, Neural Networks, Short Term Features, Speech Emotions*