Review Article | Impact Factor: 0.621 | ISSN: 2319-507X
Aparna Wanare, IJPRET, 2014; Volume 2 (8): 894-899 | IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

HUMAN EMOTION RECOGNITION FROM SPEECH: A REVIEW

MISS. APARNA P. WANARE[1], PROF. SHANKAR N. DANDARE[2]

1. Department of Electronics & Telecommunication S.G.B.A.U. Amravati (India).
2. Babasaheb Naik College of Engineering, Pusad – 445 215 (India).

Accepted Date: 15/02/2014 ; Published Date: 01/04/2014

Abstract: Field of emotional content recognition of speech signals has been gaining increasing interest during recent years. Emotion recognition from speech signals is one of the important research areas which add value to machine intelligence [2]. Several emotion recognition systems have been constructed by different researchers for recognition of human emotions in spoken utterances. Here a brief review about the work done in emotion recognition on different spectral and prosodic features using various classifiers is presented. In this paper, identification of basic emotional states such as anger, joy, neutral and sadness from human speech is addressed. The database for the speech emotion recognition system is the emotional speech samples and the features extracted from these speech samples are helpful for detection of emotion. Several features can be extracted from speech signals. However, not all features for speech recognition are of equal importance for emotion recognition. This paper proposes a systematic method on feature selection for emotion recognition from speech signals. The idea is to work on a well-selected small feature set and use it to remove irrelevant information.

Keywords: Emotion Recognition, Classifier, Spectral and Prosodic features, Features Extraction and Selection, Database

PAPER-QR CODE

Corresponding Author: MISS. APARNA P. WANARE

Access Online On:

www.ijpret.com

How to Cite This Article:

Aparna Wanare, IJPRET, 2014; Volume 2 (8): 894-899

INTRODUCTION

Emotion recognition through speech is an area which increasingly attracting attention within the engineers in the field of pattern recognition and speech signal processing in recent years. Speech is a complex signal which is a combination of information about message, speaker language and emotions. Emotion, which is a non-linguistic component of a speech, is used widely by human beings for expressing their intentions and emotions from voice signal [5]. 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 analyse emotions from speech [2]. Emotion recognition through speech is particularly useful for applications in the field of human machine interaction to make better human machine interface. Some other applications which require natural man machine interaction such as Interactive movie, storytelling and electronic machine pet, remote teach school & E-tutoring application where response of system depends on the detected emotion of users which makes it more practical. Other applications of the emotion recognition system are lie detection, in the psychiatric diagnosis, intelligent toys, in aircraft cockpits, in call centre and in the car board system [1]. Different intelligent systems have been developed by researchers on the basis of some universal emotions which includes anger, happiness, sadness, surprise, neutral, disgust, fearful, stressed etc. in last two decades. This different system also differs by different features extracted and classifiers used for classification. Prosodic features and spectral features contain most of the important information that can be used for emotion recognition from speech signal. Some of the prosodic features are pitch, energy, fundamental frequency, loudness, speech intensity and glottal parameters. Mel-frequency cepstrum coefficients (MFCC) and Linear predictive cepstral coefficients (LPCC) are some of the spectral features. Also some of the linguistic and phonetic features also used for detecting emotions through speech. Types of classifiers which are being used for emotion recognition such as Hidden Markov Model (HMM), k-nearest neighbors (KNN), Artificial Neural Network (ANN), super vector based SVM classifier and Gaussian Mixtures Model (GMM).

Basic Block Diagram of Speech Emotion Recognition

The block diagram of the emotion recognition system through speech considered in this study is illustrated in Figure 1. The block diagram consists of the emotional speech as input, feature extraction, feature selection, classifier and detection of emotion as the output. The emotional speech input to the system may contain the collection of the acted speech data the real world speech data. From collection of database necessary features such as prosodic and spectral are

extracted and feature containing most of the information are selected for processing. Then processed signals are given to classifier for classification and desired speech output is obtained.

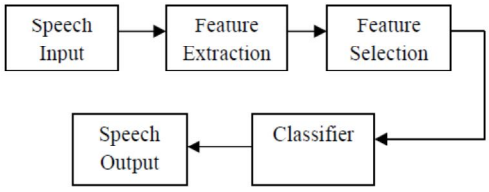

Fig. 1: Basic Block Diagram of Emotion Recognition

Extraction and selection of feature

- An important step in emotion recognition system through speech is to select a significant feature which carries large emotional information about the speech signal. Changes in these parameters indicate changes in the emotions. Speech features have two main types which are phonetic features and prosodic features. The types of sounds involving speech are considered as phonetic features, for example vowels and consonants and their pronunciation. Prosodic features on the other hand are leaning towards the musical aspects of speech, such as rising or falling tones and accents or stresses [4].

1. Energy:

The Energy is the basic and most important feature in speech signal. Energy frequently referred to the volume or intensity of the speech, where it is also known to contain valuable information. Energy provides information that can be used to differentiate sets of emotions, but this measurement alone is not sufficient to differentiate basic emotions. Scherer stated that fear, joy, and anger have increased energy level, where sadness is has low energy level [4].

$$E_n = \sum_{n=1}^{N} x(n).x^*(n)$$

2. Pitch:

The pitch signal, also known as the glottal waveform is one of the important features and has information about emotion, because it depends on the tension of the vocal folds and the sub glottal air pressure. The pitch signal is produced from the vibration of the vocal folds. Two features related to the pitch signal are widely used, namely the pitch frequency and the glottal air velocity at the vocal fold opening time instant. The time elapsed between two successive vocal fold openings is called pitch period T, while the vibration rate of the vocal folds is the

fundamental frequency of the phonation F0 or pitch frequency. Many algorithms for estimating the pitch signal exist [6].

3. Linear Prediction Cepstrum Coefficients (LPCC):

LPCC embodies the characteristics of particular channel of speech. Person with different emotional speech will have different channel characteristics, so we can extract these feature coefficients to identify the emotions contained in speech. The computational method of LPCC is usually a recurrence of computing the linear prediction coefficients (LPC), which is according to the all-pole model.

4. Mel-Frequency Cepstrum Coefficients (MFCC):

Mel frequency scale is the most widely used feature of the speech, with a simple calculation, good ability of the distinction, anti-noise and other advantages [2]. MFCC is based on the characteristics of the human ear's hearing, which uses a nonlinear frequency unit to simulate the human auditory system. The Fourier transform representation of the log magnitude spectrum called as the cepstrum coefficients. This high frequency coefficient with high efficiency are most robust and more reliable and useful set of feature for speech emotion recognition .Therefore the equation below shown using Fourier transform defined cepstrum of the signal y(n).

$$CC\,(n) = FT^{-1}\,\{\log|\,FT\,\{y(n)\}|\}$$

The Mel frequency is

$$F_{mel} = 3233\log_{10}(1+\frac{Fhz}{1000})$$

While calculating MFCC firstly pre-emphasize of speech signal from constructed emotional database has been done after this windowing is performed over pre-emphasize signal to make frames of 20 sec then the Fourier transform is calculated to obtain spectrum of speech signal and this spectrum is filtered by a filter bank in the Mel domain. After that the logs of the powers at each of the Mel frequencies is calculated. Then cosines transform in order to simplify the computation and are used to obtain the Mel frequency cepstrum coefficients.

Types of Classifier

For choice of classifier there is no fixed criterion. Selection of classifier depends on the geometry of the input feature vector. Some classifiers are more efficient with certain type of

class distributions, and some are better at dealing with many irrelevant features or with structured feature sets. Various Classifiers used by researchers are as follows:

1. HMM has been studied long time by researchers for speech emotion recognition, as has advantage on dynamic time warping capability. Moreover, it has been proved useful in dealing with the statistical and sequential aspects of the speech signal for emotion recognition. However, the classify property of HMM is not satisfactory [3].

2. Gaussian mixture model allows training the desired data set from the databases. GMM are known to capture distribution of data point from the input feature space, therefore GMM are suitable for developing emotion recognition model when large number of feature vector is available. Given a set of inputs, GMM refines the weights of each distribution through expectation-maximization algorithm. GMMs are suitable for developing emotion recognition models using spectral features, as the decision regarding the emotion category of the feature vector is taken based on its probability of coming from the feature vectors of the specific model. Gaussian Mixture Models (GMMs) are among the most statistically matured methods for clustering and for density estimation [6].

3. Another common classifier, used for many pattern recognition applications is the artificial neural network (ANN). They are known to be more effective in modelling nonlinear mappings. Also, their classification performance is usually better than HMM and GMM when the number of training examples is relatively low. Almost all ANNs can be categorized into three main basic types: MLP, recurrent neural networks (RNN) and radial basis functions (RBF) network. The classification accuracy of ANN is fairly low compared to other classifiers. The ANN based classifiers may achieve a correct classification rate of 51.19% in speaker dependent recognition, and that of 52.87% for speaker independent recognition [7].

4. One of the important classifiers is the support vector machine. SVM classifiers are mainly based on the use of kernel functions to nonlinearly map the original features to a high dimensional space where data can be well classified using a linear classifier. SVM classifiers are widely used in many pattern recognition applications and shown to outperform other well-known classifiers. SVM has shown to have better generalization performance than traditional techniques in solving classification problems. The accuracy of the SVM for the speaker independent and dependent classification are 75% and above 80% respectively [9]

CONCLUSION

In this paper, most recent work done in field of emotion recognition is studied. Most used methods of features extraction and classifier are reviewed. Performance of emotion recognition model depends highly upon combination of features such as prosodic or spectral and classifier used to classify emotions. To increase efficiency appropriate database of emotional speech sample must be used. The application area of emotion recognition from speech is expanding as it is new means of communication between machine and human.

REFERENCES

1. Akshay S. Utane, Dr. S. L. Nalbalwar, "Emotion Recognition Through Speech Using Gaussian Mixture Model And Hidden Markov Model", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, April 2013.

2. Dipti D. Joshi, Prof. M. B. Zalte, "Speech Emotion Recognition: A Review", IOSR Journal of Electronics and Communication Engineering (IOSR-JECE), Volume 4, Issue 4 (Jan. - Feb. 2013), PP 34-37.

3. A. Nogueiras, A. Moreno, A. Bonafonte, Jose B. Marino, "Speech Emotion Recognition Using Hidden Markov Model", Eurospeech, 2001.

4. M. N. Hasrul, M.Hariharan, Sazali Yaacob, "Human Affective (Emotion) Behavior Analysis using Speech Signals: A Review", International Conference on Biomedical Engineering (ICoBE), 27-28 February 2012.

5. Shahidhar G. Koolagudi, K. Sreenivasa Rao, "Recognition of Emotions from Speech using Excitation Source Features", International Conference on Modelling, Optimization and Computing (ICMOC), 2012.

6. Neiberg, D., elenius K., Laskowski K. "Emotion recognition in spontaneous speech using GMM", Proc. INTERSPEECH, 2006, Pittsburgh.

7. M. E. Ayadi, M. S. Kamel, F. Karray, "Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases", Pattern Recognition 44, PP.572-587, 2011.

8. Pathak, S., Kulkarni, A., "Recognizing emotions from speech", 3rd International Conference on Electronics Computer Technology (ICECT) (Vol. 4), Pp.107 – 109, 8-10 April 2011.

9. P. Shen, Z. Changjun, X. Chen, "Automatic Speech Emotion Recognition Using Support Vector Machine", International Conference on Electronic and Mechanical Engineering and Information Technology, 2011.