Recognizing emotions of Marathi isolated speech from energy and formants using LDA

Shaikh Nilofer Rais Ahmed, Ratnadeep R. Deshmukh, Vishal B. Waghmare

Department of Computer Science and IT,
Dr. Babasaheb Ambedkar Marathwada University,
Aurangabad-431004 (MS) India.

Abstract

Many attempts have been made for developing a system which can recognize emotions from speech signals effectively. The research shows that various features namely prosodic and spectral have been used for emotion recognition from speech. This paper is one search approach towards same. The database used for recognition purpose was developed on Marathi language using 100 speakers. Here we have extracted the prosodic features i.e. energy and formants from speech signals. We have also applied linier discriminates analysis as the classification technique. Further we have compare the results gain on the basis of extracted features and came to know that formant gives quit good results compared to that of energy.

Keywords- Database, Emotion Recognition, Formants, Feature extraction, Linier discriminant Analysis, Speech.

I. INTRODUCTION

There is an enormously important role of emotions in human life. As per human's perspective or feelings emotions are essential medium of expressing his or her psychological state to others. Humans have the natural ability to recognize the emotions of their communication partner by using all their available senses. They hear the sound, they read lips, they interpret gestures and facial expression Humans has normal ability to recognize an emotion through spoken words but since machine does not have capability to analyze emotions from speech signal for machine emotion recognition using speech signal is very difficult task. Automatic emotion recognition paid close attention in identifying emotional state of speaker from voice signal [1].

Human emotion is a visible manifestation of affective State, cognitive activity, emotional state and personality. Like face detection, human emotion analysis is also a very challenging area of research in the field of computer vision and machine learning [2].

Emotion recognition through speech is an area which is increasingly attracting the attention the field of pattern recognition and speech signal processing in recent years. Automatic emotion recognition pays close attention to identify emotional state of speaker from voice signal. It is important medium of expressing humans perspective or fillings and his or hers mental state to others. Humans have natural ability to recognize emotions through speech information but the task of emotion recognition for machine using speech signal is very difficult since machine does not have sufficient intelligence to analyze emotions from speech [3].

Till date many speech recognition systems have been proposed. Researchers have been using various techniques for identifying emotions. This includes accurate feature extraction and selection and further applying proper classifier These systems used various features viz. Prosodic and spectral where prosodic features included Pitch, Speech intensity glottal parameters and Spectral features included Mel-frequency cepstral coefficients(MFCC) and Linear Predictive cepstral coefficients LPCC. The different classifiers used for emotion recognition are Hidden Markov Model (HMM), Gaussian Mixture Model (GMM), Support Vector Machine(SVM), Artificial Neural Network(ANN). A study using the features Pitch and energy and classifier HMM was performed in where the accuracy rate achieved was 86%. [4]

II. DATABASE SAMPLES

The database consists of five emotions each having 8 words. The database was recorded in noisy environment using PRAAT software. The frequency was set 16000HZ. Each word was recorded thrice. In total there were 100 speakers 50male and 50female.

TABLE I. SOME OF THE SAMPLES OF DATABASE

Neutral Emotion	English
मग काय	What now
बोला आता	Say now
बरं बरं	Ok ok

a. Some of the samples of Neutral emotion

Happy Emotion	English
किती छान	How nice
लय भारी	Awesome
अभिनंदन	Congrats

b. Some of the samples of Happy emotion

Sad Emotion	English
अरे देवा	Oh my god
काय करु	What to do
नकोये मला	I don't want

c. Some of the samples of Sad emotion

III. FEATURES EXTRACTION

Prosodic features include Intensity, Pitch, and Energy. The mean Standard deviation, minimum, maximum, range and variance of Pitch, energy and other similar features are used for distinguishing emotions. In another study the peaks and troughs of fundamental frequency and intensity are studied which gave 55%accuracy for four emotions namely sad, fear, joy and anger [5].

A. Energy

The speech signal energy provides a representation in terms of amplitude variations. The analysis of energy is paying attention on short-term average amplitude and short-term energy. In this short time energy features estimated energy of emotional state by using variation in the energy of speech signal. To obtain the statistics of energy feature we implied short-term function to extract the value of energy in each speech frame.

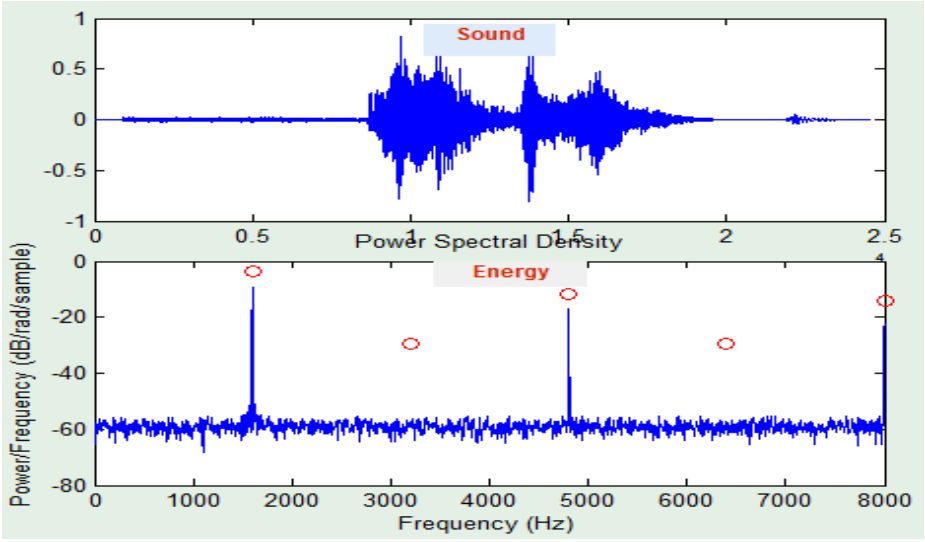

Fig. 1 Original sound signal and energy.

B. Formant

Formant is a very important parameter to reflect sound track features. First of all, linear prediction is applied to calculate the 14 order prediction coefficients, and then, those coefficients are used to estimate the sound track frequency response curve, and last, the peak picking method is adopted to calculate the frequency of every formant [6]. The average formant frequency and formant frequency changing rate of the first formant; the average peak values and the average slope of regression curve of formant peak of the first 4 formants. Choose the difference between the first average formant frequency of every frame, the average values and slopes of the peak value regression curves of the first 4 formants, and the parameters of the corresponding peaceful Sentences; also choose the rate of first formant frequency changing rate over corresponding sentence [7].

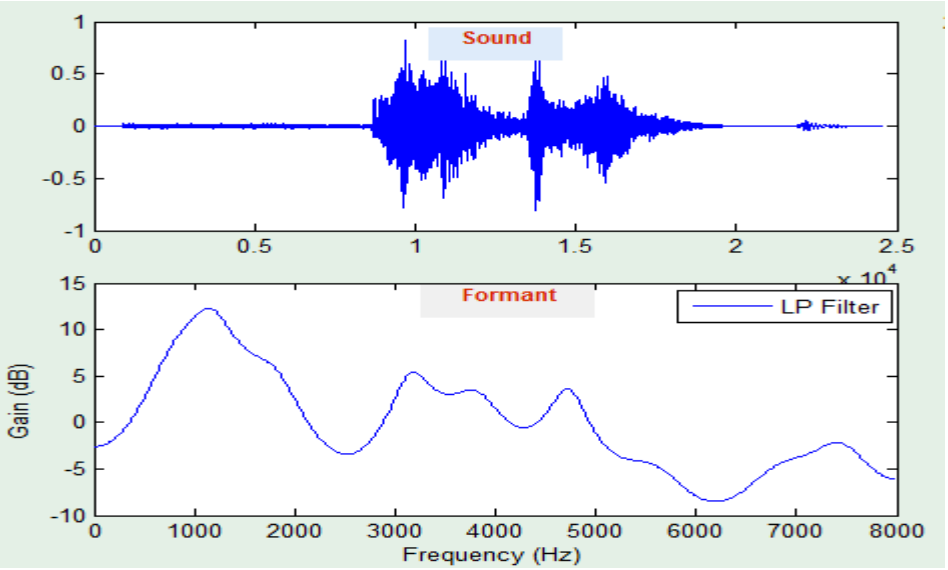

Fig. 2 Original sound signal and formant.

IV. CLASSIFIER

A. Linear Discriminate Analysis

The linear discriminate analysis (LDA) is a statistical technique to classify data into mutually exclusive and exhaustive groups based on a set of measurable data's features. It is a well-known scheme for feature extraction and dimension reduction and has been used widely in many applications involving high dimensional data, such as pattern recognition, supervised learning and data classification. In LDA, the directions that will give a good separation to the different classes of the data need to be located. This is attained by projecting the data onto a lower-dimensional vector space so that the ratios of the between class distance to the within class distance is maximized, in order to achieve maximum separation. Here, if the data contains only two features, the separators between data groups will become lines. If there are three features, the separator will become a plane and if the number of features is more than three, the separator will become a hyper-plane.

The process of LDA is summarized below:

1. Calculating the within class scatter matrix. The amount of scatter between training data in the same class is first calculated using Equation (1) where Si is scatter matrix, mi is the mean of the training data xi within the class i and Xi is the covariance matrix of n the data. The within class scatter matrix Sw is then calculated as the sum of all the scatter matrices as shown in Equation (2) where C is the number of classes. A

$$S_i = \sum_{x \in X_i} (x - m_i)\,(x - m_i)^{-1} \quad (1)$$

$$S_w = \sum_{i=1}^{C} S_i \quad (2)$$

2. Calculating the between class scatter matrix. The amount of scatter between classes Sb is measured using Equation (3) where ni is the number of data in the ith class, m is the total mean of all training data, mi is the mean value of each class and C is the number of classes.

$$S_b = \sum_{i=1}^{C} n_i (m_i - m)(m_i - m)^{-1} \quad (3)$$

3. Calculating the generalized eigenvectors and eigenvalues. The generalized eigenvectors and eigenvalues of the within class and between class scatter matrices are computed.
4. Sorting the order of eigenvectors. The eigenvectors are sorted in a descending order depending on the magnitude of the eigenvalues. Here, the sorted eigenvectors form the Fisher basis vector.
5. Projection of the training data onto the Fisher basis vectors. The training data are projected onto the Fisher basis vector by calculating the dot product of the training data with each of the Fisher basis vectors. Important feature is energy of speech signal. Speech energy is having more information about emotion in speech [8].

The following table 2 displays the approximately achieved accuracy for Angry, stress, admiration, teasing and shocking using pitch and formants as discriminate factor.

TABLE II. ACCURACY RESULT

Emotion	Energy	Formant
Neutral	70%	100%
Happy	50%	90%
Sad	60%	90%
Surprise	90%	100%
Boredom	70%	100%

V. CONCLUSION

Using energy and formant Surprise can be 100% and 90% recognized where as happy gives the lowest accuracy that is of 50% accuracy with energy. Boredom, neutral and sad gives 70%, 70% and 60 % of accuracy with energy respectively. Whereas using formant we get 100%, 100% and 90% for the same emotions. Thus formant plays a vital role in recognizing emotions and we can conclude that using formant we can accurately recognize emotions compare to that of energy.

REFERENCE

[1] Chandra Praksah , Prof. V. B. Gaikwad, "Analysis Of Emotion Recognition System Through Speech Signal Using KNN, GMM & SVM Classifier", International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume 4 Issue 6 June 2015, Page No. 12523-12528a.

[2] Suvam Chatterjee,Hao Shi, "A Novel Neuro Fuzzy Approach to Human Emotion Determination,"2010 Digital Image Computing: Techniques and Applications, 978-0-7695-4271-3/10 $26.00 © 2010 IEEE,DOI 10.1109/DICTA.2010.114

[3] Rani P. Gadhe, R. R. Deshmukh, V. B. Waghmare, P. P. Shrishrimal, "Emotion Recognition From Speech:A Survey," ISSN 2229-5518, International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015.

[4] Shaikh Nilofer R. A.,Rani P. Gadhe, R. R. Deshmukh, V. B. Waghmare, P. P. Shrishrimal ," Automatic emotion recognition from speech signals: A Review," ISSN 2229-5518,International Journal of Scientific & Engineering Research, Volume 6, Issue 4, April-2015

[5] Rao, K. S., & Yegnanarayana, B." Prosody modification using instants of significant excitation," IEEE Transactions on Audio,Speech, andLanguage Processing, 14, 972–98, .2006.

[6] Xin Min Cheng ,Pei Ying Cheng, Li Zhao," A Study On Emotional Feature Analysis And Recognition In Speech Signal," International Conference On Measuring Technology And Mechatronics Automation, 978-0-7695-3583-8/09 © 2009 IEEE Doi 10.1109/Icmtma.2009.89.

[7] McGilloway, S., Cowie, R., Douglas-Cowie, E., Gielen, S.,Westerdijk, M., & Stroeve, S. (2000)."Approaching automatic recognition of emotion from voice," A rough benchmark. In ISCA workshop on speech and emotion, Belfast.

[8] Li Wern Chew, Kah Phooi Seng, Li-Minn Ang, Vish Ramakonar3 and Amalan Gnanasegaran" Audio-Emotion Recognition System using Parallel Classifiers and Audio Feature Analyzer", 2011 Third International Conference on Computational Intelligence, Modelling & Simulation, 978-0-7695-4562-2/11 $26.00 © 2011 IEEE DOI 10.1109/CIMSim.2011.44