www.ijemr.net ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

Volume-5, Issue-3, June-2015

International Journal of Engineering and Management Research

Page Number: 275-279

Comparative Study of Statistical Moments and Entropies of Wavelet Coefficients for Speech Emotion Recognition

Prashant P. Patil[1], Dr. A. S. Bhalchandra[2]
[1, 2]Department of Electronics & Telecommunication, INDIA

ABSTRACT

Emotion Recognition plays an important role in robust speech recognition. In this paper, for emotion recognition the Wavelet decomposed coefficients of speech are used. The Wavelet approximate and detail coefficients are improved using Teager Energy Operator. This improved coefficient's entropy is calculated in feature extraction stage and used for emotion classification. The analysis is carried out on Polish Emotional Database. The four emotions namely anger, joy, neutral and sad are considered which creates a four class problem. The Euclidean distance is applied as a feature classifier and giving the nearest emotion that matches to test input speech. The performance is evaluated based on the ability of system to recognize emotion independent of speaker. Teager Energy Operator which reflects the nonlinear vortex flow interaction of speech and entropy as a feature vector truly minimizes the calculations for emotion recognition. The entropy as a feature outperforms the other statistical features extracted from coefficients.

Keywords— DWT, Entropy, Euclidean distance, Teager Energy Operator(TEO).

I. INTRODUCTION

The classification of emotions such as angry, sad, happy has application in robust speech recognition, lie detectors, video games, psychiatric aid, emergency call sorting, stressful environment (aircraft cockpit, battlefield) and animal behavior understanding. The existing systems for emotion recognition shows higher accuracy rate for Speaker dependent systems than that of speaker independent systems [1] [2]. In this paper two methods are compared, one using statistical features extracted from Wavelet coefficients and another proposed method based on entropy as a feature. The feature extracted in first are statistical parameters such as Mean, Kurtosis, Skewness, Standard Deviation of decomposed Wavelet coefficients and in second case the feature used is improved wavelet coefficient's entropy. The entropy as a feature gives comparable results with less number of calculations. The systems are developed for speaker independent emotion recognition. No prior gender classification or hierarchical approach of emotions is done. The simple Euclidean classifier for feature classification gives better result with lesser complexity.

II. SYSTEM DEVELOPMENT

The proposed Emotion Recognition Systems mainly consist of Feature extraction stage and Feature classifier stage. The feature classification stage remains the same for both methods which is based on simple Euclidean Distance. The first system uses entropy of improved Wavelet coefficients as a feature vector and in another case feature vector is derived from wavelet coefficients statistical parameters such as Mean, Kurtosis, Skewness, Standard Deviation. Fig 1 shows the different blocks that are associated with general emotion recognition system.

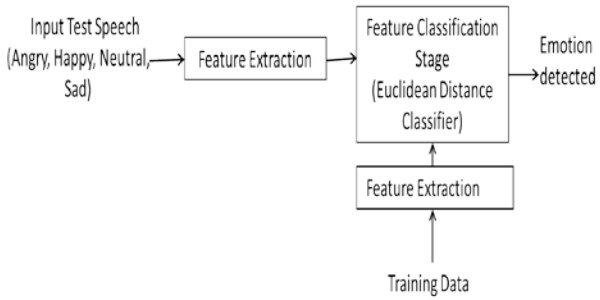

Figure1.Emotion Recognition System

A.Feature Extraction Stage:

The accuracy of correctly recognizing given emotion is depending on features extracted from speech. The effective features give distinct values for different

emotion class and hence provide correct emotion detection. In this system we calculated the discrete wavelet transform of given speech. The Teager Energy Operator is then applied on approximate and detail coefficients of wavelet decomposition. The entropy values for coefficients are concatenated which is then used as a feature vector. Fig2 explains the flow of processes that to be followed during entropy based feature extraction stage.

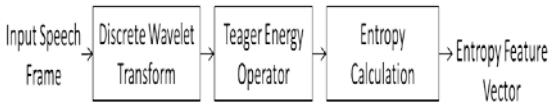

Figure 2.Feature Extraction Process for Entropy based Emotion Recognition

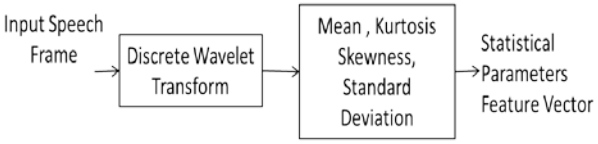

Figure 3.Feature Extraction Process for Statistical Parameter based Emotion Recognition

***1. Discrete Wavelet Transform*:**

To analyze non stationary speech signal, time and frequency both resolution are important. Fourier transforms which gives frequency information of the signal but does not provide exact time location of frequency component. The wavelet transform is best suited for speech signal giving better resolution in time and frequency domain by varying scale and translation parameter. Discrete wavelet transform analyzes speech signal at different frequency bands with different resolutions by decomposing it into approximate and detail coefficients [4]. This is achieved by implementing successive high pass and low pass filtering of the time domain speech signal. In wavelet transform after each filtering stage the numbers of samples taken into consideration are halved which results in less resolution in time domain but increases resolution in frequency domain of signal.

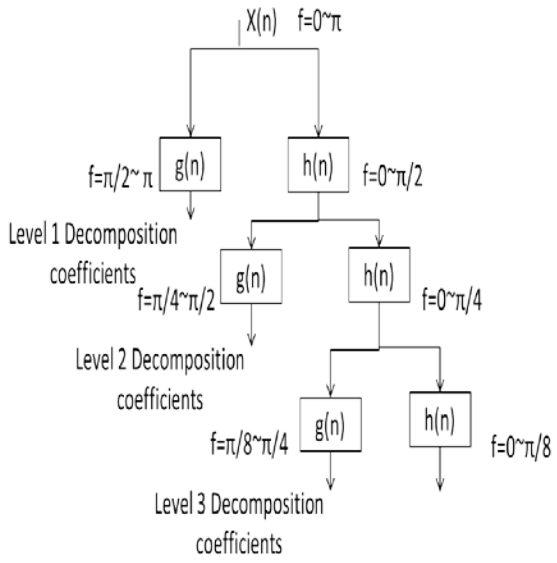

Figure 3.Three levels Wavelet Decomposition using High pass and Low pass filters

The frequency is represented in radian. Further decomposition takes into account the lower frequency component of the signal. Figure 3 gives three level decomposition using high pass and low pass filters. The six level wavelength decomposition using Discrete Meyer (dmey) Wavelet is applied for both systems.

2. Teager Energy Operator (TEO):

This energy operator maps the behavior of instantaneous energy of non-linear vortex flow interactions [5]. It supports the concept that hearing is nothing but detection of energy. The Teager Energy Operator in discrete time form is given by

$$\text{TE}\{X[n]\}=X[n]^2 - X[n+1]X[n-1] \quad (1)$$

Where $X[n]$ is the band limited discrete time speech signal. As the amplitudes of approximate and detail coefficients have reduced scale. Figure 4 shows the approximate coefficients before applying the Teager Energy Operation. The energy operator is applied to each coefficient giving enhanced Wavelet coefficients. The amplitude of the approximate and detail coefficients is improved by TEO.

ISSN (ONLINE): 2250-0758, ISSN (PRINT): 2394-6962

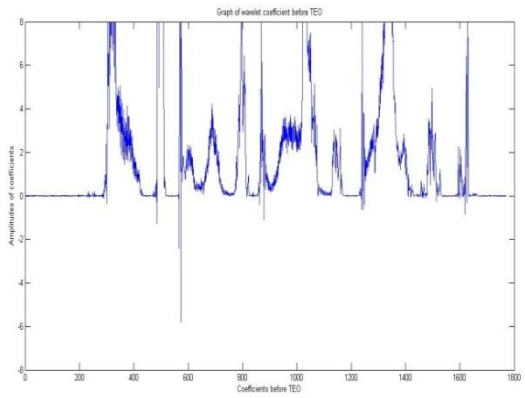

Figure 4.Plot of Wavelet Approximate Coefficients before Teager Energy Operation

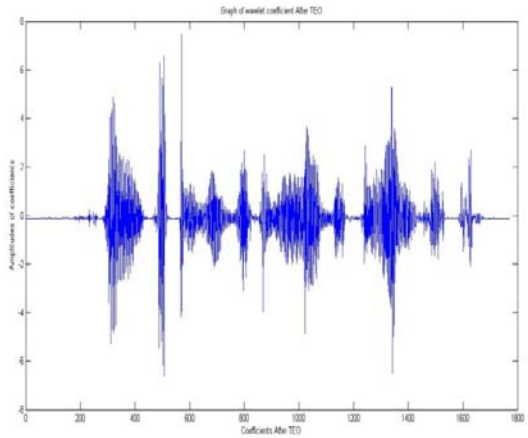

Figure 5.Plot of Wavelet Approximate Coefficients after Teager Energy Operation

3. Entropy of The signal:

The entropy is nothing but information content of the signal. More the unpredictability of the random variable more the information the signal has. The entropy of the signal can be given as

$$\text{Entropy} = -\sum_{j=1}^{N} P(Y_j)\, log_2(P(Y_j)) \quad (2)$$

Entropy is chosen as a feature vector instead of energy of the signal because energy gives the output which depends only on the amplitudes of the signal [1] [3]. Though some positive results are observed using energy as a feature to classify sad and angry or loud and sad it is observed that energy of the signal cannot find significant difference between angry and joy speech. Entropy is in the form of probabilistic model of the signal which can be seen from the equation no (2). Entropy provides closed values for the same class emotion and distinct values among different emotions. Considering Teager Energy Operated approximate and detail discrete wavelet coefficients the proposed feature vector is prepared as

$$\text{Feature Vector for EntropyMethod} = [C_{Approx}\ C_{Detail}] \quad (3)$$

Where, C_{Approx} stands for entropy value for approximate Wavelet coefficients and C_{Detail} stands for entropy values for detail Wavelet coefficients for each speech frame. Instead of choosing all Wavelet coefficients as feature vector, entropy greatly reduces the size of feature vector and hence the less number of mathematical calculations. In this stage feature vector is calculated for test input and for all four emotions training speech samples. Table I shows entropy values of different wavelet coefficients for different class of emotions. The distinguish values of entropies for different emotion classes can be observed from table.

	CA6	CD1	CD2	CD3	CD4	CD5	CD6
Angry1	-294009	1.82E-09	1.82E-05	0.271068	5.038337	-2227.29	-4669.1
Angry2	-249994	1.77E-09	0.000215	2.633521	14.57984	23.20466	-66355.7
Joy1	-1939.4	4.29E-09	0.000118	0.810541	1.412028	-3187.98	-3732.73
Joy2	-463.838	5.10E-09	1.07E-05	2.10476	22.35007	-118.67	-1200.9
Neu1	60.43979	7.16E-09	9.22E-07	0.001752	0.018672	2.110966	5.743531
Neu2	140.3813	5.46E-09	1.88E-06	0.006584	0.23349	0.277856	-531.874
Sad1	-931.982	1.20E-08	4.95E-06	0.005294	0.000338	0.025408	11.24879
Sad2	-184.834	5.09E-09	4.53E-06	0.347646	0.107517	0.51232	44.15075

4. Mean of the Wavelet Coefficients:

The average value of absolute wavelet coefficients is taken as one of the statistical feature. The mean (μ) is needed for calculations of higher statistical moments. It is given by formula,

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

5. Kurtosis of the Wavelet Coefficients:

The Kurtosis of the Wavelet Coefficients distribution is defined as

$$\text{Kurtosis (k)} = \frac{E(x-\mu)^4}{\sigma^4} \quad (4)$$

Where, μ is mean of the x, σ is standard deviation of x, E denotes the expected value of the bracketed term. Kurtosis value indicates how outlier-prone the coefficient distribution is.

6. Skewness of the Wavelet Coefficients:

The Skewness of the Wavelet Coefficients distribution is defined as

$$\text{Skewness (s)} = \frac{E(x-\mu)^3}{\sigma^3} \quad (4)$$

Where, μ is mean of x, σ is standard deviation of x, E denotes the expected value of the bracketed term. The Skewness value indicates how asymmetrical the coefficients are around the sample mean.

7. Standard Deviation of the Wavelet Coefficients:

The Standard deviation(s) of the Wavelet Coefficients distribution is defined as

$$s = \left(\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu)^2 \right)^{\frac{1}{2}}$$

Where,

$\mu$is the mean of *x*, N is the number of coefficients in *x* .The standard deviation how the coefficients are spread with respect to the mean value. The feature vector for Statistical parameter based system is obtained by concatenating Mean, Kurtosis, Skewness and Standard Deviation values of Wavelet approximate and detail coefficients.

Feature Vector for Statistical Parameter Method
= [Mean Kurtosis Skewness Standard_Deviation]

B.Feature Classification Stage:

The task of classifier is to find most similar speech from training data that matches to the test input. The simplest Euclidean Distance between test speech feature vector and training data feature vectors is used for classifying emotions. The minimum value of Euclidean distance will indicate the nearest emotion from all emotion classes that matches to the test input speech. The test input in our case is taken from database only and not to consider it in training data while classifying. Consider the speech is divided into L number of frames and feature vector consist of m number of features representing entropies of Wavelet coefficients for each frame. Let K^{th} frame of Test input speech T has feature vector $\{K_{Tn}(1), K_{Tn}(2), K_{Tn}(3), \ldots . K_{Tn}(\mathrm{m}),\}$ with the database suppose emotion speech D with vector $\{L_D(1), L_D(2), L_D(3) \ldots L_D(m)\}$ is given by

$$Euclidean = \sum_{n=1}^{q} \sum_{j=1}^{m} |K_{Tn}(j) - L_D(j)|^2.$$

Using above equation, Euclidean distance is calculated between test input and all class of emotions. Now the minimum distance is located and corresponding class is treated as most matching emotion class for test input.

IV. SIMULATION RESULTS

Database and Simulation Conditions:

Speech sentences from Polish Emotional Database are used for Emotion Recognition System. The database consists of six emotions, eight speakers (4 female and 4 male) who are speaking five different sentences in emotion class. Speaker's acted voice in different emotion is taken that is not real time inputs. The database is recorded at 44.1 KHz sampling rate with 16 bits for each sample. The angry (40 speech data), joy(40), neutral(40) and sad(40) are considered for experiments. Each 40 speech data consist of 5 different sentences spoken by 8 speakers in that emotional voice.

Performance Evaluation:

The comparison between two methods is shown with confusion matrices. In confusion matrix rows indicate the actual emotion class to be tested and columns indicate the classification done by the system into different categories.

The results are shown in percentage form that is out of 100 actual emotions how much is the result.

Table II Confusion Matrix for Entropy based Emotion Recognition

	Angry (%)	Joy (%)	Neutral (%)	Sad (%)
Angry	62.5	25	12.5	0
Joy	22.5	65	10	2.5
Neutral	7.5	7.5	60	25
Sad	5	7.5	15	72.5

Average Accuracy=65%

Table II shows the confusion matrix for statistical parameter based emotion recognition. The accuracies obtained for emotions anger, joy, neutral and sad are 52.5%, 60%, 67.5%, and 72.5% respectively.

Table III Confusion Matrix for Statistical Features based Emotion Recognition

	Angry (%)	Joy (%)	Neutral (%)	Sad (%)
Angry	52.5	35	7.5	5
Joy	35	60	5	0
Neutral	5	0	67.5	20
Sad	5	0	22.5	72.5

Average Accuracy=63.125%

Table III shows confusion matrix for entropy based emotion recognition system. From table the proposed system is able to recognize emotions viz. angry, joy, neutral and sad with accuracies of 62.5%, 65%, 60% and 72.5% respectively.

V. CONCLUSION

In proposed system speech signal is analyzed using Discrete Wavelet Transform. The non-stationary behavior of speech is mapped efficiently in time and frequency domain. The selection of entropy as a feature instead of large TE operated Wavelet decomposed coefficients greatly minimizes the length of feature vector. Hence with simple Euclidean distance based feature classifier provides improved emotion recognition. Because

of reduced calculations and improved accuracy results, the entropy as a feature outperforms over other statistical features calculated from Wavelet coefficients.

REFERENCES

[1] S. Sultana, C. Shahnaz, S. A. Fattah, I. Ahmmed, W. P. Zhu and M. O. Ahmad "Speech Emotion Recognition based on Entropy of enhanced wavelet coefficients" IEEE 2014, pp 137-140.
[2] E. H. Kim, K. H. Hyun, S. H. Kim, and Y. K. Kwak "Improved Emotion Recognition with a novel speaker independent feature" , IEEE Trans. on Mechatronics vol 14,no 3, pp 317-325, 2009.
[3] C. Shahnaz and S. Sultana "A Feature Extraction scheme based on Enhanced wavelet coefficients for speech Emotion Recognition"IEEE2014, pp1093-1096
[4] Kidae Kim, Dae Hee Yaun and Chulhee Lee "Evaluation of wavelet filters for speech recognition"IEEE2000, pp 2891 2894
[5] Guojun Zhou, John H. L. Hansen and James F. Kaiser "Nonlinear Feature Based Classification of Speech Under Stress" IEEE Transaction on Speech and Audio Process. Vol 9, No.3, pp 201-216
.