Variational Mode Decomposition based Emotion Recognition Speech Features from Voiced Regions using Thresholding Technique

Lakshmi Srinivas. D, Shaik Jakeer Hussain

Abstract*—Emotion recognition from speech signals is one of the latest research topics involving various emotional speech features for its classification. In this work, a variety of emotional speech features are extracted only from the voiced regions of the emotional speech signal. This algorithm includes the average energy distributed over the frequency range in wavelet domain and zero crossing rate for voiced region detection. The median of ratio of highest sub-band to lowest sub-bands energy and the average zero crossing rate of all segments is considered as thresholds for voiced region detection in speech signal. The voiced regions of speech are filtered to the low frequency range and divided into smaller voiced regions. Intrinsic mode functions and mean frequency components are calculated using Variational mode decomposition (VMD) and Hilbert transform in iterative way. Mean of mean frequencies and mean of inter-quartile range of intrinsic mode function of all speech segments are extracted as features, which provide variations in emotional speech for classification. Statistical parameters are calculated on these extracted features only from voiced regions of speech which provide easy process of classification.

***Index Terms*—Speech classification, Wavelet transforms, Empirical mode decomposition, Intrinsic mode functions, Hilbert Huang transform, Variational mode decomposition.**

I. INTRODUCTION

Speech is basic form of communication between humans and also for human machine interactions. It not only provides the information or message content alone, it also provides additional information related to speaker's gender, age and emotional state. Recent advances have been developed to find the nature of emotion of a speaker from speech signals. Emotion recognition from speech signals also provides useful information in identification of speaker recognition. Various speech features are used in different algorithms to determine the actual emotion of a person from speech [1]. As the number of predominant features increase in large number the algorithm tries to provide better accuracy. Speech can be divided into voiced and unvoiced region, which plays a vital role in many speech processing applications. Speech can be classified based on the vibration of vocal cords which determines the voicing and non-voicing regions of speech. The voiced regions and the unvoiced regions can be further defined based on the position of velum [2].

If the velum is closed, the production of speech is done by vocal cavity, and if open it is by nasal cavity. An important issue to be considered in the speech emotion recognition is the choice of features that play a vital role in the classification of emotions of a person from their speech signal. Emotion recognition from speech is a challenging task because, speech is a non stationary signal and the features extracted from it are not clear. The difference in acoustic signals from person to person and from sentence to sentence also adds a challenging task in recognition process. The selection of boundaries in speech signals is also an important task and the emotion of a person mainly depends on the speaker [3]. The features required for emotion classification can be categorized into continuous features, spectral features, qualitative features and Teager energy operator (TEO) features. The continuous features are like pitch related features, formant features, energy related features, timing features and articulation features. The various acoustic or speech features that are considered in the emotion recognition are pitch, fundamental frequency, energy, Mel-frequency cepstral coefficients (MFCC), linear predictive coefficients (LPC), linear predictive cepstral coefficients (LPCC) [3], [4].

Pitch, formant, fundamental frequency and energy related features provide the basic essential information regarding emotions. MFCC, LPC, LPCC are the time dependent acoustic features which provide more vital information regarding emotion of the person [5], [6], [7], [8]. Energy, degree of predominance and frequency are used as features in classifying positive and negative emotions of a person in emotion recognition [9], [10]. Wavelet packet, analysis coefficients and detailed coefficients of speech signal are also used as the features for emotion recognition [11], [12]. Fourier parameters are also used along with MFCC as feature vector for emotion recognition [13], [14], [15].

The rest of the paper focuses on: section II provides the details of the speech corpus that is adopted in this experimental work. Section III provides the basic knowledge of the discrete wavelet transform. Section IV introduces the concept of Variational mode decomposition of signals. Section V introduces the proposed algorithm, section VI provides experimental results and section VII provides conclusion.

Revised Manuscript Received on April 07, 2019.

Lakshmi Srinivas, D, Department of Electronics and Communication Engineering, VFSTR, Guntur, India.

Shaik Jakeer Hussain, Department of Electronics and Communication Engineering, VFSTR, Guntur, India.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

II. CORPUS

An emotional speech corpus, recorded by SMART Lab Ryerson University, Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) [16], is considered in this experimental work. It contains 1440 utterances of 24 speakers, 12 speakers are male and 12 speakers are female. Each speaker utters two statements two times in eight archetypal emotions: neutral, sad, angry, fearful, disgust and surprise, calm, happy. The neutral emotional speech features are considered as the reference for classifying the remaining emotional speech features. The repetition of the sentences in different emotions is also considered, to provide a detailed information regarding utterances in various emotions. This provides additional features for recognition of emotions from speech signal.

III. DISCRETE WAVELET TRANSFORM

The wavelet transform provides not only multi resolution analysis or multi-scale decomposition of signals, it also provides time frequency, information at a single time. It provides time frequency analysis by correlating the given set of filters from the family of basis wavelets with the given signal. The basis of wavelet analyses the complete signal by the property of dilation and shifting, which gives more information about the signal characteristics. Using the process of discrediting, the 2-D parameterization by a set of basis wavelets of mother wavelet. Following the function $\psi_{a,\tau}(t)$ is given by:

$$\psi_{a,\tau}(t)=\frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right) \quad a\in\Re,\tau\in\Re \qquad (1)$$

In discrete domain, the C (a, τ) coefficients of the discrete wavelet transform are divided into the approximation and the detail coefficients of the signal. The approximation coefficients provide the low frequency components and the detail coefficients provide the high frequency components of the signal $f(k)$ at level j [17], [18], [19]. The discrete wavelet transform approximate coefficients of the signal $f(k)$ at level j is given by

$$A_j=\sum_{n=0}^{\infty} f(n)\phi_{j,k}(n)$$
$$=\sum_{n=0}^{\infty} f(n)\frac{1}{\sqrt{2}}\phi\left(\frac{n-k2^j}{2^j}\right) \qquad (2)$$

where $\phi_{j,k}(n)$ is the scaling function of wavelet, $\psi_{a,\tau}(n)$. In same way, the detail coefficients of the signal $f(k)$ is given as

$$D_j=\sum_{n=0}^{\infty} f(n)\frac{1}{\sqrt{2^j}}\psi\left(\frac{n-k2^j}{2^j}\right) \qquad (3)$$

The decomposition of the signal $f(k)$ can be iterated based on the above equations as the number of level increases. A well defined approximation and detail coefficient set can be constructed through different levels of decomposition, which is known as multi-resolution analysis (MRA). Useful information related to the signal $f(k)$ can be extracted from these wavelet level decompositions. Considering the finite set of stages, introducing the scaling function $\phi(t)$and the reconstructed function $f(t)$ can be defined as

$$f(t)=\sum_k c_{l,k}\phi_{l,k}(t)+\sum_k d_{j,k}\psi_{j,k}(t) \qquad (4)$$

IV. VARIATIONAL MODE DECOMPOSITION

The basic concept of VMD is an extended version of Empirical Mode Decomposition (EMD) technique. This technique decomposes the given nonlinear, non-stationary signal into its finite number of zero mean components called as Intrinsic Mode Functions (IMF) [20], [21], [22]. The basis functions for the EMD technique are derived from the data itself. The instantaneous frequency of these IMFs can be calculated through the analytic signal method called as Hilbert Huang Transform [23]. EMD based technique extracts the energy associated with the various intrinsic scales from finer to coarser. Each of the IMF satisfies two properties:

1) From whole dataset, number of extrema and number of zero crossings must be equal or may differ by one.
2) At any time, the mean value of the envelope function defined by the local maxima and the minima is zero.

The steps to compute EMD technique are:

1) Compute all the local maxima and minima of the signal$x(t)$.
2) By using cubic spline interpolation of the maxima and minima points computed in previous step, create an upper envelope $e_u(t)$ and lower envelope $e_l(t)$.
3) Calculate the envelope mean of the two, upper envelope $e_u(t)$ and lower envelope $e_l(t)$ which is $m_1(t)=\frac{e_u(t)+e_l(t)}{2}$.
4) Compute the new signal, $d_1(t)=x(t)-m_1(t)$. If $d_1(t)$ is a zero mean function, then consider it as an IMF. If not consider $d_1(t)$ as new data and find the new IMF by repeating the steps.

This process is repeated until the new signal has no maxima or minima. $x(t)=\sum_{i=0}^{N} a_i(t)+res(t)$, where $res(t)$is considered as the residue function.

EMD and EMD- wavelet based analysis of signals provide better results for the speech and biomedical signals [24], [25], [26]. The selection of predominant IMF is done based on the energy content associated with it. In EMD-wavelet analysis, the dominant IMF are selected and passed through the wavelet filters for analysis purposes. Variational Mode Decomposition (VMD) is the latest technique used to analyze the non linear, non stationary signals along with its bandwidth. The main aim of VMD technique is to decompose the given real signal into the number of discrete sub-signals called as IMFs or modes, where each mode is compact around the center pulsation w_k, which is determined in the decomposition process. The instantaneous frequency and the amplitude of each IMF are calculated by Hilbert transform, which provides a unilateral spectrum [27], [28], [29].

The bandwidth of each mode can be calculated by following steps:

1) Obtain the unilateral spectrum of each mode by calculating the Hilbert transform.
2) Shift the unilateral spectrum to the baseband by mixing with exponential which is tuned to the center frequency.
3) H^1 Gaussian smoothness is performed to obtain the bandwidth of the signal. The resulting constrained problem is defined as:

$$\min_{u_k, w_k} \left\{ \sum_k \left| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * d_k(t) \right] e^{-j\omega_k t} \right|^2 \right\} \quad (5)$$

subject to: $\sum_k d_k = x(t)$ where $d_k(t)$ are IMFs and w_k are the center frequencies respectively.

V. PROPOSED ALGORITHM

The block diagram representation of the proposed algorithm is given in Figure1. The speech signal which is sampled at 48 KHz is normalized, so that all the signals have same power. The speech signal after normalization is segmented into frames of 20msec which contains both voiced regions and unvoiced regions of speech.

A. Median Energy in Frequency Domain

All the segments of the speech signal are decomposed into eight different bands using a 4-level dyadic DWT and average energy is computed in band wise in the last five sub-bands. Generally, unvoiced regions of speech segment represent their energy in high frequency region while voiced regions of speech segment represent their energy in the low frequency regions. Per segment ratio R_i is calculated by ratio of the average energy of the wavelet low-bands to that of the wavelet high-bands.

Let us consider E_i be the total energy content in the i^{th} segment of the signal and E_j be the energy in the j^{th} wavelet band. $\sum_i = \sum_j = \sum_k$ where j=1 represents the highest frequency band and j=2-8 represents the subsequent bands in the wavelet domain. In this paper, j=5 is considered as the highest frequency region in the wavelet domain and j=8 as the lowest frequency band in wavelet domain and per segment ratio R_i is calculated as $R_i = \frac{\sum_{j=6}^{8} E_j}{E_5}$. By calculating the per segment ratio for each frame, the median energy for frame is considered to be threshold for classifying into voiced and unvoiced regions based on energy, $Th_1 = median\left(\sum_k R_k \right)$. A particular segment is said to be voiced if its corresponding $R_i > Th_1$.

B. Median Zero Crossing Rate

Zero crossing is defined as change in the sign of amplitudes between two successive samples of the speech segment, which determines the frequency content present in that particular speech segment. This is usually defined by the expression:

$$ZCR_i = \sum_{n=0}^{N-1} \left| \mathrm{sgn}\left(x_i(n)\right) - \mathrm{sgn}\left(x_i(n+1)\right) \right| \quad (6)$$

where N denotes the samples in the i^{th} speech segment of emotional speech signal. The number of zero crossing for each segment of speech are calculated and the median of overall zero crossings is considered as another threshold for identifying voiced regions of speech, $Th_2 = median\left(\sum_k ZCR_k \right)$. A segment is said to be voiced if its $ZCR_i < Th_2$.

C. Feature Based Voiced and Unvoiced Segmentation

Even though, individual features provide a great result in the classification of the voiced and the unvoiced regions of speech, the combination of per segment energy ratio and the zero crossing rate provide a better accurate results in the classification. A particular segment is said to be voiced segment of the speech signal if satisfies the following condition:

$$\left[\left(R_i > Th_1 \right) \& \left(ZCR_i \le Th_2 \right) \right] \Rightarrow \left(i \in V \right) \quad (7)$$

D. Data Smoothing

The voiced regions of the speech signal are segmented based on above criteria and each segment is given a frame id number to identify the voiced segments easily. The frame id numbers are used and arranged in an ascending order, there by the new speech signal contains only the voiced regions in a predefined order, which does not change the information content or meaning in the actual speech signal. The obtained segments are passed through a basic smoothing filter so as to avoid any mismatches between the two adjacent segments, thereby providing a smooth transition over the voiced segments of speech signal.

E. Determining the Intrinsic Mode Functions

Variational mode decomposition algorithm is applied after logical smoothing of the voiced regions of the speech signal in order to determine the intrinsic mode functions of various emotional speech signals. This parameter varies from one emotion to another emotion. The number of mode functions that are created is 5 with an initial ω value to be 1. The weight updations are done there by providing the optimal weights for extracting the bandwidth of each intrinsic mode functions. Applying Hilbert transform as said in the VMD technique, the phase and instantaneous frequencies are calculated, the mean instantaneous frequencies of each intrinsic mode function is also calculated and are used as a parameters.

F. Statistical Parameter Calculations

The statistical parameters like mean and interquartile are calculated from the extracted features for each frame. The mean is going to provide the overall information in a precise value. Since the mean cannot provide the information in outliers, the interquartile range is calculated which provides the spread of the values in different frames of speech. These two statistics provide better results for analysis. Let us consider the instantaneous frequencies to be as, $x_1, x_2, \ldots\ldots\ldots x_n$. where n is the number of voiced frames in the speech signal.

$$X\_mean = \frac{x_1, x_2, \ldots\ldots\ldots\ldots x_n.}{n} \quad \text{- (8)}$$

The interquartile range is calculated by following steps.

1. Arrange the values $x_1, x_2, \ldots\ldots\ldots\ldots x_n$ in ascending order.
2. Calculate the median for the arranged values by the centre value of arranged set.
3. Divide the total values into twoas lower half of median and higher half of median and calculate the median for two .
4. The IQR is subtracting the median values of the two sets.

The mean and the interquartile are the two statistical properties which provide the average or most approximate value of the given values and the interquartile range provides the spread and it provides the information of given values which includes outliers for calculation. The median value is calculated by calculating the interquartile range. Inter quartile range of all the 5 intrinsic mode functions is calculated and quartile range of each intrinsic mode is normalized by multiplying with a constant as γ, which boots up the signal.. Inter quartile ranges of 5 modes vary from emotion to emotion in speech signal which provides better accuracy in differentiating emotions based on these values.

In this algorithm, overlapping is avoided so as to avoid the confusion in smoothing all the voiced regions off the speech signals. The above steps are computed in an order to extract the features of various emotional speech signals, which provide better accuracy in basic classification of various emotions. The complete algorithm is given in the figure 1 and the voiced segments of various emotional speech waveforms are given in figure2.

In this work, different emotional speech signals like neutral, calm, happy, sad, angry, fearful, disgust and surprise are considered. Each speech signal is segmented into frames of 20 millisecond duration. First, Zero crossing is calculated for each frame and discrete wavelet transform, Daubechies6 wavelet is applied on each segment of each speech signal. According to the defined thresholds in this work, based on the median value of energy per segment ratio and total zero crossings of all the segments of the speech signal, each segment of emotion speech signal is separated into voiced regions and the unvoiced regions of speech. Various wavelets like Haar, Daubechies2 and Daubechies6 wavelets are applied for the classification of voiced regions of speech and Daubechies6 wavelet provides better results compared to the Haar, Daubechies2 wavelets.

So in this work, Daubechies6 wavelet is considered for the classification of voiced regions.

Variational Mode Decomposition technique is one of the best decomposition techniques for the analysis of non stationary signals. VMD technique is applied on the each segment of voiced regions of emotion speech signal to extract intrinsic mode functions, mean frequencies and center frequencies for all voiced segments of emotional speech. Again mean value of the above calculated parameters of voiced segments is calculated and these features are used as statistical feature set, to classify different emotions. This provides the reduction of feature vector for classification of different emotions from speech signal. The process of calculation of these statistical parameters repeated till the end of voiced regions of speech is identified.

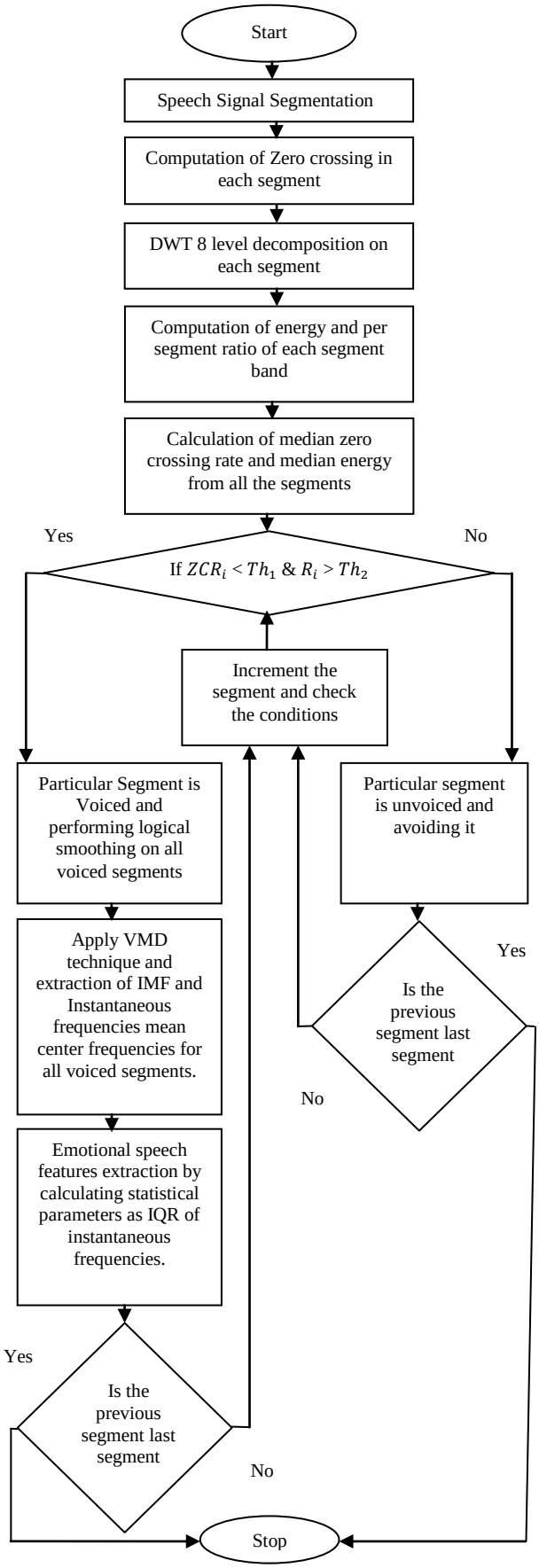
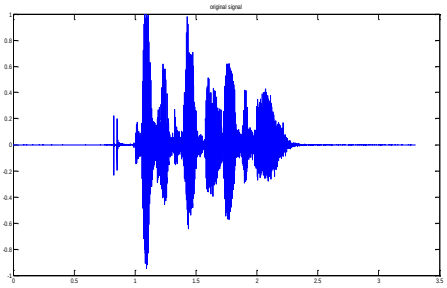
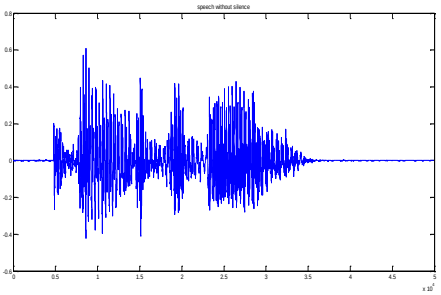


Fig: 1 Algorithm of VMD based emotional speech feature extraction from voiced regions.

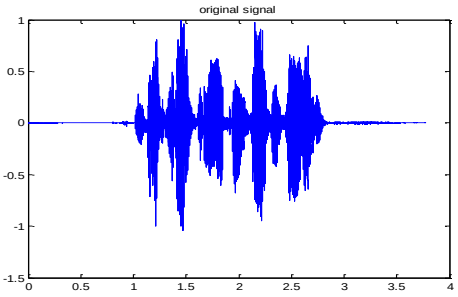
VI. EXPERIMENTAL RESULTS



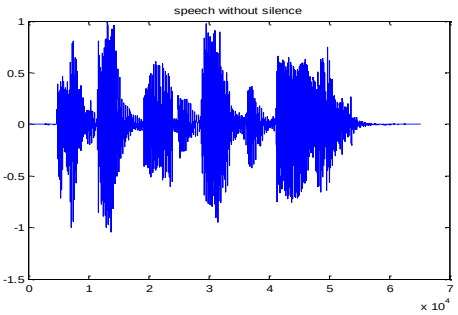
a) Actual neutral Speech.



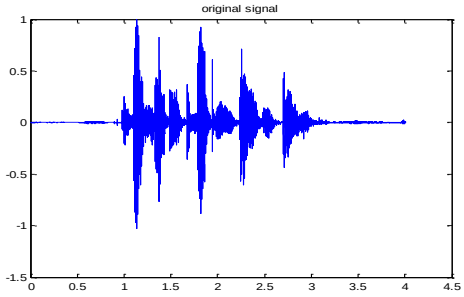
b) Voiced regions of neutral speech.



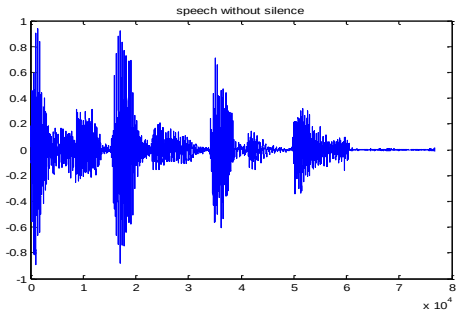
c) Actual Happy speech signal.



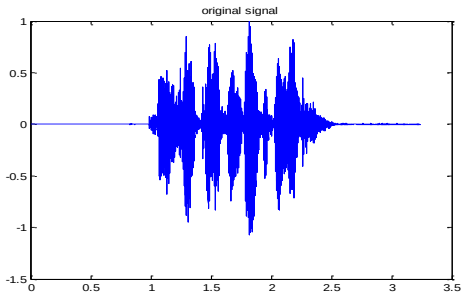
d) Voiced regions of happy speech.



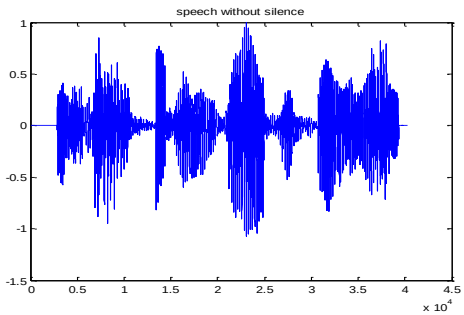
e) Actual angry speech signal.



f) Voiced regions of angry speech.



g) Actual surprise speech signal.



h) Voiced regions of surprise speech.

Fig.2 Results of voiced regions segmentation of emotional speech.

In this experiment, RAVDESS speech data base is considered and it has shown the variations in the identification of emotions of various speakers based on the emotional speech features.

Emotions	Persons	Mean_MF1	Mean_MF2	Mean_MF3	Mean_Iqr_IMF1	Mean_Iqr_IMF2	Mean_Iqr_IMF3
Neutral	Person1	215.006	3194.6	8746.91	8616.52	2133.62	131.075
	Person2	225.157	3705.51	8817.27	6993.12	973.289	113.971
	Person3	296.356	4183.9	8213.8	9667.03	736.584	137.75
Calm	Person1	210.318	3020.78	8518.31	11177.4	2686.27	194.107
	Person2	247.54	3535.88	8017.89	8479.18	2162.6	416.563
	Person3	242.987	3813.65	9285.27	12424	1592.91	162.881
Happy	Person1	224.154	3343.94	8133.59	9145.59	2601.52	618.165
	Person2	265.75	4066.54	8614.43	3905.45	361.143	86.1299
	Person3	356.237	3711.26	8308.04	20568.7	1975.07	630.598
Sad	Person1	261.056	3101.9	8452.54	8677.87	1086.59	215.552
	Person2	343.536	2849.39	7577.88	18413.5	5512.17	523.457
	Person3	253.13	3343.56	8776.27	7272.66	1184.62	.72.6975
Angry	Person1	304.379	3726.86	8967.03	9385.54	1434.43	104.984
	Person2	292.673	3926.03	8725.08	8335.69	887.705	274.284
	Person3	279.107	4042.12	9509.9	7443.52	652.043	123.32
Fearful	Person1	250.207	3565.94	9383.69	10620.3	1256.6	75.0438
	Person2	438.233	4105.45	8067.13	16639.1	1110.69	628.315
	Person3	303.041	4050.97	9370.35	10965.1	1131.84	138.468
Disgust	Person1	275.021	3129.17	8591.45	17583	4198.19	927.995
	Person2	416.243	3755.78	7262.87	18659.8	1922.39	881.714
	Person3	334.503	3755.66	9404.69	5670.98	773.569	94.6126
Surprise	Person1	222.644	3501.2	8990.06	7049.43	802.672	50.7918
	Person2	323.688	3805.6	8558.68	11668.9	1645.05	457.703
	Person3	369.368	4098.59	8255.79	24313.6	2994.02	316.232

The emotional speech feature vector of speaker, both male and female is shown in the table, person 1 and 3 are male and person 2 is female, which contains the variations in the mean frequency of each intrinsic mode functions and the mean of inter-quartile range of intrinsic mode functions of all segments of each emotion. Each speaker has unique emotional speech vector which can classify the various emotions easily. The inter-quartile range feature of the intrinsic mode functions is normalized by multiplying with a higher value to classify in large extent. The difference between the various emotional speech features is made higher which provides the difference to be significant. The tabulated values are also plotted in the graph and it represents the basic difference in the emotional speeches of the speaker. The eight emotions neutral, sad, angry, fearful, disgust and surprise, calm, happy

feature vectors are plotted in the graph and is given in the figure 3.

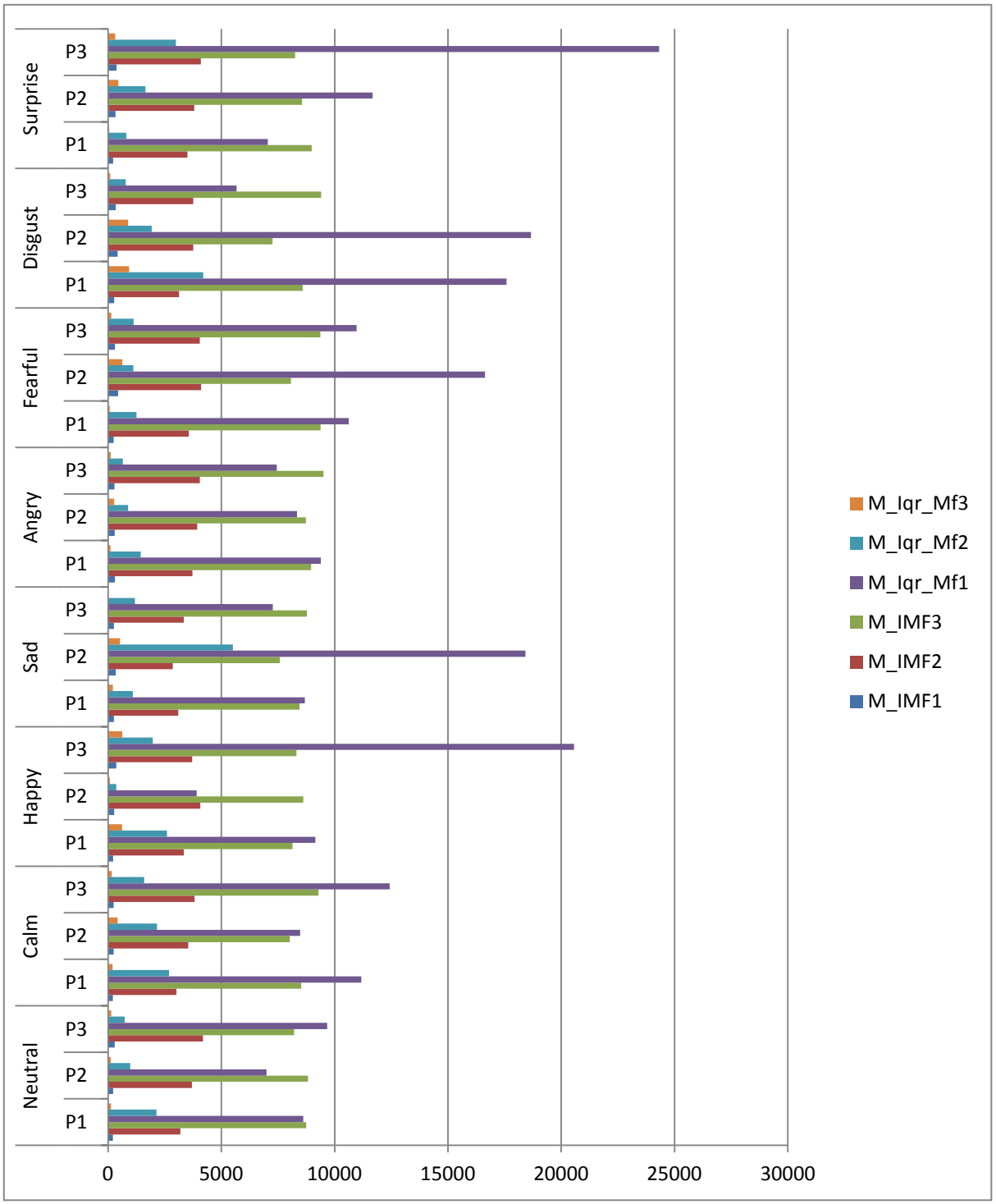


Fig. 3 Graphical representation of feature vectors of emotional speech signals of different persons.

VII. CONCLUSION

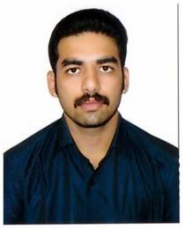
It can be seen that, the proposed algorithm can be used in clean and noisy environments for the feature extraction. The emotional speech signal is decomposed and the features extracted from the variational mode decomposition are more like to be unique, as the intrinsic mode functions and mean frequencies are extracted from the speech signal directly by decomposing the signal without applying transformation techniques. Since most of the information lies in the voiced regions of speech, the features are extracted only from the voiced regions of the speech signal. This provides better stability for the classification of different emotions and the features are directly obtained from the basic signal without any separate basis functions as in Fourier transform. The features are not only speaker independent and also different for different emotions without any gender bias. The higher order intrinsic mode functions and mean frequencies are only

considered and since the feature vector contains also the statistical values of these parameters, the classification along with the MFFC's also provides better accuracy.

REFERENCES

1. Koolagudi, Shashidhar G., and K. Sreenivasa Rao. "Emotion recognition from speech: a review." International journal of speech technology 15, no. 2 (2012): 99-117.
2. Rabiner, Lawrence R., and Ronald W. Schafer. Theory and applications of digital speech processing. Vol. 64. Upper Saddle River, NJ: Pearson, 2011.
3. El Ayadi, Moataz, Mohamed S. Kamel, and Fakhri Karray. "Survey on speech emotion recognition: Features, classification schemes, and databases." Pattern Recognition44, no. 3 (2011): 572-587.
4. Xu, Xin, Ya Li, Xiaoying Xu, Zhengqi Wen, Hao Che, Shanfeng Liu, and Jianhua Tao. "Survey on discriminative feature selection for speech emotion recognition." In Chinese Spoken Language Processing (ISCSLP), 2014 9th International Symposium on, pp. 345-349. IEEE, 2014.
5. Chen, Lijiang, Xia Mao, Yuli Xue, and Lee Lung Cheng. "Speech emotion recognition: Features and classification models." Digital signal processing 22, no. 6 (2012): 1154-1160.
6. Kerkeni, Leila, Youssef Serrestou, Mohamed Mbarki, Kosai Raoof, and Mohamed Ali Mahjoub. "A review on speech emotion recognition: Case of pedagogical interaction in classroom." In Advanced Technologies for Signal and Image Processing (ATSIP), 2017 International Conference on, IEEE, 2017: 1-7.
7. Renjith, S., and K. G. Manju. "Speech based emotion recognition in Tamil and Telugu using LPCC and hurst parameters—A comparitive study using KNN and ANN classifiers." In Circuit, Power and Computing Technologies (ICCPCT), 2017 International Conference on, IEEE, 2017: 1-6.
8. Basu, Saikat, Jaybrata Chakraborty, Arnab Bag, and Md Aftabuddin. "A review on emotion recognition using speech." In Inventive Communication and Computational Technologies (ICICCT), 2017 International Conference on, IEEE, 2017: 109-114.
9. Pavaloi, I., A. Ciobanu, M. Luca, E. Musca, T. Barbu, and Anca Ignat. "A study on automatic recognition of positive and negative emotions in speech." In System Theory, Control and Computing (ICSTCC), 2014 18th International Conference, IEEE, 2014: 221-224.
10. Pavaloi, Ioan, Elena Musca, and Florin Rotaru. "Emotion recognition in audio records." In Signals, Circuits and Systems (ISSCS), 2013 International Symposium on, IEEE, 2013: 1-4.
11. Wang, Kunxia, Ning An, and Lian Li. "Speech emotion recognition based on wavelet packet coefficient model." In Chinese Spoken Language Processing (ISCSLP), 2014 9th International Symposium on, IEEE, 2014: 478-482.
12. Shahnaz, C., and S. Sultana. "A feature extraction scheme based on enhanced wavelet coefficients for speech emotion recognition." In Circuits and Systems (MWSCAS), 2014 IEEE 57th International Midwest Symposium on, IEEE, 2014: 1093-1096.
13. Wang, Kunxia, Ning An, Bing Nan Li, Yanyong Zhang, and Lian Li. "Speech emotion recognition using Fourier parameters." IEEE Transactions on Affective Computing 6, no. 1 (2015): 69-75.
14. Gupta, Shikha, Jafreezal Jaafar, WF Wan Ahmad, and Arpit Bansal. "Feature extraction using MFCC." Signal & Image Processing: An International Journal (SIPIJ) 4, no. 4 (2013): 101.
15. Ververidis, Dimitrios, and Constantine Kotropoulos. "Emotional speech recognition: Resources, features, and methods." Speech communication 48, no. 9 (2006): 1162-1181.
16. Livingstone, Steven R., and Frank A. Russo. "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English." PloS one 13, no. 5 (2018): e0196391.
17. Mallat, Stéphane. A wavelet tour of signal processing. Elsevier, 1999.
18. Jafer, E., and A. E. Mahdi. "Wavelet-based voiced/unvoiced classification algorithm." In Video/Image Processing and Multimedia Communications, 2003. 4th EURASIP Conference focused on, vol. 2, IEEE, 2003: 667-672.
19. Uribe, Alejandro, Alejandro Gómez, Manuela Bastidas, O. Lucia Quintero, and Damian Campo. "A novel emotion recognition technique from voiced-speech." In Automatic Control (CCAC), 2017 IEEE 3rd Colombian Conference on, IEEE, 2017: 1-4.
20. Maheshwari, Sonam, and Ankur Kumar. "Empirical Mode Decomposition: Theory & Applications." International Journal of Electronic and Electrical Engineering. ISSN (2014): 0974-2174.
21. Boudraa, Abdel-Ouahab, Jean-Christophe Cexus, Fabien Salzenstein, and Laurent Guillon. "IF estimation using empirical mode decomposition and nonlinear Teager energy operator." Proc. IEEE ISCCSP (2004): 45-48.
22. Xin, Li, and Li Xiang. "Novel Hilbert Energy Spectrum Based Features for Speech Emotion Recognition." In Information Engineering (ICIE), 2010 WASE International Conference on, vol. 1, IEEE, 2010: 189-193.
23. He, Ling, Margaret Lech, Namunu C. Maddage, and Nicholas B. Allen. "Study of empirical mode decomposition and spectral analysis for stress and emotion classification in natural speech." Biomedical Signal Processing and Control 6, no. 2 (2011): 139-146.
24. Kabir, Md Ashfanoor, and Celia Shahnaz. "Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains." Biomedical Signal Processing and Control7, no. 5 (2012): 481-489.
25. Chu, Yun Yun, Wei Hua Xiong, and Wei Chen. "Speech emotion recognition based on EMD in noisy environments." In Advanced materials research, vol. 831, Trans Tech Publications, 2014: 460-464.
26. Shahnaz, Celia, Sharifa Sultana, Shaikh Anowarul Fattah, RH Md Rafi, Istak Ahmmed, W-P. Zhu, and M. Omair Ahmad. "Emotion recognition based on EMD-Wavelet analysis of speech signals." In Digital Signal Processing (DSP), 2015 IEEE International Conference on, IEEE, 2015: 307-310.
27. Dragomiretskiy, Konstantin, and Dominique Zosso. "Variational mode decomposition." IEEE transactions on signal processing 62, no. 3 (2014): 531-544.
28. Sai Chaitanya kumar. "Variational Mode Decomposition and Multiple Feature Segmentation on Microarray Images." In International Journal of Computational Intelligence Research(IJCIR), no. 7(2017), 1777-1787.
29. Upadhyay, Abhay, Manish Sharma, and Ram Bilas Pachori. "Determination of instantaneous fundamental frequency of speech signals using variational mode decomposition." Computers & Electrical Engineering 62 (2017): 630-647.

AUTHORS PROFILE



Lakshmi Srinivas D was born on August, 1992. He received the Bachelor of Technology and Master of Technology degrees from VFSTR University and KL University, in 2013 and 2015, respectively.

In 2014, he did his Master's degree project in Indian Space Research Organization on Jet Acoustic signals. He published an article on Noise Source Localization in Jet Acoustics using Advanced Signal Processing Techniques (2016: LAMSYS- India). His research study includes Acoustic Signal Processing and Speech Processing. He is presently working in speech processing.

Jakeer Hussain SK was born on November, 1975. He received his Bachelor of Engineering degree in 1996 from Andhra University College of Engineering. He received Master of Business Administration and Master of Technology degrees from B. R. Ambedkar University and Jawaharlal Nehru Technological University, Hyderabad in 2006 and 2008 respectively. He received Doctor of Philosophy degree from Jawaharlal Nehru Technological University, Hyderabad in 2016.

He worked in industry till 2006 and moved to teaching. He had 14 publications in various International Conferences and 5 articles in reputed Journals. His research works include Speech Processing and Epileptic Seizure. His interests are Digital Signal Processing and Bio Medical Signal Processing.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication