ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

Classification of Emotion using Different No. of MFCCs

[1] Pramod Mehra, [2] Parag Jain
[1][2] UTU, Dehradun

Abstract - **In this paper, the spectral features are separated from speech and used to perceive the feelings from speech. As the speech has been utilized as a vital method of correspondence since the time is immemorial. Feelings are a basic piece of normal speech correspondence. The vast majority of the present speech frameworks can process studio recorded nonpartisan speech with more prominent precision. Hence, a need is felt to refresh speech preparing frameworks with the ability to process feelings. The part of feeling handling makes the current speech frameworks more practical and significant. In this work, spectral highlights are extricated from speech to perform feeling grouping. mel recurrence cepstral coefficients and their subsidiaries (speed and increasing speed coefficients) are investigated as highlights. Gaussian blend models are proposed as classifiers. The feelings considered in this examination are outrage, satisfaction, unbiased, pity and astonishment. The speech feeling database utilized as a part of this work is semi-common in nature, which has been gathered from the exchanges of performing artists/on-screen characters.**

Index Terms — Feeling grouping, GMM, MFCC, Spectral highlights, Text dependent emotion recognition, Text independent emotion recognition.

I. INTRODUCTION

Speech is the vocalized type of human correspondence. Each talked word is made out of the phonetic blend of a restricted arrangement of vowel and consonant speech sound units. These vocabularies, the linguistic structure which structures them, and their arrangement of speech sound units vary, making the presence of a huge number of various sorts of commonly incoherent human dialect Emotions. Most human speakers (polyglots) can convey in at least two of them. The vocal capacities that empower people to deliver speech likewise furnish people with the capacity to sing. Feeling acknowledgment from speech articulations might be valuable in various applications, for example, call focus discussion investigation, stimulation, ordering of sound documents in light of feelings, improvement of compelling human PC collaboration et cetera. Speech highlights might be fundamentally extricated from excitation source, vocal tract or prosodic perspectives, to finish distinctive speech tasks. This work shows its extension to the utilization of spectral highlights for perceiving feelings. Spectral highlights speak to vocal tract data, for example, formant frequencies successive variety in the shapes/sizes of vocal tracts, spectral bandwidths, spectral roll-off et cetera. By and large, the vast majority of the speech undertakings are refined effectively utilizing spectral highlights. Different spectral highlights have likewise been investigated for feeling examination. To recognize outrage from nonpartisan speech in Mandarin dialect, a mix of MFCCs, LPCCs, Rasta PLP coefficients and log frequency power coefficients (LFPCs) has been utilized as the highlights [3]. MFCC highlights from bring down recurrence parts (20 Hz to 300 Hz) of speech signal have been utilized to display pitch varieties.

Naturally, spectral highlights are separated through piece handling approach. Entire speech signal is processed frame by frame, considering the frame size of around 20 ms, and a shift of 10 ms. It is accepted that with in this frame, speech signal is stationary in nature. Mel recurrence cepstral coefficients are removed as spectral highlights and utilized as a part of this work for feeling investigation. In this work, UTU-semi-natural database (UTU-SNESC), gathered from Hindi motion pictures has been utilized.

II. DATABASE

From the accessible writing, three sorts of databases are utilized for investigation of speech feelings. They are simulated, elicited and natural and semi-natural speech databases. Enthusiastic speech separated from exchanges of performing artists and on-screen characters of Hindi films has been utilized to make semi common feeling speech corpus known as Uttarakhand Technical University Semi Natural Emotion Speech Corpus (UTU-SNESC). The feelings communicated by on-screen characters in Hindi motion pictures are near genuine feeling articulation saw on account of typical Hindi speaking Indian populace.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

For the formation of the database, exchanges of the famous performers have been extricated from Hindi films. This database contains single and multi-speaker enthusiastic expressions for male speakers. The feelings gathered for this database are outrage, joy, unbiased, trouble and shock. For single speaker database, video clasps of various Hindi motion pictures acted by a similar on-screen character are utilized. Afterward, sound tracks are isolated and connected to make a solitary document. Adobe Audition is utilized to separate sound with mono channel recurrence of 16 KHz and 16 bit determination. Enthusiastic speech has been separated deliberately, containing no ambient sounds and unsettling influences. Long quiet districts have been expelled with the assistance of wavesurfer without influencing the installed feelings. Fifteen minutes of successful information is gathered along these lines for every feeling for male speakers. For multi-speaker database, the video clasps of various Hindi motion pictures are picked independent of performers, yet of a similar sex and feelings.

III. EXTRACTION OF FEATURES

Legitimate component extraction dispenses with unimportant highlights that upset the acknowledgment rates; it decreases the info dimensionality (and thusly enhances speculation); it sets aside the computational assets. Highlight vectors can be long-lasting or brief time in nature. Long-lasting highlights are assessed over the whole expression length. Brief time highlights are resolved in a littler time window (normally 20 to 30 msec). Con-transitory research approach supports the long-lasting highlights for investigation of feelings [4], since the long time highlights correspond feelings superior to brief time ones. The extraction of individual highlights and their utilization in building up the models has been talked about in the accompanying sections. Mel frequency cepstral coefficients (MFCCs), their subordinates, known as speed (Δ) and quickening (Δ - Δ) coefficients are sepa-rately utilized for feeling examination. Different MFCCs are removed from speech signal. The previously mentioned number of highlights utilized independently to develop feeling acknowledgment models. Δ and Δ-Δ coefficients are utilized as a part of connection with individual essential highlights to frame the component vectors. Hamming window has been utilized while confining the speech signal. The general square graph of advancement of feeling acknowledgment models (FAMs) is given in Fig. 1.

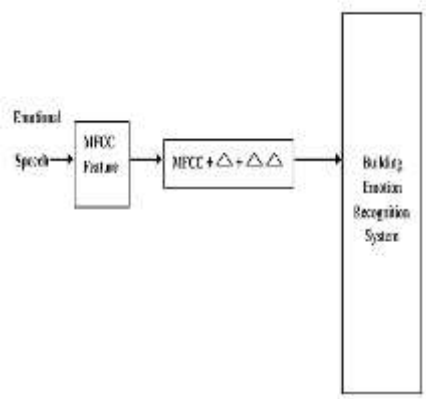

Fig. 1. Plan of creating feeling acknowledgment models

Human sound-related framework is expected to process speech motion in a nonlinear manner. It is all around contemplated that lower recurrence parts of speech signal contain more infor-mation. Along these lines nonlinear mel scale channel has been intended to accentuate bring down recurrence parts over higher ones. In speech preparing mel recurrence cepstrum is a portrayal of the brief span control range of a speech outline utilizing direct cosine change of log control range on a non-straight mel recurrence scale. Transformation from ordinary recurrence 'f' to mel recurrence 'm' is given by the condition :

$$m = 2595 \log10(f/700+ 1) \quad (1)$$

The calculation utilized as a part of this work for acquiring mel recurrence cepstral coefficients (MFCCs) from speech signal is as per the following :

1. Acquire Fourier change of a speech fragment to get brief time range.
2. Register forces of the above range inside the triangular covering win-dows set by mel scale.
3. Take logs of the power at each of the mel frequencies.
4. . Register the discrete cosine transform (DCT) of the rundown of mel log controls as though it were a signal. (Note: Basically one of the imperative uses of DCT is to dispose of unimportant number of high recurrence parts amid pressure.)
5. The amplitudes of coming about range give MFCC's.

IV. EMOTION RECOGNITION MODELS (ERMS)

The feeling acknowledgment models are produced utilizing male speaker expressions. 90% of the information is utilized for preparing the feeling

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

acknowledgment models and 10% is utilized for approval. Figure 2 represents the general strategy of a feeling acknowledgment framework.

The procedure is isolated into two sections: include extraction stage and feeling recog-nition stage. From every speech articulation highlights (MFCCs) are separated. Subse-quently, highlight vectors are framed. These component vectors are given as contribution to the feeling acknowledgment advancement stage. In the preparation organize, the component vectors are utilized to prepare the GMM models. In the acknowledgment organize, the element vectors of test articulations are given to officially prepared models. Approval is finished by giving test articulate to officially prepared models. Feeling Recognition Models are created utilizing MFCCs and their Δ and Δ - Δ highlights acquired from the speech signal.

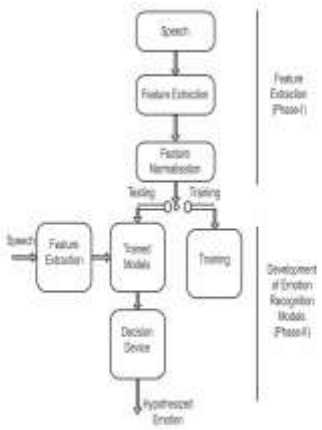

Fig. 2. Improvement of a feeling acknowledgment framework

V. RESULTS AND DICUSSION

While utilizing UTU-SNESC, the content of the exchanges recorded from various motion pictures was not same. Subsequently, the enthusiastic expressions utilized for preparing and testing the GMM models contain diverse content. This is the reason that the acknowledgment execution is cited as content autonomous. For this situation, it might be noticed that the impact of phonetic data on feeling acknowledgment is slightest.

Table I demonstrates the consequences of feeling acknowledgment of UTU – SNESC database utilizing 6, 8, 13, 21, 29 MFCCs as perplexity network

The average emotion recognition is found 49.2%, 51.2%, 46.8%, 46.6%, 55.6% respectively. Highest emotion recognition performance of 55.6% is achieved with 29 MFCC's.

The emotion recognition performance for neutral emotions is worst as in real life, conversations among humans are never neutral (unbiased) but are always interlaced with some emotion. Also Astonishment is mostly expressed with cheerful and outrage.

The other possible reasons for low recognition rate are use of less data for training the model, presence of silence zones in speech, presence of background music.

Table 1. Average emotion classification performance with semi-natural database using varying no of spectral features

Emotion Recognition Model	No of MFCCs				
	6	8	13	21	29
	Recognition performance in %				
Outrage	35	39	26	30	43
Cheerful	58	58	71	71	62
Unbiased	13	17	9	17	40
Pity	88	92	74	66	74
Astonishment	52	50	54	49	59
Average	49.2	51.2	46.8	46.6	55.6

SUMMARY AND CONCLUSIONS OF PRESENT WORK

In this work, semi common speech corpus, gathered from Hindi films has been utilized to describe and group the feelings. MFCC includes alongside their speed and increasing speed coefficients are utilized to catch feeling particular data from vocal tract of the speaker. The motivation behind this examination is to explore the

spectral highlights for their regulation of feeling particular data. Gaussian Mixture models, known to catch the circulation example of highlight vectors are utilized as feeling classifiers. From the acquired outcomes, it might be watched that 29 MFCCs, alongside Δ and Δ - Δ highlights have given better outcomes. It demonstrates that higher request otherworldly highlights contain better feeling particular data. The normal feeling acknowledgment execution for similar feelings on account of same UTU – SNESC database is impressively high contrasted with the aftereffects of various UTU - SNESC for the most part because of the impact of phonetic data on feeling acknowledgment. As duration of these investigations, prosodic highlights might be utilized as a part of mix with spectral highlights to additionally enhance the feeling acknowledgment execution of the models. Feeling acknowledgment in genuine situation expects dialect free feeling acknowledgment. In this manner, there is a need to create dialect autonomous feeling acknowledgment frameworks too.

REFERENCES

[1] D. Ververidis, C. Kotropoulos, and I. Pitas, "Automatic emotional speech classification," in ICASSP, pp. I593–I596, IEEE, 2016.

[2] Shashidhar G Koolagudi and K. Sreenivasa Rao, "Emotion Recognition from Speech: A Review", International Journal of Speech Technology, Volume 15, Issue 2, pp 99-117, Springer, June 2012.

[3] Van Bezooijen, "The Characterisitcs and Recognizability of Vocal Expression of Emotions", Drodrecht,The Netherlands: Foris, 1994

[4] Hansen, J. H. L., Cairns, D. A. ICARUS: Source generator based real-time recognition of speech in noisy stressful and Lombard effect environments, Speech Communication 16, pp 391–422, 1995.

[5] D. O'Shaughnessy, Speech Communication Human and Mechine, Addison- Wesley publishing company, 1999.

[6] M. Schroder, "Emotional speech synthesis: A review," in 7th European Conference on Speech Communication and Technology, (Aalborg, Denmark), Sept. 2001.

[7] T. L. Pao, Y. T. Chen, J. H. Yeh, and W. Y. Liao, "Combining acoustic features for improved emotion recognition in mandarin speech," in ACII (J. Tao, T. Tan, and R. Picard, eds.), (LNCS 3784), pp. 279–285, Springer-Verlag Berlin Heidelberg, 2005.

[8] Shashidhar G. Koolagudi, Sudhamay Maity, Vuppala Anil Kumar, Saswat Chakrabarti, and K. Sreenivasa Rao, "IITKGP-SESC: Speech Database for Emotion Analysis" in IC3, CCIS 40, pp. 485-492, Springer-Verlag, Berlin, Heidelberg 2009.

[9] D. Neiberg, K. Elenius, and K. Laskowski, "Emotion recognition in spontaneous speech using gmms," in INTERSPEECH - ICSLP, (Pittsburgh, Pennsylvania), pp. 809–812, 17-19 September 2006.

[10] [10] Li Y. and Zhao Y. Recognizing emotions in speech using short-term and long-term features. Proc. of the international conference on speech and language processing. pp. 2255-2258, 1998.

[11] S. G. Koolagudi, S. Maity, V. A. Kumar, S. Chakrabarti, and K. S. Rao, Emotions of ser. Communications in Computer and Information Science. Springer, August vol. 40 2009.

[12] K. S. Rao and B. Yegnanarayana, "Duration modification using glottal closure instants and vowel onset points," Speech Communication, vol. 51, p. 1263-1269, doi:10.1016/j.specom.2009.

[13] S. G. Koolagudi, S. Maity, V. A. Kumar, S. Chakrabarti, and K. S. Rao, IITKGP-SESC : Speech Database for Emotion Analysis. Communications in Computer and Information Science, JIIT University, Noida, India: Springer, issn: 1865-0929 ed., August 17-19 2015.

[14] Rahul Chauhan, Jainath Yadav, S. G. Koolagudi and K. Sreenivasa Rao, Text independent emotion recognition using spectral features, Communications in Computer and Information Science (CCIS): Contemporary Computing, Vol. 168, Part-2, pp. 359-370, Springer, 2015.

[15] L. R. Rabiner and B. H. Juang, Fundamentals of Speech Recognition. Englewood Cliffs, New Jersy: Prentice-Hall, 1993.