INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Speech based Emotion Recognition using various Features and SVM Classifier

Supriya jagtap[1], Dr. K. R. Desai[2]

[1]P. G Scholar, [2]Prof, H.O.D. Dept. of Electronics and Telecommunication Engineering, Bharati Vidyapeeth's college of Engineering, Kolhapur, (India)

Abstract: In this paper methodology for human emotion recognizes by extracting the speech signal. This speaker-based emotion recognition system recognizes the four emotions namely happiness, sadness, fear and angry. Basically, aim of this system to recognize the emotions and estimate the various features namely formant frequency, energy, pitch and MFCC from speech signal. accuracy of emotion detection system using speech signal depends on types of feature used to extract unique characteristics in case of individual emotion recognition. emotions are showing effective changes in specific features and hence SVM (Support Vector Machine) classifier can show better performance. Individual emotion classification shows better performance in our work but in case of average it is considerably less i.e. 68%.

Keywords: Berlin database, emotion recognition system, Feature extraction, PCA algorithm, SVM classifier

I. INTRODUCTION

Now a days in several application human machine interaction is widely used. Detection of feelings from speech may be a main challenge in human and machine. People express their emotions through different modalities like facial expressions, body poses and speech signal. We tend to are completely focused on speech signal, during this paper varied reasonably options that carries data regarding message, speaker, language and emotion.

The approach is to calculate numerous options that carries a lot of data and so mix these features to search out a strong recognition rate [2]. There are many applications in speech emotion recognition its helpful in healthcare center for detective work frustration. This technique is useful in E-learning field, detecting student's emotion timely and provides useful feedback to the teacher, it will facilitate the teacher to regulate the teaching set up and improve the training potency [3].

Speech based emotion recognition system is the best way to understand the feelings and actual intensions of the speaker. This technique additionally helpful in identification of mental standing like in criminal investigation, lie detection, intelligent help, disappointment, surprise/ amusement, healthcare and a better human computer interface [1][4]. This system involves feature extraction and analyze the parameters then determine the feelings based on its features such as energy, formant, pitch frequency and MFCC.

II. RELATED WORK

Under this point the focus is on the literature that is available for speech and emotions. There are so many people they done their work on emotion recognition by speech signal using different features and classifiers. The speech emotion recognition system has used various processes and some papers have been introduced here.

Anagha Sonawane et al. [2] have shown a method for emotion recognition which consist of MFCC based speech signal features extraction. After this, features are classified according to labels using Multiple Support Vector Machine (SVM). Emotional speech signals for anger, happy, neutral, surprise, sad and disgust are used in database. The results obtained by author are satisfactory.

H.K. Palo et al. [5] focused on a feature extraction by finding different cepstral coefficient-based features like perceptual linear prediction (PLP), Linear prediction coefficients (LPC), Mel frequency cepstral coefficient (MFCC) and linear prediction cepstral coefficient (LPCC) and identifies different classes of emotions by using multilayer perception (MLP). This paper acknowledges two categorize of speech emotions as induces like unhappy and bore at low responses and surprise and angry provides sturdy response.

Zhaocheng Huang [6] during this analysis, centered on feeling transition drawback are going to be investigate, as well as localizing feeling amendment points, identified feeling transition patterns and estimating or recognizing emotion changes.

Steven A. Rieger Jr et al. [7] the main target of this paper is on emotion recognition employing a pattern recognition paradigm with spectral feature extraction associated and gathering of k nearest neighbor (KNN) classifiers.

III.PROPOSED WORK

This proposed system consists a three major parts of emotion recognition system feature extraction, PCA algorithmic rule and classification.

Block diagram for speech emotion recognition

The mainstreamed diagram of speech-based emotion recognition is as shown below fig.1. In this system pre-processing of speech signal, feature extraction and classification of different classes are done. In this work our completely focus is on using different features. we've used formant frequency, pitch, energy and MFCC for the extraction of features option and classification of different emotions is completed with Support Vector Machine (SVM). During this work, we tend to use the berlin database with different emotions, such as happy, anger, worry and unhappy.
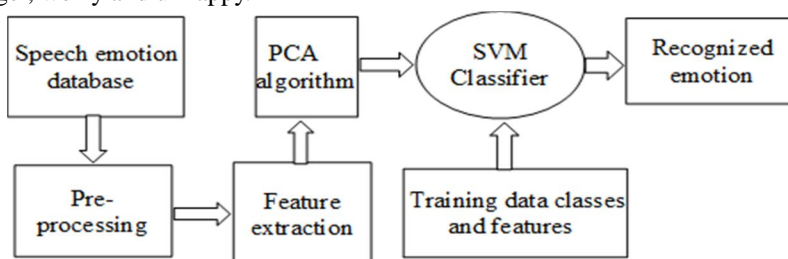

Figure.1Block diagram of speech emotion recognition.

A. Database

In our work, for the sake of evolution of emotions we have used German language-based Berlin database. Berlin database was recorded by F. Burkhardt et al. in German and this dataset was created by 800 statements (seven emotions* 10 actors* ten sentences + some second versions). Knowledgeable expert listeners selected ten individuals (five males and 5 females) for the establishment of database [8]. In this dataset seven classes of emotions are presented were fear, happiness, anger, neutral, happiness, disgust, dissatisfaction and disappointment. throughout this paper, we tend to acknowledge the four emotions like anger, fear, happy and unhappy [8].

B. Feature Extraction

Feature extraction block is most significant stage of feeling recognition. Speech contains lots of data regarding speaker throughout speech like message, language and actual intention of the speaker. during this technique while not losing any info and intention of speaker we discover here many parameters by using this feature extraction technique. This method reduces process complexness of signal. during this paper, we tend to completely center on feature extraction and analysis the results of formant frequency, energy, pitch and MFCC.

1) *Energy:* Energy is calculated from formant frequency that is nothing but square discrete time signal. The value of energy may be obtaining by calculating mean energy, maximum energy, minimum energy, energy range, standard deviation of energy and derivative of energy

$$E = \sum_{n=0}^{N-1} |x(n)|^2$$

2) *Formant Frequency:* Formants are nothing but the spectral peaks of the sound spectral peaks of the acoustic spectrum of the voice. The value of formant frequency using LPC may be obtained by calculating mean formant, maximum formant, minimum formant, formant frequency range, standard deviation of formant frequency and formant frequency bandwidth.
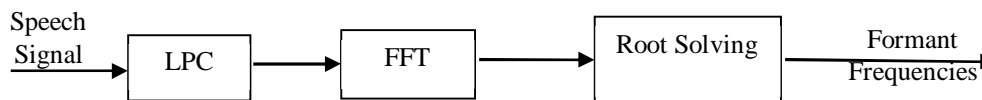

Figure.2 diagram of formant frequency detection using LPC

.

3) *Pitch Using Cepstrum:* Pitch in speech suggests that relative highness and lowness of tone is estimated by the ear. Pitch may be measured in time domain and frequency domain. number of vibrations per second produces by vocal cord is named pitch frequency. The value of pitch cepstrum obtained by finding mean pitch, maximum pitch, minimum pitch, energy range, standard deviation of pitch and derivatives of pitch.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 7 Issue III, Mar 2019- Available at www.ijraset.com

4) *MFCC Feature Extraction:* There are 2 sorts of filter that are spaced linearly at low frequency below one thousand Hertz and power (logarithmic) spacing on top of 1000Hz. during this technique Pre-emphasis could be a method which passes input audio through a filter because speech signals are continuously varying that filter emphasizes the signals range of frequency. And the next is framing, were speech samples prerecorded then divided signal into frame i.e. varying from 20-40msec time unit. once framing method done, windowing method is performed. during this technique, acting window is employed that reduces signals discontinuities from the beginning and end of every frame and that frame is shifted with 10ms span. the method of FFT converts every frame of N samples from time domain into frequency domain. The frequencies direct FFT spectrum is extremely wide then FFT algorithmic rule is employed for changing n samples from time domain to frequency domain Mel frequency reduces this scale and identifies what proportion energy is exists during a frame.

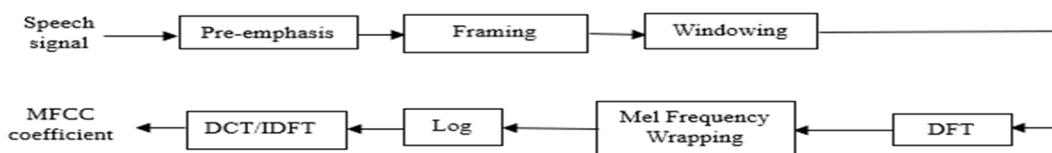F (Mel) = [2595 *log 10 [1+ f] * 700] DCT convert the log Mel spectrum into time domain [2] [10].

Figure.3 Calculation of MFCC Coefficients.

C. SVM Classifier

SVM classifier is one amongst the sturdy tools that uses a discriminative approach used for pattern recognition that associate SVM. SVMs use nonlinear and linear separating hyper-planes for knowledge classification [2][7] [11] [12]. Variable length knowledge is required to be reworked to mounted length vectors before using SVMs. it's a generalized linear classifier with maximum-margin fitting functions. This fitting operate provides regularization that helps the classifier to be generalized higher and the classifier ignores several of the options [2][7].

D. PCA Optimization

Principal component Analysis (PCA)is permits us to perform on a dimensionality reduction. PCA may be wont to cut back an outsized set of variables to a little set that also contains most of the data within the large set [13]. It maintains distinctive vector for unique PCA that's why process speed is additionally ultimately reduces.

IV. RESULT AND ANALYSIS

GUI shows steps performed within the system. during this system initial preprocessing of speech signal and have extraction is completed then optimize the features by using PCA and classification is done using SVM.
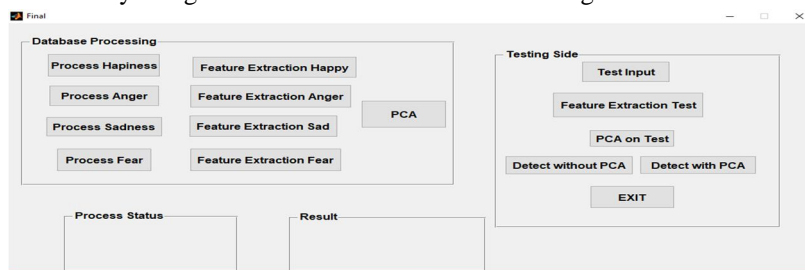
This are the results of speech signal, we will decide a speech signal is unvoiced or not and acquire voiced signals. Figure 3. shows the results of a voice handled by the preprocessing and have extraction process.
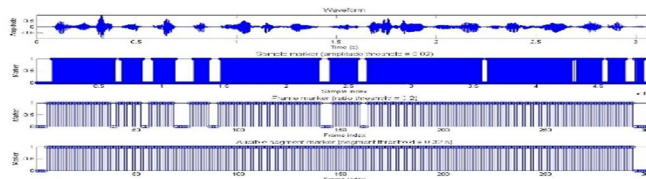

Figure 4. Result of a voice handled by the preprocessing and feature extraction processing.

Performance of the multiple SVM is measured based on total cross validation accuracy.

Accuracy = TP/TP+TN

Parameter	Happy	Sad	Angry	Fear	Average
Sensitivity	0.91	0.00	0.91	0.01	0.46
Specificity	0.92	0.58	0.91	0.55	0.74
Correct rate	0.92	0.44	0.74	0.42	0.68
Error rate	0.07	0.55	0.07	0.57	0.31
Accuracy in (%)	92.59	44.44	92.59	42.58	68

Table1. Performance analysis of multiple SVM

Individual emotion classification shows better performance in our work however just in case of average it's significantly less i.e. 68%.

V. CONCLUSIONS

In this work, emotion detection based on speech analysis using SVM classifier is bestowed. The performance evaluated shows better accuracy. Also, combination of features like formant frequency, pitch frequency, energy and MFCC are accountable to urge distinctive unique features and helps better accuracy. The performance evolution is completed on German speech database. just in case of worry and unhappy detection accuracy fails because of energy feature thought. If energy options are neglected during this 2 case higher performance is seen. This shows dynamic feature choice needed with relevance sort of feeling to be recognized. The work more may be extended for different language sorts having database recorded with correct acquisition.

REFERENCES

[1] M.S. Likitha, Sri Raksha R. Gupta, K. Hasitha and A. Upendra Raju, "Speech Based Human Emotion Recognition Using MFCC",This full-text paper was peer-reviewed and accepted to be presented at the IEEE WiSPNET 2017 conference

[2] Anagha Sonawane, M.U.Inamdar,Kishor B. Bhangale, "Sound based Human Emotion Recognition using MFCC & Multiple SVM",IEEE International Conference on Information, Communication, Intsrtumentation and Contol(ICICIC 2017)

[3] Yixiong Pan, Peipei Shen and Liping Shen,"speech emotion recognition using support vector machine",International Journal of smart home,vol.1,21 January 2014.

[4] A. Milton,S. Sharmy Roy,S. Tamil Selvi, PhD., "SVM Scheme for Speech Emotion Recognition using MFCC Feature",International Journal of Computer Applications (0975 – 8887) Volume 69– No.9, May 2013

[5] H.K. Palo, Mihir Narayana Mohanty and Mahesh Chandra, "Use of Different Features for Emotion Recognition Using MLP Network", © Springer India 2015 I.K. Sethi (ed.), Computational Vision and Robotics, Advances in Intelligent Systems and Computing 332, DOI 10.1007/978-81-322-2196-8_2

[6] Zhaocheng Huang, "An Investigation of Emotion Changes from Speech", 2015 International Conference on Affective Computing and Intelligent Interaction (ACII)

[7] Steven A. Rieger Jr., Rajani Muraleedharan, Ravi P. Ramachandran, "Speech Based Emotion Recognition Using Spectral Feature Extraction and an Ensemble of Knn Classifiers",2014 9th InternationalSymposium on Chinese Spoken LanguageProcessing (ISCSLP)

[8] Qingli Zhang, Ning An, Kunxia Wang, Fuji Ren and Lian Li , "Speech Emotion Recognition using Combination of Features", 2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP) June 9 – 11, 2013, Beijing, China

[9] Supriya B. Jagtap, Dr.K.R.Desai, Ms.J.K.patil, " A Survey on Speech Emotion Recognition Using MFCC and Different classifier", 8th National conference on emerging trends in engineering and technollogy(NCETET-2018),ISBN:978-93-87793-03-3

[10] Leh Luoh,Yu-zhe Su and Chih-Fan Hsu, "Speech signal processing based emotion recognition", 2010 International Coriference on System Science and Engineering

[11] F. Yu, E. Chang, Y.Q. Xu, and H.Y. Shum, "Emotion detection from speech to enrich multimedia content", in Proc. 2nd IEEE Pacific-Rim Conference on Multimedia 2001, pp.550-557, Beijing, China, October 2001.

[12] Tsuyoshi Moriyama, Shinya Mori, and Shinji Ozawa. "A synthesis method of emotional speech using subspace constraints in prosody". Journal of Information Processing Society of Japan, 50(3):1181–1191, 2009. (in Japanese).

[13] Introduction to Principal Components and Factor Analysis, AtchleyOct19

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)