International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 06 | June 2019 www.irjet.net p-ISSN: 2395-0072

Study of Effect of PCA on Speech Emotion Recognition

Supriya B. Jagtap[1], Dr. K.R. Desai[2]

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

Abstract: *In smart environment control based on human moods may involve human emotion recognition with variety of techniques. Speech based emotion recognition can be considered in variety of applications where voice command-based controls are depleted. Speech emotion recognition means extracting the emotional state and synthesizing actual intension of a speaker. In this paper we focused on emotion recognition by extracting various features of speech signal also we compare the results by discriminating features using PCA techniques. This voice command-based emotion recognition system recognizes the four emotions namely angry, sad, happy and fear. We evaluate the results using accuracy, sensitivity and specificity parameters for SVM based classifier results.*

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

I. INTRODUCTION

In human life emotion plays vital role, these emotions people can be express through different modalities like facial expression, body poses and speech signal. Speech is one of the natural forms of communication in day to day life. This system aims to grasp feeling and acknowledges the emotions which are exactly present in speech and automatically determine this emotion [1]. Speech based emotion recognition highly depends on speaker and culture so different people have different style to express their emotions. It's a voice-based tool that can express your desires, feelings, thoughts and needs. This paper is mainly focused on audio and audio-based system is built on different aspects. Extensively, these audio signals give information about who has spoken, what has been said, information about which language is spoken and gives information about the true feelings of speech. Voice based emotion recognition aims to recognize and analyze the actual emotional state of the speaker and it recognize the emotions like Sad, Angry, Happy and fear.

The trend of globalization and also the widespread quality of the Web emphasizes the capability of human-computer interaction strategies [2]. This technique is enforced within the numerous fields like health care centers for monitoring the mental conditions of patients, in a telecom communication centers for determine the person's actual intention according to the reaction and identifying the exact mood. This system can also be useful for detecting psychological studies, the sense of identity in the field of forensics is one of the key uses, understanding the person's feelings and analyzing them for medical treatment or finding the person statement whether the truth or lie in specific cases.

In speech-based emotion recognition system involves four main steps first is database which is present in three types namely natural, simulated and induced database. Second step is feature extraction, we have considered four features like MFCC, pitch frequency, formant frequency and energy. And then third step is feature vector reduction by using PCA algorithm, it shows working of PCA algorithm in the system and shown the results of system with PCA and without PCA which is followed by experimental evaluation [1]. And last step is recognizing the various categories of emotion by using SVM classifier. In the next section, the previous related work on speech emotion recognition system is explained.

II. LITERATURE REVIEW

Under this point the focus is on the literature that is available for speech and emotions. There are some people they done their work on speech-based emotion recognition by using feature extraction and PCA algorithm. The speech emotion recognition system has used various processes and some papers have been introduced here.

S. Radha Krishna et. al. [2] have shown a method for emotion recognition which is focused on spectral features like skewness, Pitch chroma, MFCC and Centroid. And then SVM classifier is used for identification of different classes of emotion and analyze the results of recognition performance for feature sets after applying a PCA algorithm.

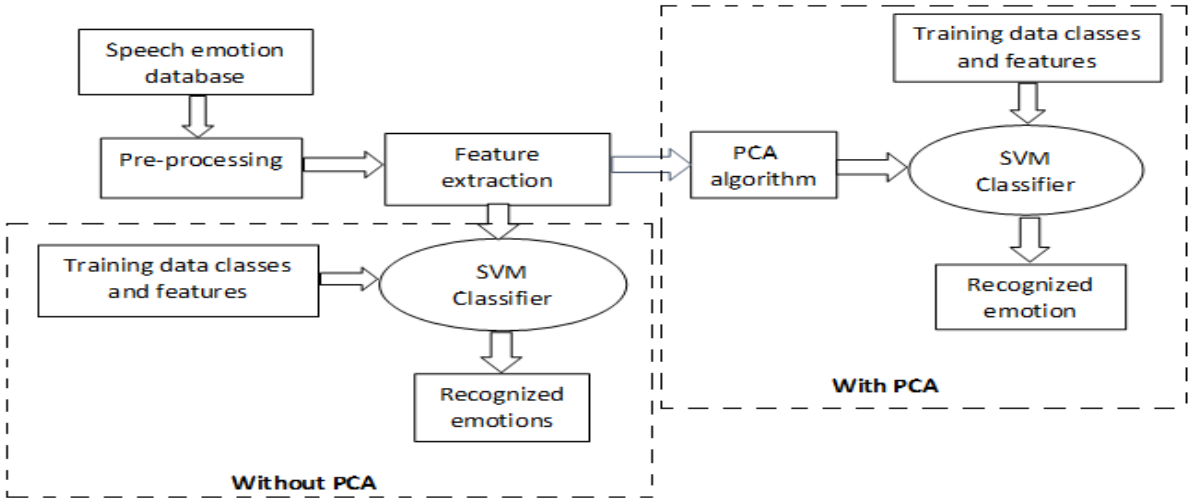
Manjushree B. Aithal et al. [3] focused on speech and emotion recognition using distorted speech signal. During this analysis, for feature extraction MFCC is used on distorted speech signal and PCA is used for evaluating eigen values.

They analyze highest Eigen values contain important information which are keep hold on to these highest values, and others are discarded as noise. And then recognize the emotions by using Hidden Markov Models.

III. PROPOSED WORK

This planned system consists a three major parts of feeling recognition system feature extraction, PCA algorithmic rule and classification by using multiclass SVM.

Block diagram for speech emotion recognition



Block dig1. Speech emotion recognition system

The generalized diagram of voice-based emotion recognition is as shown below fig.1. Main stages of sound-based emotion recognition are pre-processing of speech signal, feature extraction and classification of different classes by using multiclass SVM. For our implementation completely focus is on using different features and evaluating the better results by using PCA algorithmic rule. For feature extraction we've used formant frequency, pitch, energy and MFCC and classification of different emotions is completed with multiclass Support Vector Machine (SVM). During this work, we tend to use the berlin database with varied emotions, such as anger, worry, unhappy and happy

A. Database:

In this system database is used for emotion recognition research. Database is divided into three types 1. Natural database this database is the collection of real-world conversation like call center conversation. 2. Simulated database this database created from actors or artists, expressing different emotions by using linguistic neutral sentences. 3. Induced database is recorded by artificially creating the emotional situations without the knowledge of speaker.in this paper we have used (4). 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) (5).

Table 1. Number of emotional speech files in Berlin Database

Emotions	No. of speech files
Anger	123
Fear	34
Happy	63
Sad	24

B. FEATURE EXTRACTION

Feature extraction is a main step in speech emotion recognition system. There is number of parameters are present in speech signal. Speech signals contains lots of information about speech, by using this technique extracts and selects best features from speech signal. Selecting the suitable features which brings more information about emotions from voice signal. It is not clear which speech feature are most powerful in distinguishing between emotions. In this work, we focused on four types of features namely energy, pitch, MFCC and formant frequency.

i. Energy:

Energy is that the basic and most significant feature in speech signal. To get the statistics of energy feature, we tend to use short run perform to extract the value of energy in every speech frame. Then we are going to acquire the statistics of energy inside the total speech sample by conniving the energy, like mean energy, max value, energy range, standard deviation and derivatives of energy.

ii. Pitch:

The pitch signal has data regarding feeling, as a result of it depends on the stress of the vocal folds and therefore the sub speech organ gas pressure, that the mean of pitch, maximum pitch, minimum pitch, energy range, standard deviation of pitch and derivatives of pitch.

iii. MFCC features:

MFCC is the spectral type of feature extraction technique. MFCC features Mel-Frequency Cepstrum coefficients is that the most vital feature of speech with straightforward calculation, smart ability of distinction, anti-noise [6]. MFCC within the low frequency region contains a smart frequency resolution, and the strength to noise is additionally superb. MFCC is taking index once Mel Filter bank and Frequency wrapping. After that, we tend to conjointly cipher first and 2d distinction concerning this feature.

iv. Formant frequency:

In audio signal processing for representing the spectral envelope of digital signal of speech in compressed form, we are using the linear predictive coding (LPC). This one is also most powerful speech analysis technique, and it provides accurate results of speech parameters. 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

C. PRINCIPAL COMPONENT ANALYSIS

Principal Component analysis (PCA) is usually used as technique for data reduction/compression with no loss of data. It's a way accustomed rework one set of variables into another smaller set, wherever the new created variable isn't straightforward to interpret. In many applications, PCA is employed solely to supply data on actuality spatial property of an information set. If the information set includes M variables, all M variables don't represent needed data. PCA transforms a collection of related variables into a brand-new replacement set of uncorrelated variables that are referred to as principal elements. However, but if the information is already unrelated the PCA is of no use. Together with the unrelated information, the principal elements are orthogonal and are ordered in terms of the variability they represent. That is, the primary principal part represents, for one dimension, the best quantity of variability within the original information set. PCA may be applied to information sets containing any number of variety variables [3].

To decorrelate the variables, we want to rotate the variables data set till the information points are distributed symmetrically regarding the mean. Within the decorrelated condition, the variance is maximally distributed on the orthogonal axes. It's additionally typically necessary to center the information by removing the mean before rotation. In applied mathematics sense, if 2 variables are freelance, they're going to even be unrelated however reverse isn't true. The rotation is therefore performed that the covariance (or correlation) goes to zero. A higher thanks to come through zero correlation is to use a method from algebra that generates a rotation matrix that reduces the covariance to zero. One well-known methodology is by pre- or post-multiplication with the orthonormal matrix:

U'CU= D…….. (3.1)

where, C is m-by-m covariance matrix, D is a diagonal matrix, U is an orthogonal matrix that does transformation. The diagonal parts of D are the variances of the new information, usually referred to as the characteristic's roots, or eigenvalues of C: λ_1, λ_2..... λ_n. The eigenvalues of the new variance matrix correspond to the variances of the revolved variables. The eigenvalues will be obtained as:

$\det|C - \lambda| = 0$ ..... (3.2)

where, I is that the unit matrix. once getting, the eigenvectors are obtained as:

$\det|C - \lambda I|Vi = 0$ .....(3.3)

where, the eigenvectors are obtained from by the equation given below,

$u_i = v_i/\sqrt{V_i' v_i}$ ..... (3.4)

These eigenvectors are Feature vector that is increased with the input file because of which we tend to get the new data set is : information adjusts= Feature vector' x Final data

This is often however PCA reduces redundant data. just in case of speech signal.

D. SVM CLASSIFER

In this work we are using a multiclass support vector machine (SVM) classifier to classify four different classes of emotions like angry, surprise, happy and sad. This is a multiclass SVM classifier, classifier firstly trained with four different classes and then also trained by extracted features of speech signals and selects only best features from speech signals. After training a multiclass SVM then test input entered and this input signal is matched with trained speech features and four different classes. Then finally recognize the emotion after matching the parameters.

IV. RESULT AND ANALYSIS

This are the results evaluated with PCA algorithmic rule which evaluates the overall emotion recognition in terms of parameters like sensitivity, specificity, correct rate, error rate and accuracy of the system is presented. The recognized emotions are anger, fear, sad and happy.

The results are reported in two states the first part shows the performance of the system with PCA algorithmic rule. The second part of the result provides the performance of the system without the PCA rule. The experimental results highlight that the PCA based results allows incrementing accuracy of emotion recognition system on average. And as shown below the results of Before and after PCA Obtained values of accuracy tables followed by Graphs of corresponding tables. The emotions been considered are angry, fear, happy and sad.

Table: 1 Performance evaluation of Emotion Recognition with PCA

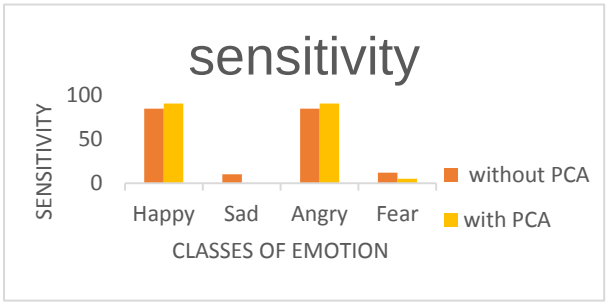
Parameter	Happy	Sad	Angry	Fear	Average
Sensitivity	91	0	91	1	46
Specificity	92	58	91	55	74
Correct rate	92	44	74	42	68
Error rate	7	55	7	57	31
Accuracy in (%)	92.59	44.44	92.59	42.58	68

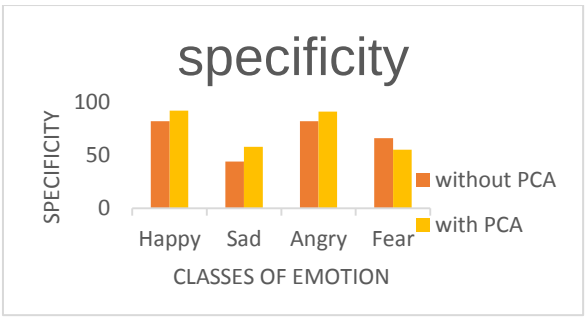
Table: 2 Performance evaluation of Emotion Recognition after PCA

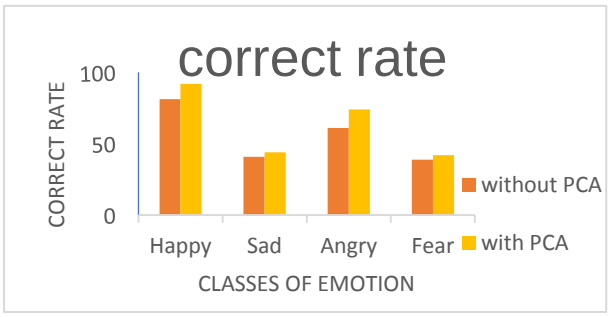
Parameter	Happy	Sad	Angry	Fear	Average
Sensitivity	85	10	85	12	48
Specificity	82	44	82	66	68.50
Correct rate	81.19	43.59	61.19	39	56.24
Error rate	12	60	14	66	38
Accuracy in (%)	88	40	86	44	64.50

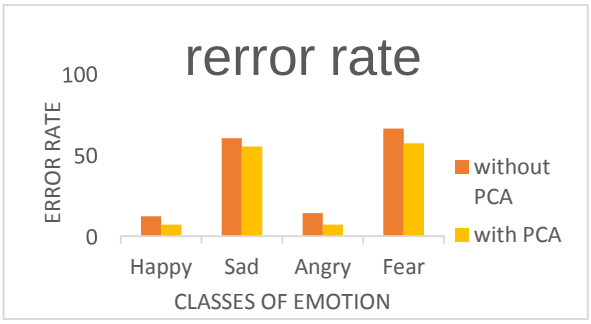
In this work performance evolution is done with and without PCA and the angry emotion is confused with the happy emotion. The bar graphs given shows number of emotions recognized with PCA and without PCA and with considering different parameters like sensitivity, specificity, correct rate, error rate and accuracy.

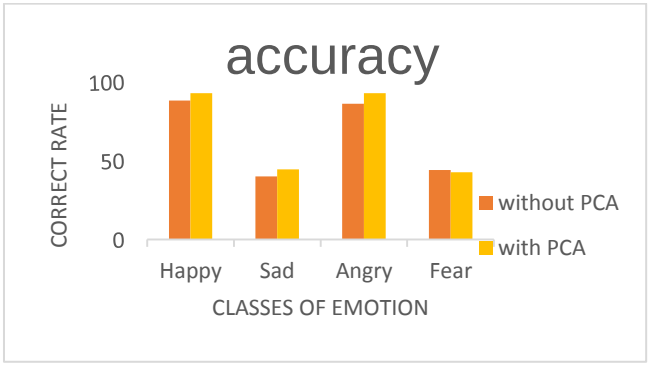
Graph1. Performance evaluation of Emotion Recognition before PCA and after PCA with different parameters.











V. CONCLUTION

The purpose of user discrimination is to reduce computational complex of SVM as a discrimination leads to reduction in size without affecting features uniqueness with respect to specific speech signals when done by using PCA. Based on our revolution we can conclude that the required optimization for specific features using multiclass SVM for desired accuracy requires more computational complexity and hence PCA can solve the problem and is responsible for considerable reduction in computations and better accuracy.

VI. REFRENCES

[1] Anuja Bombatkar, Gayatri Bhayar, Khushbu Morjani, Shalaka Gautam, Vikas Gupta, “Emotion recognition using k-nearest neighbor algorithm”, IJERA, ISSN:2248-9622, April 2014

[2] S. Radha Krishna1 and R. Rajeswara Rao 2 1JNTUK-University College of engineering Vizianagaram, Vizianagaram, “SVM based Emotion Recognition using Spectral Features and PCA”, International Journal of Pure and Applied Mathematics, Volume 114 No. 9 2017, 227-235

[3] Manjushree B. Aithal, Pooja R. Gaikwad, & Shashikant L. Sahare, “Speech Enhancement Using PCA for Speech and Emotion Recognition”, G.J. E.D.T, Vol.4(3):6-12, (May-June, 2015), ISSN: 2319 – 7293

[4] 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:978-93-87793-033

[5] Supriya jagtap[1], Dr. K.R. Desai, "Speech based emotion recognition using various features and SVM classifier", International Journal for Research in Applied Science &Engineering Technology (IJRASET), ISSN:2321-9653; IC Value45.98; Volume 7 III,Mar 2019

[6] Akalpita Das, Laba Kr. Thakuria, Purnendu Acharjee, Prof. P.H. Talukdar, "A BRIEF STUDY ON SPEECH EMOTION RECOGNITION", International Jouranal of Scientific & Engineering Research. Volume5, Issue1, January 2014, ISSN 2229-5518

[7] Bageshree V. Sathe-Pathak, Ashish R. Panat, " Extraction of Pitch and Formants and its Analysis to identify 3 different emotional states of a person ", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 1, July 2012 ISSN (Online): 1694-0814

[8] Sheguo Wang,Xuxiong Ling,Fuliang Zhang,Jianing Tong, "Speech Emotion Recognition Based on Principal Component Analysis and Back Propagation Neural Network", 2010 International Conference on Measuring Technology and Mechatronics Automatio, 978-0-7695-3962-1/10 $26.00 © 2010 IEEE DOI 10.1109/ICMTMA.2010.523

[9] Bo-Chang Chiou and Chia-Ping Chen, "Feature Space Dimension Reduction in Speech Emotion Recognition Using Support Vector Machine",

[10] Sheguo Wang,Xuxiong Ling,Fuliang Zhang,Jianing Tong, "Speech Emotion Recognition Based on Principal Component Analysis and Back Propagation Neural Network, 2010 International Conference on Measuring Technology and Mechatronics Automation "

[11] 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).

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

[13] M. Morales-Perez, J. Echeverry-Correa, A. Orozco-Gutierrez and G. Castellanos-Dominguez, "Feature extraction of speech signals in emotion identification", 30th Annual International IEEE EMBS Conference Vancouver, British Columbia, Canada, August 20-24, 2008

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