© 2018 IJSRST | Volume 4 | Issue 2 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

Review on Detection and Analysis of Emotion from Speech Signals

Yuvraj M. Umak[1], Dr. Pritesh R. Gumble[2]

[1]M.E Scholar, Department of Electronics & Telecommunication Engineering, Sipna College of Engineering and Technology, Amravati, Maharashtra, India

[2]Associate Professor, Department of Electronics & Telecommunication Engineering, Sipna College of Engineering and Technology, Amravati, Maharashtra, India

ABSTRACT

Perceiving feeling from discourse has turned out to be one the dynamic research topics in discourse handling and in applications in view of human-PC cooperation. The feelings considered for the tests incorporate happy, sad, fear, anger, boredom and neutral. The recognize capacity of passionate highlights in discourse were examined first took after by feeling characterization performed on a custom dataset. The arrangement was performed for various classifiers. One of the primary component quality considered in the arranged dataset was the crest to-top separation got from the graphical portrayal of the discourse signals. Feeling is characterized as the constructive or adverse condition of a man's mind which is connected with an example of physiological exercises. Feelings portray the psychological condition of a man. Using MFCC based parameters show the energy migration in frequency domain and also helps in identifying phonetic characteristics of speech. Feature extraction process done by using MFCC.

Keywords: FFT/DCT, MFCC, Emotion Analysis, Emotion Classification, Speech Processing, Mel-Frequency Cepstral Coefficients, Human-Computer Interface.

I. INTRODUCTION

Feeling grouping is a standout amongst the most difficult assignments in a discourse flag handling space. The issue of speaker or discourse acknowledgment turns out to be generally a less demanding one when contrasted and perceiving feeling from discourse. Sound flag is one of the fundamental medium of correspondence and it can be handled to perceive the speaker, discourse or even feeling. The essential guideline behind feeling acknowledgment lies with breaking down the acoustic contrast that happens while expressing a similar thing under various enthusiastic circumstances. Notwithstanding the highlights relating to the speaker or potentially the discourse, the sound signs do have a few highlights that speak to the passionate condition of the speaker. The investigation is gone for investigating conditions the idea of articulation have with the human enthusiastic state. Since the feelings impact the sensory system, the heart rate likewise is influenced by them. The work in says that if there is a negative boosts that causes negative feeling the heart rate decelerate more effectively than when there is certain jolts . So the feeling acknowledgment can likewise prompt recognize the focused on circumstance of a man. Now and then in numerous applications, for example, military and regular citizen applications, in police office, it's important to get to whether a speaker is talking bona fide or not.

MFCCs are coefficients which speak to sound, in light of view of human sound-related frameworks. The fundamental distinction between the activity of

FFT/DCT and the MFCC is that in the MFCC, the recurrence groups are situated logarithmically which approximates the human sound-related framework's reaction more intently than the straightly separated recurrence groups of FFT or DCT.

The principle point to comprehend about discourse is that the sounds produced by a human are separated by the state of the vocal tract including tongue, teeth and so on. This shape figures out what sound turns out. In the event that we can decide the shape precisely, this should give us an exact portrayal of the phoneme being created. The state of the vocal tract shows itself in the envelope of the brief timeframe control range, and the activity of MFCCs is to precisely speak to this envelope.

II. LITERATURE REVIEW

Darren M. Haddada, Roy Ratleyb "Voice Stress Analysis and Evaluation" Enabling Technologies for Law Enforcement and Security, Simon K. Bramble, Edward M. Carapezza, Leonid I. Rudin, Editors, Proceedings of SPIE Vol. 4232 (2001) © 2001 SPIE · 0277-786X/01

After reviewing the three technical tests performed, it could be stated that these two VSA units do recognize stress. Although these systems state they detect deception, this was not proven. This study does shows, from a number of speech under stress studies, that linear and non-linear features are useful for stress classification. Due to the lack of deceptive stress data available, classification of deceptive stress versus emotional stress or physical stress was not been tested. This is a vital role in the detection and classification of stress. Many suspects are under an extreme amount of stress under being interrogated. Do these VSA systems actually differentiate between the different types of stress? This still needs to be proven. It was proven that the Diogenes Lantern system detects stress via the amount of energy in the speech envelope. Even though this system performed well under the technical and the field tests, it seems from an engineering point of view, that one feature is insufficient to detect and classify deceptive stress. In the study under Dr. Hansen, it was shown that fusion of features help to increase the accuracy of your stress classification. It was proven that the systems tested will and do give the same response when the inputted audio is recorded as opposed to live. The only criterion is when recording using a cassette player it is the up most importance to set the AGC. To eliminate any possibility of errors, recording with a DAT is the safest way to go.

Assel Davletcharovaa, Sherin Sugathanb, Bibia Abrahamc, Alex Pappachen James "Detection and Analysis of Emotion From Speech Signals" Second International Symposium on Computer Vision and the Internet (VisionNet'15) © 2015

Recognizing emotion from speech has become one the active research themes in speech processing and in applications based on human-computer interaction. This paper conducts an experimental study on recognizing emotions from human speech. The emotions considered for the experiments include neutral, anger, joy and sadness. The distinuishability of emotional features in speech were studied first followed by emotion classification performed on a custom dataset. The classification was performed for different classifiers. One of the main feature attribute considered in the prepared dataset was the peak-to-peak distance obtained from the graphical representation of the speech signals. After performing the classification tests on a dataset formed from 30 different subjects, it was found that for getting better accuracy, one should consider the data collected from one person rather than considering the data from a group of people.

The paper explored the idea of detecting the emotional state of a person by speech processing techniques. The study on words and letters under

different emotional situations proved that the emotional state can alter the speech signal. It was observed that there are distinguishable features in a speech segment that characterizes each emotion state. After performing the classification tests on a dataset prepared from 30 subjects, it was observed that it is better considering data from an individual subject rather than a group of people. The development of a software based agent for emotion detection and heart rate analysis can greatly improve telemedicine based systems.

Neha P. Dhole, Dr. Ajay A.Gurjar "Detection Of Speech Under Stress Using Spectral Analysis" Volume: 02 Issue: 04 | Apr-2013

This paper deals with an approach to the detection of speech in English language. The stress detection is necessary which provides real time information of state of mind of a person. Voice features from the speech signal is influenced by stress is MFCC is considered is this paper. To examine the effect of Exam Stress on speech production an experiment was designed. First Year students of age group 18 to 20 were selected and assignment was given to them and instructs them that have viva on that assignment and their performance in the viva will decide their final internal marks in the examination. The experiment and the analysis of the test results are reported in this paper.

The Spectral Analysis of speech signal is aimed at extracting spectral features such as MFCC Changes in spectrum of speech signal have shown to be a indicator of the internal emotional state of a person. In this research work, we have extracted these spectral features of some speakers in neutral condition and under stress condition. We have formed the feature matrix of the feature vectors obtained. For classification of the speech signal for stress Artificial Neural Network and ANFIS plays main role.. Thus, we could conclude that spectral analysis is an efficient tool for detecting stress in speech.

Jarosław Cichosz, Krzysztof Ślot, "Emotion recognition in speech signal using emotion extracting binary decision trees", supported by the Polish State Fund for Research Support,pp:211-215, may 2012.
The presented paper is concerned with emotion recognition based on speech signal. Two novel elements introduced in the method are an introduction of novel set of emotional speech descriptors and an application of a binary-tree based classifier, where consecutive emotions are extracted at each node, based on an assessment of feature triplets. The method has been verified using two databases of emotional speech on German and Polish, yielding very high recognition rates (72 %) for speaker-independent recognition.

The presented method of emotional speech classification through dichotomy-based decision-trees has been proved to produce very promising results. The experiments performed on two different databases, which comprised recordings of emotional speech, uttered in two different languages, yielded high recognition rates. This suggests that the proposed emotion-extraction approach can appear appropriate for the task realization. We have also shown that the proposed, new emotional speech signal descriptors are very useful in emotion modeling. Among these descriptors, regression parameters of pitch and mean energy in low-frequency sub-bands consistently appeared to be among the best performing features. We believe that linear or nonlinear regression parameters are inherently good ways of emotion characterization, since they combine frequency and temporal description of the signal in a compact way.

Pooja Yadav, Gaurav Aggarwal "Speech Emotion Classification using Machine Learning" International Journal of Computer Applications (0975 – 8887) Volume 118 – No. 13, May 2015.

In this paper, the study of how emotions are recognized from speech using various machine learning algorithms is discussed. The recognition rate is calculated by applying various classification algorithms and the algorithms which provides the best recognition rate is identified. The emotion recognition rate is dependent on the types of features extracted and the selection of the classification algorithm. From the study, it has been evaluated that the SVM and the SMO algorithms are better classification algorithms which gives higher accuracy in emotion recognition rates. Our future work will include the recording of speech samples of small children to adult speech samples and then extracting the features like pitch and fundamental frequency from those recorded data set. After extraction of the feature vectors we will be applying the classification algorithms to recognize emotions from the recorded speech dataset. The classification algorithms will help in evaluating that how emotional state of a person changes from a child to adult. Our proposed work is based on using Machine learning technique to develop a speech emotion recognition system with more correctness and efficient than the already existing systems. For identification of voice of children and adults, we need to create a database for making it more robust.

Without database it will be difficult for the system to differentiate between emotions of child and adults. The proposed topic on which I have decided to work is "Characterization of emotion from specch using machine learning algorithm". We will be recording audio messages of small children from 4 to 8 years and also of some adult males and females. After that the audio samples are recorded, the samples are converted into monowave format. The next step will be extraction of features from the speech samples. After the features are extracted from the speech samples, we will be applying one of the above classification algorithms. This will help us in classifying how the emotion of human beings is affected as tha age increases.

J. Sirisha Devi, Dr. Srinivas Yarramalle, Siva Prasad Nandyala "Speaker Emotion Recognition Based on Speech Features and Classification Techniques" I.J. Computer Network and Information Security, 2014.

Speech Processing has been developed as one of the vital provision region of Digital Signal Processing. Speaker recognition is the methodology of immediately distinguishing who is talking dependent upon special aspects held in discourse waves. This strategy makes it conceivable to utilize the speaker's voice to check their character and control access to administrations, for example voice dialing, data administrations, voice send, and security control for secret information. A review on speaker recognition and emotion recognition is performed based on past ten years of research work. So far iari is done on text independent and dependent speaker recognition. There are many prosodic features of speech signal that depict the emotion of a speaker. A detailed study on these issues is presented in this paper. This paper attempts to provide a complete survey of research on speaker recognition and emotion recognition. Survey says that till date we could not achieve 100% accuracy in recognizing either a speaker or his emotion.

When the emotional state of speaker differs in the testing phase the recognition rate decreases significantly. The table shows that the accuracy rate of speaker recognition has been considerably affected when the emotional state of the speaker was not considered. Pitch is not particularly good for the recognition of neutral tones. Survey of this paper gives us a conclusion the accuracy rate of speaker indirectly depends on the accuracy of emotion recognition.

A. A. Khulageand Prof. B. V. Pathak. "Analysis Of Speech Under Stress Using Linear Techniques And Non-Linear Techniques For Emotion Recognition System" © CS & IT-CSCP 2012 DOI : 10.5121/csit.2012.2328.

Analysis of speech for recognition of stress is important for identification of emotional state of person. This can be done using 'Linear Techniques', which has different parameters like pitch, vocal tract spectrum, formant frequencies, Duration, MFCC etc. which are used for extraction of features from speech. TEO-CB-Auto-Env is the method which is non-linear method of features extraction. Analysis is done using TU-Berlin (Technical University of Berlin) German database. Here emotion recognition is done for different emotions like neutral, happy, disgust, sad, boredom and anger. Emotion recognition is used in lie detector, database access systems, and in military for recognition of soldiers' emotion identification during the war. Hierarchical algorithm is used for classification of different emotions. After differentiating anger and disgust emotions from TEO-CB-Auto method, formant frequencies estimation method is used to classify angry and disgust emotion recognition. First formant frequencies range is has higher values in angry emotion as compared to disgust emotion. Vocal tract spectrum estimation is used for identification of happy emotion. Sad emotion is separate out from boredom and neutral emotion using duration calculation. Boredom and neutral emotion is differentiated under the consideration of mean values of MFCC. Mean values of MFCC in boredom emotion are greater than mean values of MFCC obtained in case of neutral emotion.

III. PROPOSED WORK

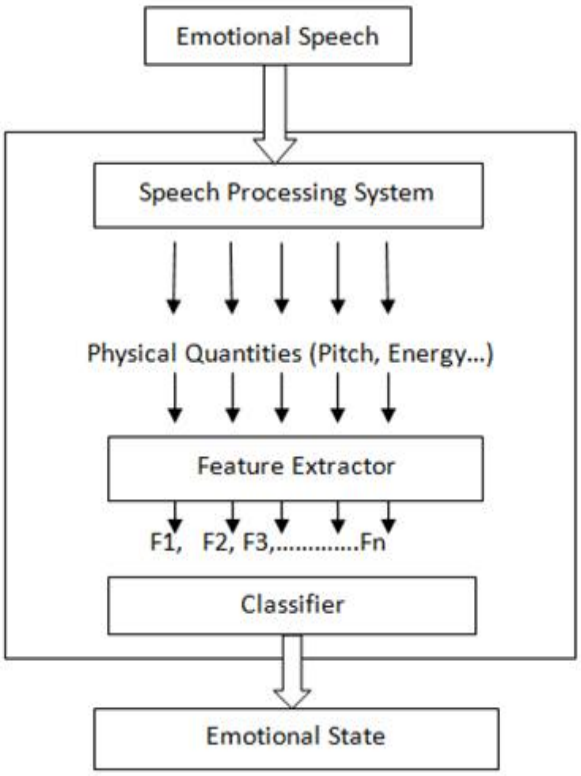

Figure 1. System square outline

To start with, the passionate discourse of certain feeling record utilizing an amplifier and changed over into .wave document arrange. At that point discourse handling framework is utilized to gained physical amount like Pitch, Energy, from which helpful element are separated. MFCC is most generally utilized as a part of sound order tests because of its great execution. It removes and speak to highlight of discourse flag. The Mel-cepstra takes brief time ghastly shape with essential information about the nature of voice and creation impacts. To ascertain these coefficients the cosine change of genuine logarithm of the transient range of vitality must be finished. At that point it is performed in mel-recurrence scale. Further, after pre-underlining the discourse sections are windowed. The Hamming window utilized for this procedure is a straightforward window in view of lessening of spillage impact. It smears vitality from genuine flag recurrence into neighboring ones along these lines

contrarily influencing the execution. It additionally adds to maintaining a strategic distance from the intermittence of the discourse motion in time area that may happen amid Fast Fourier Transform. The idea of windowing depends on duplicating the flag outlines by window work. At last identified and ordered feeling will be appeared in plain view screen.

Feature Extraction

Discourse is divided into little interims known as edges. The way toward dividing discourse into outlines in light of the data they are conveying about feeling is known as highlight extraction. Highlight extraction is the essential advance in SER (discourse feeling acknowledgment) framework. A portion of the highlights that makes sense of feelings from discourse are-

Pitch

It is the primary part of any discourse which is characterized as the lowness or height of a voice as distinguished by the human ears. Pitch is reliant on the vibrations every second. The estimation of pitch parameter is extricated by utilizing cepstrum in the recurrence area. Contribute helps recognizing the impartial and furious feelings from discourse test.

Energy

Force of the discourse characterizes the vitality level of discourse. Vitality level for each edge is computed as first the square of all example sufficiency is done and after that summing up the estimations of all the squared example amplitudes.

Pitch Difference and Energy Difference

The contrast between the estimations of pitch or vitality level of neighboring sections is utilized to sort the discourse parameters into feelings. The more the variance the more it is less demanding to uncover the energetic feelings like bliss and outrage

Formats

Formants are represented by the state of the vocal tract and are controlled by various feelings for eg. The condition of energy brings about acquiring the higher mean estimations of the primary formant recurrence. The essential recurrence (f0) helps in distinguishing cheerful feeling from discourse tests.

Mel-recurrence cepstral coefficient

The Mel-recurrence cepstral coefficients (MFCC) are broadly utilized as a part of sound characterization tries because of its great execution. It extricates and speaks to highlights of discourse flag. The Mel-cepstra takes brief time phantom shape with essential information about the nature of voice and generation impacts .To ascertain these coefficients the cosine change of genuine logarithm of the fleeting range of vitality must be finished. At that point it is performed in mel recurrence scale.

Further, after pre-accentuating the discourse fragments are windowed .The Hamming window utilized for this procedure is a straightforward window in light of diminishment of spillage impact. It smears vitality from genuine flag recurrence into neighboring ones in this way adversely influencing the execution. It additionally adds to maintaining a strategic distance from the irregularity of the discourse motion in time space that may happen amid Fast Fourier Transform. The idea of windowing depends on increasing the flag outlines by window work.

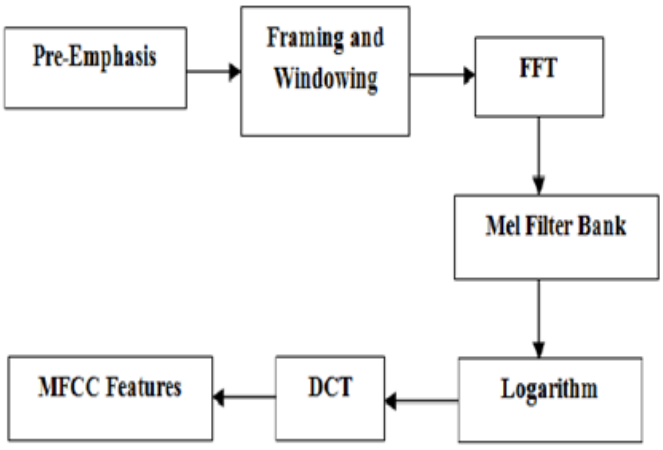

Figure 2. MFCC Generation

Pre-Emphasis

In handling sound signs, pre-accentuation alludes to a framework procedure intended to expand the size of some higher frequencies regarding the extent of other lower frequencies keeping in mind the end goal to enhance the general flag to-clamor proportion by limiting the antagonistic impacts of such marvels as lessening contortion or immersion of recording media in resulting parts of the framework

Encircling and windowing

Discourse signals are separated into casings of wanted length and are examined and Hamming window is connected to each casing to expel discontinuities in flag and guarantee coherence amongst first and last information focuses. Along these lines, Each casing must be duplicated with a hamming window to keep the coherence of the first and the last focuses in the casing. On the off chance that the flag in an edge is signified by x(n), n = 0,… N-1, at that point the flag subsequent to Hamming windowing is, x(n) * w(n) , where w(n) is the Hamming window characterized by

Quick fourier change

It changes over each casing from time area signals into recurrence space and gets recurrence reaction of each edge. and afterward performed Mel recurrence wrapping on this range in which range is sifted by a channel bank in the Mel area. at that point taking the logs of the forces at each of the Mel frequencies to get Mel range. At that point this Mel range is changed over by taking log to cepstrum and acquires the Mel recurrence cepstrum coefficients.

Logarithm

It takes the log of the power at each of the mel frequencies

Discrete cosine change

A discrete cosine change (DCT) communicates a limited arrangement of information focuses as far as an aggregate of cosine capacities wavering at various frequencies The preferred standpoint the DCT has over the Fourier change is that the subsequent coefficients are genuine esteemed, which makes ensuing preparing and capacity less demanding. DCT method expels certain frequencies from sound information to such an extent that the size is decreased with sensible quality.

IV. CONCLUSION

Feeling acknowledgment framework is a vital research zone in the present fields. There are the few applications where discourse feeling acknowledgment can be sent. An appropriately and all around composed database is fundamental for building up the feeling acknowledgment framework. This audit paper covers the current work of discourse feeling acknowledgment for filling some essential re-look holes. This paper contains the audit of late works in discourse feeling acknowledgment from the purposes of perspectives of enthusiastic databases, discourse highlights, and grouping models. The paper investigates distinguishing the enthusiastic condition of a man by discourse handling methods. The investigation on words and letters under various passionate circumstances demonstrated that the enthusiastic state can modify the discourse flag. The advancement of a product based operator for feeling location and heart rate examination can incredibly

enhance telemedicine based frameworks can likewise be progressed.

V. REFERENCES

[1]. Darren M. Haddada, Roy Ratleyb "Voice Stress Analysis and Evaluation" Enabling Technologies for Law Enforcement and Security, Simon K. Bramble, Edward M. Carapezza, Leonid I. Rudin, Editors, Proceedings of SPIE Vol. 4232 (2001) © 2001 SPIE · 0277-786X/01

[2]. Assel Davletcharovaa, Sherin Sugathanb, Bibia Abrahamc, Alex Pappachen James "Detection and Analysis of Emotion From Speech Signals" Second International Symposium on Computer Vision and the Internet (VisionNet'15) © 2015

[3]. Neha P. Dhole, Dr. Ajay A.Gurjar "Detection Of Speech Under Stress Using Spectral Analysis" Volume: 02 Issue: 04 | Apr-2013

[4]. Jarosław Cichosz, Krzysztof Ślot, "Emotion recognition in speech signal using emotion extracting binary decision trees", supported by the Polish State Fund for Research Support,pp:211-215, may 2012.

[5]. Pooja Yadav, Gaurav Aggarwal "Speech Emotion Classification using Machine Learning" International Journal of Computer Applications (0975 – 8887) Volume 118 – No. 13, May 2015.

[6]. J. Sirisha Devi, Dr. Srinivas Yarramalle, Siva Prasad Nandyala "Speaker Emotion Recognition Based on Speech Features and Classification Techniques" I.J. Computer Network and Information Security, 2014.

[7]. A. A. Khulageand Prof. B. V. Pathak. "Analysis Of Speech Under Stress Using Linear Techniques And Non-Linear Techniques For Emotion Recognition System" © CS & IT-CSCP 2012 DOI : 10.5121/csit.2012.2328.

[8]. Hassan, E. A., El Gayar, N., &Moustafa, M. G. "Emotions analysis of speech for call classification" In Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on (pp.242-247). IEEE (2010,November).

[9]. Nwe, T. L., Wei, F. S., & De Silva, L.C.(2001). Speech based emotion classification. In TENCON 2001 Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology(Vol. 1. pp. 297-301),IEEE.

[10]. Casale, S., Russo, A., Scebba, G., & Serrano, S. (2008, August). Speech emotion classification using machine learning algorithms. In Semantic Computing, 2008 IEEE International Conference on (pp. 158-165). IEEE.

[11]. "Emotion Recognition from Speech: A Survey" by Rani P. Gadhe,Shaikh Nilofer R. A., V. B.Waghmare, P. P. Shrishrimal, R. R. Deshmukh on April-2015.