ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

Survey on ECG Data Compression

Yasmin P. Shaikh[1], Sunita Deshmukh[2]

Department E and TC, NBN School of Engineering, Sinhgad, Pune, India[1, 2].

ABSTRACT: Emotions play a very important role in our day to day life. Emotions are the natural physiological response of the human body which can be recognized by the facial expression. In the proposed system research has been done in the field of Human Computer Interaction(HCI). The entire project is divided into three major steps i.e. Face detection, facial feature extraction and classification. In the first phase face detection has been done using Haar Cascaded frontal face algorithm. The system detects and crops the lip region for further classifications, then the features are extracted into vectorized form. Extracted features are compared with trained database using Logistic Regression. The main hardware used in this project is Raspberry pi with linux based OS to neglect the drawbacks of personal computer and for better picture quality 8MP USB Camera is used.

KEYWORDS: Facial expressions, Haar cascade, Logistic regression, Human computer Interaction(HCI).

I. INTRODUCTION

Although emotion detection from speech is a relatively new field of research, it has many potential applications. In human-computer or human-human interaction systems, emotion recognition systems could provide users with improved services by being adaptive to their emotions. In virtual worlds, emotion recognition could help simulate more realistic avatar interaction. The body of work on detecting emotion in speech is quite limited. Currently, researchers are still debating what features influence the recognition of emotion in speech. There is also considerable uncertainty as to the best algorithm for classifying emotion, and which emotions to class together.
A speech signal is a logical arrangement of sounds. Our brain performs a complex set of analyses of auditory input (i.e. sounds). It converts the sounds into some conceptual ideas and thoughts which forms the basis of instructions, commands, information & entertainment. Automatic recognition is often studied in sense of identifying emotion among some fixed set of classes. Speech emotion recognition is a kind of analyzing vocal behavior. The speech processing involves three main steps i.e. pre-processing, feature extraction and pattern recognition. In case of speech signal, vowels carry the most of the informative part. Vowels are mainly voiced part of the spoken words. Therefore it is customary to separate out voiced part from unvoiced part of the information spoken and proceed further with signal processing on only voiced part. For an effective and natural HMI, emotion recognition plays a vital role. Emotions reflect the mental state of the person through speech, facial expressions, body postures and gestures and also other physical parameters like body temperature, blood pressure, muscle action, etc. The mental state of the person indirectly affects the speech produced by the person. E.g. in human-human interaction, speech rate is faster in case of anger/ joy and pitch range is also wider while in case of sadness, speech is slower with lower pitch range. Therefore, emotion detection in speech is advantageous in various applications.

II. LITERATURE SURVEY

In this paper we have designed a system which is used to recognize human emotions from audio clip generated by speaker. In this system we have employed two statistical models such as SVM and HMM to classify emotions .In order to recognize emotions we extracted four acoustic features such as spectral centroid, spread, flatness and projection. This system is divided in to five different stages-audio preprocessing, feature extraction, segmentation, model training and classification. Audio preprocessing is used to remove noise present in the signal. In the feature extraction part, we extracted four acoustic features. Segmentation is used to divide audio clips in to voiced and unvoiced category. [1]
Emotions play a very important role in our day to day life. Emotions are the natural physiological response of the

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

human body which can be recognized by the facial expression. In the proposed system research has been done in the field of Human Computer Interaction(HCI). The entire project is divided into three major steps i.e. Face detection, facial feature extraction and classification. In the first phase face detection has been done using Haar Cascaded frontal face algorithm. The system detects and crops the lip region for further classifications, then the features are extracted into vectorized form. Extracted features are compared with trained database using Logistic Regression. the main hardware used in this project is Raspberry pi with linux based OS to neglect the drawbacks of personal computer and for better picture quality 8MP USB Camera is used. [2]

Emotion recognition is a rapidly growing research domain in recent years. Unlike humans, machines lack the abilities to perceive and show emotions. But human-computer interaction can be improved by automated emotions recognition, thereby reducing the need of human intervention. In this paper, four basic emotions (Anger, Happy, Fear and Neutral) are analyzed from emotional speech signals. Signal processing methods are used for obtaining the production features from these signals. Source feature the instantaneous fundamental frequency (F0), system features the formants and dominant frequencies, zero-crossing rate (ZCR), and the combined features signal energy are used for the analyses. F0 is obtained using zero-frequency filtering (ZFF), and formants and dominant frequencies using LP spectrum. Short-time signal energy (STE) and ZCR are obtained in the voiced and unvoiced regions using a rectangular window of 200 samples. Two databases, German and Telugu Emotion Databases are used to cross-validate the results. Distinct differences are observed between high-arousal emotions (Anger and Happy) and Neutral emotion. Results indicate overlap between Anger and Happy emotions. But distinct differences are observed in the features for Happy/Anger and Fear, and between Happy and Anger emotions which is otherwise a challenging problem. The insights gained may be helpful in range of applications. [3]

Speech is a spontaneous medium of perceiving emotions which provide in-depth information related to different cognitive states of a human being. In this context, we introduce a novel approach using a combination of prosody features (i.e. pitch, energy, Zero crossing rate), quality features (i.e. Formant Frequencies, Spectral features etc.), derived features ((i.e.) Mel-Frequency Cepstral Coefficient (MFCC), Linear Predictive Coding Coefficients (LPCC)) and dynamic feature (Mel-Energy spectrum dynamic Coefficients (MEDC)) for robust automatic recognition of speaker's emotional states. Multilevel SVM classifier is used for identification of seven discrete emotional states namely angry, disgust, fear, happy, neutral, and sad and surprise [4]. In this project emotion from Hindi speech is developed. The database used was collected from various speakers belonging to different genders and age group. This work basically focused on eight emotions which comprises of a combination of fundamental emotions with some advance emotions and are listed as: Happy, Angry, Sad, Depressed, Bored, Anxiety, Fear and Nervous. These signals were preprocessed and analyzed using various techniques like: cepstral, linear prediction coefficient etc. In feature extraction various parameters used to form a feature vector are: fundamental frequency, pitch contour, formants, duration (pause length ratio) etc. These features are classified by using K Nearest Neighbor (KNN) classifier and Neural Network based classifiers. [5]

This paper presents guidelines to address the technical challenges in vocal emotion recognition in human machine interfaces which includes audio pre-processing, extraction of emotion relevant features and classification of it. Emotion recognition is the most challengeable and interesting topic of research which is so far is dealt with offline evolution. Paper demonstrates the different issues related to online processing, database preparation, features dominancy according to emotion and psychological changes during the emotion production. The overall objective of this paper is to help the reader to access the feasibility of human computer Interaction [6]

The pitch contour is one of the most significant properties of speech, which is affected by the emotional state. Therefore pitch features have been commonly used in systems for automatic emotion detection. In this work different intensities of emotions and their influence on pitch features have been studied. This understanding is important to develop such a system. Intensities of emotions are presented on Plutchik's cone-shaped 3D model. The k Nearest Neighbor algorithm has been used for classification. [7]

This paper presents a novel method for movie affective scene classification that outputs the emotion (in the form of labels) that the scene is likely to arouse in viewers. Since the affective preferences of users play an important role in movie selection, affective scene classification has the potential to develop more attractive user-centric movie search and browsing applications. Two main issues in designing movie affective scene classification are considered. One is "how to extract features that are strongly related to the viewer's emotions", and the other is "how to map the extracted

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

features to the emotion categories" [8]
In this paper, we propose an emotion detection system with facial features using a Bayesian network. In actual communication, it is possible that some parts of the face will be occluded by adornments such as glasses or a hat. In previous studies on facial recognition, these studies have been had the process to fill in the gaps of occluded features after capturing facial features from each image. However, not all occluded features can always be filled in the gaps accurately. Therefore, it is difficult for robots to detect emotions accurately in real-time communication. For this reason, we propose an emotion detection system taking into consideration partial occlusion of the face using causal relations between facial features. Bayesian network classifiers infer from the dependencies among the target attribute and explanatory variables. This characteristic of Bayesian network makes our proposed system can detect emotions without filling in the gaps of occluded features. In the experiments, the proposed system succeeded in detecting emotions with high recognition rates even though some facial features were occluded. In this paper, we propose an emotion detection system with facial features using a Bayesian network. [9]
This paper reports on methods for automatic classification of spoken utterances based on the emotional state of the speaker. The data set used for the analysis comes from a corpus of human-machine dialogs recorded from a commercial application deployed by Speechworks. Linear discriminant classification with Gaussian class-.conditional probability distribution and knearestneighbourhood methods are used to classify utterances into two basic emotion states, negativeand non-negative. The features used by the classifiers are utterance-level statistics of the fundamental frequency and energy of the speech signal. To improve classification performance, two specific feature s lection methods are used; namely, promising first selection and fomard feature selection. Principal component analysis is used to reduce the dimensionality of the features while maximizing classification accuracy [10]

IIII. PROPOSED SYSTEM

The pitch contour is one of the most significant properties of speech, which is affected by the emotional state. Therefore pitch features have been commonly used in systems for automatic emotion detection. Fig shows algorithm for audio signal processing used in this system.

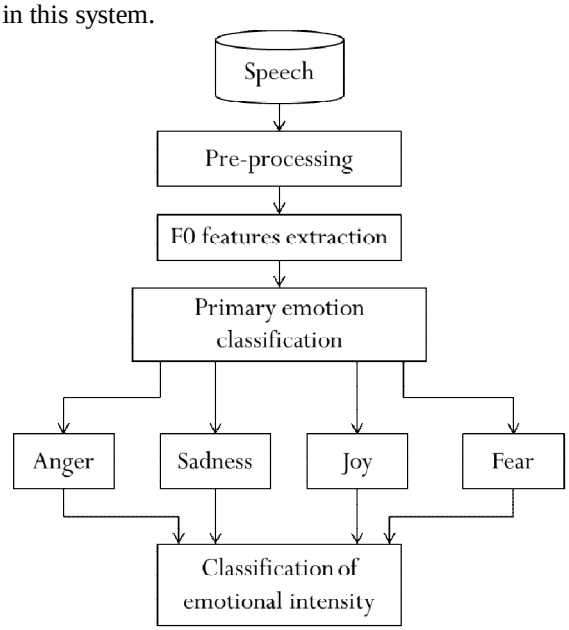

Fig1. Algorithm for audio signal processing

Speech: Speech is an audio signal. It is input from user

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

Pre-processing:

Pre-processing images commonly involves removing low-frequency background noise, normalizing the intensity of the individual particles images, removing reflections, and masking portions of images. Image pre-processing is the technique of enhancing data images prior to computational processing. Pre-processing of audio signal is divided in two parts:

1) The first stage being the pre-edit and processing of raw audio, to a common standard, before applying FX processing. This typically involves the removal of unwanted sections, such as chatter between takes, coughs, sneezes and any aberrant peaks, such as clicks, thumps, paper rustling, to leave a clean audio file, and then measuring the RMS level and normalising the audio to a predetermined RMS level so that the audio files are at the same RMS level prior to any FX processing.
2) The second stage is the use of FX to remove unwanted noise, rumble and hum and tonal and overall levelling, using Noise reduction, EQ, harmonic enhancement, dynamics etc to provide a set of clean, leveled audio assets.

All collected utterances may contain background and microphone noise. Wavelet thresholding was used to the de-noising recorded utterances. Moreover, for further analysis, all collected utterances have been segmented into 20 ms frames using Hamming window with 50% overlap.

F0 features extraction

Feature extraction is the estimation of variables, called a feature vector, from another set of variables (e.g., an observed speech signal time series). Feature selection is the transformation of these observation vectors to feature vectors. The goal of feature selection is to find a transformation to a relatively low-dimensional feature space that preserves the information pertinent to the application while enabling meaningful comparisons to be performed using simple measures of similarity.

Selection of efficient acoustic features is a critical point. It is quite difficult to create a non-numerous vector, which describes the object of analysis well. In this paper the influence of demonstrated emotional states on F0 contour has been presented. Following are typical F0 contours for four basic emotions and their intensities.

1) There are three anger intensities: rage, anger and annoyance. For rage F0 increases noticeably in relation to neutral speech and also to its intensities. This emotion appears to progress on a higher level in voice pitch. The lowest values were obtained for annoyance. Along with increase of emotion intensity the pitch range becomes much wider and its rises have a greater steepness.
2) According Plutchik's model joy has three intensities: ecstasy, joy and serenity. These vocal emotional states (similar with rage, anger and annoyance) characterized by increases in F0 mean, range and variability. However, pitch changes are smoother compared to the previous group. Although, increases are still proportional to the intensity of articulated emotion.
3) Grief, sadness and pensiveness have very similar F0 contours, also similar with the neutral speech. There is general decrease in F0 mean, range and variability and also downward directed intonation contour. All of them are spoken with a small amount of change, F0 is almost constant. As in previous cases increases are proportional to the intensity of emotion.
4) The last group of emotion consists of terror, fear and apprehension. During the examination higher F0 mean and wider F0 range were found in comparison with neutral speech contour.

The effect of the intensity for the fundamental frequency is the same as in other emotional groups.

Classification

Firstly, all emotions were assigned to four groups representing primary emotions: anger, fear, sadness, and joy. It is difficult to exactly recognize emotion basing only on F0 features even with such a small set of emotions. Best results were obtained, as well as in many other researchers, for anger.

Second step was classification of intensities inside each group. Both classifications were carried out using k-NN algorithm. For recognition of emotion intensities in a specific group accuracy performance greatly improves.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

IV. CONCLUSIONS

In this paper, a new approach for recognizing emotions from speech signal has been proposed. The results of this investigation show that expression of emotion affects F0 contour. However, usage of features related solely to F0 does not provide satisfactory results. The average recognition accuracy of emotion group recognition is about 50%. For recognition of emotion intensities in a specific group accuracy performance greatly improves. One can observe some regularity for each group of emotions: best results were achieved for the weakest and strongest intensities, the worst results for primary emotions. Moreover, analysis of confusion matrix shows that if the classification is incorrect, results point at adjacent emotion of the same group.

REFERENCES

[1] Recognition Of Emotions From Audio Signals Swapnali Tandel1, Shital Patil2, Vikrant Kadam3, Srijitabhattacharjee International Journal Of Current Engineering And Scientific Research (IJCESR) Volume-5, Issue-2, 2018

[2] Prof. V.D. Bharate1, Shubham Sunil Phadatare2, Suhas Panchbhai3, Vishal Pawar, "Emotion Detection using Raspberry Pi", International Research Journal of Engineering and Technology (IRJET) Volume: 04 Issue: 05 | May -2017 Page 780

[3] Esther Ramdinmawii1, Abhijit Mohanta2 and Vinay Kumar Mittal "Emotion Recognition from Speech Signal", Proc. of the 2017 IEEE Region 10 Conference (TENCON), Malaysia, November 5-8, 2017

[4] Emotion Recognition in Speech by MFCC and SVM AkhileshWatile, Vilas Alagdeve, Swapnil Jain International Journal of Science, Engineering and Technology Research (IJSETR) Volume 6, Issue 3, March 2017, ISSN: 2278 -7798

[5] Rajni, Dr.Nripendra Narayan DasEmotion Recognition from Audio Signal International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5, Issue 6, June 2016

[6] Development in Emotion Recognition in Human Speech: A ReviewProf. SandeepKhannaMr.Vipin J. Gawande, International Journal of Engineering Research & Technology (IJERT), Vol. 3 Issue 4, April - 2014

[7] Recognition of Human Emotion from a Speech Signal Based on Plutchik's Model INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2012, VOL. 58, NO. 2, PP. 165–170 August 24, 2011; revised May, 2012

[8] Miyakoshi, Y., & Kato, S. (2011). Facial emotion detection considering partial occlusion of face using Bayesian network. 2011 IEEE Symposium on Computers & Informatics.

[9] Irie, G., Satou, T., Kojima, A., Yamasaki, T., &Aizawa, K. (2010). Affective Audio-Visual Words and Latent Topic Driving Model for Realizing Movie Affective Scene Classification. IEEE Transactions on Multimedia, 12(6), 523–535.

[10] C. M. Lee', S. Narayanan'. R. Pieraccini, "Recognition of Negative Emotions from the Speech Signal", 0-7803-7343-X/02/$17.00 Q 2002 IEEE