© 2018 IJSRSET | Volume 4 | Issue 7 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Categorization and Analysis of Emotion from Speech Signals

Vishal.A.Wankhade[1], Renuka Vijayrao Kukade[2]

[1]Assistant Professor, Electronics and Telecommunication Engineering, DMIETR, Wardha, Maharashtra, India

[2]Lecturer, Electronics and Telecommunication Engineering, VSPD, Amravati, Maharashtra, India

ABSTRACT

Recognizing emotion from speech has become one the active research themes in speech processing and in applications based on human-computer interaction. This paper conducts an experimental study on recognizing emotions from human speech. The emotions considered for the experiments include neutral, anger, joy and sadness. The distinguish ability of emotional features in speech were studied first followed by emotion classification performed on a custom dataset. The classification was performed for different classifiers. One of the main feature attribute considered in the prepared dataset was the peak-to-peak distance obtained from the graphical representation of the speech signals. Emotion is defined as the positive or negative state of a person’s mind which is related with a pattern of physiological activities. Emotions describe the mental state of a person. Sometimes in many applications such as military & civilian applications , in police department , its necessary to access whether a speaker is talking genuine or not and becoming increasingly important in security systems. So this project deals with the conditions like , if the speaker is involved in a stressful activity then the speech signal will be the significant indicator of the psychological stress. In this project speakers speech will be analysed depending on short time spectrum of vowels. For that we will have to take sample of some speech signals since the factors such as mood , emosion , physical characteristics are contained in the speech signal

Keywords : Emotion Analysis, Emotion Classification, Speech Processing, Mel-Frequency Cepstral Coefficients.

I. INTRODUCTION

Emotion classification is one of the most challenging tasks in a speech signal processing domain. The problem of speaker or speech recognition becomes relatively an easier one when compared with recognizing emotion from speech. Sound signal is one of the main medium of communication and it can be processed to recognize the speaker, speech or even emotion. The basic principle behind emotion recognition lies with analyzing the acoustic difference that occurs when uttering the same thing under different emotional situations. The mood of children is identified using speech signals. In addition to the features corresponding to the speaker and/or the speech, the sound signals do have some features that represent the emotional state of the speaker. The paper addresses the problem of emotion classification for human speech. The study is aimed at exploring dependencies the nature of utterance have with the human emotional state. Since the emotions have a direct influence on the nervous system, the heart rate also is affected by them. So the heart rate of a person can also be measured to get information about the emotional status of person. It is interesting to note that the speech signals are also a representative of the heart rate of the speaker since the heart rate also affects the speech. The work in says that if there is a

negative stimuli that causes negative emotion the heart rate decelerate more actively than when there is positive stimuli

II. SPEECH EMOTION RECOGNISION SYSTEM

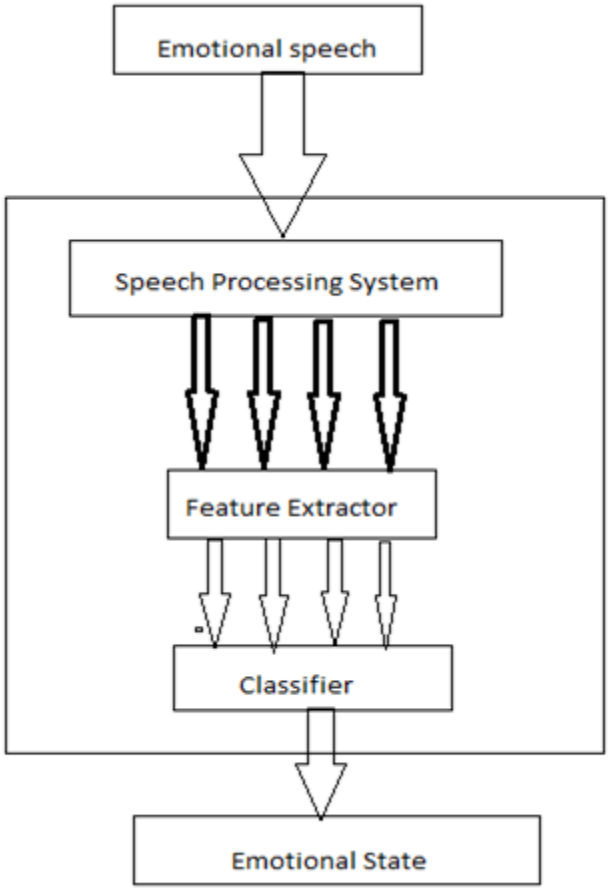

Fig -1: proposed system block diagram

The Mel-frequency cepstral coefficients (MFCC) are widely used in audio classification experiments due to its good performance. It extracts and represents features of speech signal. The Melcepstra takes short-time spectral shape with important data about the quality of voice and production effects . To calculate these coefficients the cosine transform of real logarithm of the short-term spectrum of energy must be done. Then it is performed in melfrequency scale. Further, after pre-emphasizing the speech segments are windowed . The Hamming window used for this process is a simple window based on reduction of leakage effect. It smears energy from true signal frequency into neighboring ones thus negatively affecting the performance. It also contributes to avoiding the discontinuity of the speech signal in time domain that might occur during Fast Fourier Transform. The concept of windowing is based on multiplying the signal frames by window function

III. FEATURE EXTRACTION

Speech is partitioned into small intervals known as frames. The process of partitioning speech into frames based on the information they are carrying about emotion is known as feature extraction. Feature extraction is the vital step in SER (speech emotion recognition) system. Some of the features that helps to figure out emotions from speech are-

3.1 PITCH:

It is the main component of any speech which is defined as the lowness or highness of a voice as identified by the human ears. Pitch is dependent on the vibrations per second. The value of pitch parameter is extracted by using cepstrum in the frequency domain. Pitch helps in identifying the nutral and angry emotions from speech sample.

3.2 ENERGY: -

Intensity of the speech defines the energy level of speech. Energy level for each frame is calculated as first the square of all sample amplitude is done and then summing up the values of all the squared sample amplitudes

3.3 PITCH DIFFERENCE AND ENERGY DIFFERENCE

The difference between values of pitch or energy level of neighboring segments is use to categorized speech parameters into emotions. The more the fluctuation the more it is easier to reveal the lively emotions like happiness and anger

3.4 FORMANTS: -

Formants are governed by the shape of the vocal tract and are manipulated by different emotions for eg, the state of excitement results in obtaining the higher mean values of the first formant frequency. The fundamental frequency(f0) helps in identifying happy emotion from speech samples.

3.5 MEL-FREQUENCY CEPSTRUM COEFFICIENT (MFCC): -

MFCC is the most vital parameter in which best describes the emotional state by using simple calculations. Mfcc also provides good frequency resolution when the speech frequency is low. MFCC based parameters show the energy migration in frequency domain and also helps in identifying phonetic characteristics of speech.

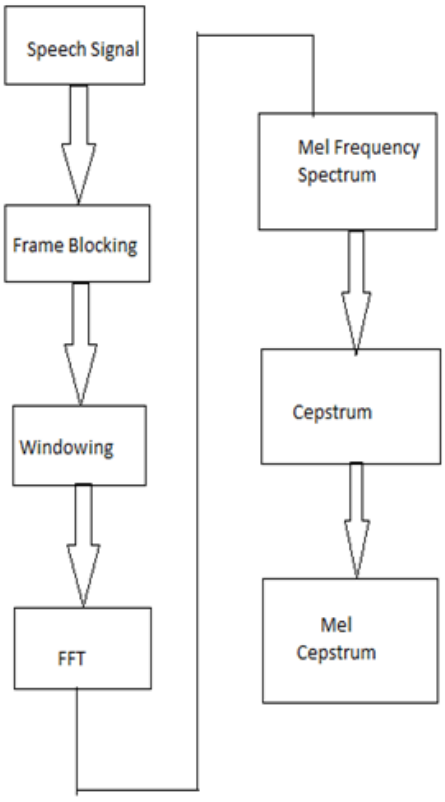

Fig -2: MFCC Block Diagram

IV. CLASSIFICATION ALGORITHM

An emotional state classification has a vital role in emotion recognision system using speech. The accuracy of classification, on the basis of different features extracted from the speech samples of different emotional state.The classifier is provided by proper features values to classify emotions. There are various types of classifiers such as K Nearest Neighbour (KNN) and Gaussian mixture Model (GMM) & Support Vector machine (SVM). Using any of these algorithm, emotional state can classified.

4.1 Support Vector Machine (SVM) Classifier :

The main motive of the SVM classifier is to track down the hyper-planes with maximum obtainable margin that sets apart the data points into classes by identifying a weight vector and an offset. Support Vector Machine (SVM) classifier uses binary classification based on statistical learning theory.SVM transforms the original input set to a high dimensional feature space with the help of kernel function. This renovation can also be used for transforming non-linear problems. SVM can have a very good classification performance even when there is a limited training data set. SVM has the capability to generalize new and accurate data by using the trained models designed in the learning phase. An adjustable weighted segmentation (AWS) is proposed to improve the accuracy rate of SVM classifier. AWS is a very simple approach in which each segment is assigned with a weight vector based on the type of emotion and the weights assigned are adjustable according to the input data.

V. CONCLUSIONS

The paper explores the idea of detecting the emotional state of a person by speech processing techniques. The study on words and letters under different emotional situations proved that the

emotional state can alter the speech signal. The development of a software based agent for emotion detection and heart rate analysis can greatly improve telemedicine based systems can also be improved.

VI. REFERENCES

[1]. Nwe, T. L., Wei, F. S., & De Silva, L.C.(2001). Speech based emotion classification. In TENCON 2001 Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology(Vol. 1. pp. 297-301),IEEE .

[2]. Casale, S., Russo, A., Scebba, G., & Serrano, S. (2008, August). Speech emotion classification using machine learning algorithms. In Semantic Computing, 2008 IEEE International Conference on (pp. 158-165). IEEE.

[3]. " Hassan, E. A., El Gayar, N., &Moustafa, M. G.(2010,November). Emotions analysis of speech for call classification. In Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on (pp.242-247). IEEE.[4]. "Theories of Emotion". Psychology.about.com. 13 September 2013. Reteieved 11 November 2013.