19th International Computer Science and Engineering Conference (ICSEC)
Chiang Mai, Thailand, 23-26 November, 2015

Fisher Feature Selection for Emotion Recognition

Piyatragoon Boonthong[1], Pusit Kulkasem[2], Suwanna Rasmequan[3], Annupan Rodtook[4] and Krisana Chinnasarn[5]
[1,2,3,5]Knowledge and Smart Technology Research Lab, Faculty of Informatics, Burapha University, Chonburi, Thailand
[4]Department of Computer Science, Faculty of Science, Ramkhamhaeng University, Bangkok, Thailand

Abstract— **Emotion Recognition is one of the major areas of Affective Computing. Affective Computing is intended to reduce the communication gap between human and machine. Many research works in this area are focus on how to make a machine properly response to different mood of human. In this research, we propose Fisher Feature Selection (*F-Score*) for Emotion Recognition of Thai Speech to classify 4 different emotions from Thai utterance: Sad, Angry, Happy and Fear. The essence of our work lies on the inherit difficulties of Thai Language properties that intonation, articulation and accentuation can made different meanings. The proposed method is divided into two steps: First step, the human sound is extracted to get the 14 dominant features using *F-Score* Feature Selection. Then, in step two, two different learning networks are used to compare the classification performance. The results show that with the use of *F-Score* Feature Selection as a feature selection method to combine with Back Propagation Neural Network as a learning network offers a distinctive recognition rate of 95.13%.**

Keywords:- Affective Computing; Thai Speech Emotion Recognition; Fisher Feature Selection; Machine Learning

I. INTRODUCTION

Affective Computing is very helpful for computer and human interaction. It can help to reduce the gap between human and machine. Mood classification feature makes computer response to the human feeling or recognize the changing of emotion at real-time. Moreover, the efficient system should be able to learn human emotions and provide the proper interactions that can be help to improve the system performance. If there is any computer that can detect human feeling and response with an appropriate solution, it can then increase user friendly and may make people feel enjoyable with the computer system.

The emotion is an important criterion for Affective Computing. The human sound can express human emotion. The characteristics of the sound will be changed when human have difference feelings. The changing of sound and human emotion has some relations. Researches on emotion recognitions are focus on the variation of the sounds as effected by emotion. Researchers in this areas try to desire algorithms that can lead to advance improvement in this aspect of Human-Machine Interactions. In order to perform effective emotion recognition, we need a computer system that could detect the high variance features and then recognize the mood of particular sound with high accuracy performance.

Yun Jin et al. [1] proposed a feature selection and feature fusion combination method for recognizing speaker-independent speech emotion. Signal processing steps proposed in this paper are multiple kernel learning for extracting feature subsets, feature fusion and combination feature fusion. The final recognition rate for 7 kinds of emotions on Berlin Database was 83.10%.

Jun-Seok Park and Soo-Hong Kim [2] proposed a method for recognizing speech emotion using Fractal Dimension Features. The fractal feature is suitable for non-linearity and self-similarity of a speech signal. This approach use Support Vector Machine technique as classification and recognition tools. In their research experiment, a standard database, the Berlin Emotional Speech Database was used as input to measure the effectiveness. The approach provided a recognition rate of 77%.

Dipti D. Joshi and M.B. Zalte [3] proposed system to recognize emotions from Marathi Speech samples which is one of the Indian Languages. The features extracted from the speech samples are the energy, pitch, formants and Mel frequency cepstrum coefficient (MFCC). Discrete wavelet transform was used for feature vector extraction. The SVM classifier was used to differentiate emotions such as anger, happiness, sadness and neutral state.

Muhanmad Sanaullah et al. [4] proposed learning method for classification of stressed speech. They used Bark Band Spectral Energy and Significant Spectral Energy as input features and compare the results with MFCC features. Neural networks with Levenberg-Marquardt algorithm was used as learning tools. Experimental results confirmed that Bark Band Spectral Energy provided better results than significant spectral energy and MFCC.

In this research, we found that Thai utterance recognition is interesting because of the special characteristics of Thai language that consists of, for example, intonation, articulation and accentuation. Because of these special characteristics together with the varieties of Thai utterance speech behavior made the sample space quite large. The sparse space sample made the selection process difficult which led to difficulty of the recognition process. We propose 14 feature extraction methods to work with those special characteristics of Thai utterance and then deploy Fisher Score as a feature selection method. Recognition is based on BPNN and RBF using WEKA machine learning tool. Recognition performances of the extracted and selected features have been evaluated by performance matrix and root mean square error (RMSE).

978-1-4673-7825-3/15/$31.00

The paper is organized as follows. Section II discusses the related work and theoretical background. Section III explains the proposed methods. Section IV shows experimental results and recognition performances. Section V provides conclusion and discussion.

II. MATERIAL AND THEORETICAL BACKGROUND

A. Speech Processing

The nature of speech is established on a varying of time, as a result, they are non-stationary. Normal speech signal processing will include all data points as shown in the following equation of speech energy:

$$E_T = \sum_{n=-\infty}^{\infty} S^2(n) \tag{1}$$

However, as per the conduct of process engineering which has proposed the guideline for speech processing by applying the existing signal in the customizing tool. This tool is based on the assumption that the human speech is normally not change for an estimated window range of 10 - 30 msec. This typical process is called Short Time Processing (STP).

The short time processing of speech can be applied for both time and frequency domains. Typically, length of each segment depends upon domain target. In addition, we can combine different types of STP such that: Short Time energy, Short Time zero crossing rate, and Short Time autocorrelation can be combined to improve performance accuracy of speech signal classification. Furthermore, we can also consolidate the extracted frequency by Short Time Fourier transform with those time domain features. However, each extracted feature represents difference information which will be used for automated processing.

B. Feature Extraction

In this research, we propose Fisher Score (*F-Score*) as a feature selection method for Thai Emotion Recognition. In our experiment, we focus the recognition of emotion using sound wave as input objects. The principal assumption for our work is that once the emotion or temper is changed, the characteristic of sound wave will be altered. Under this assumption, we start our experiment by extracting the feature of sound wave into 14 dimensions as follows:

1) Energy Entropy Block (E_e)

Energy Entropy Block represents the entropy of each sub-block. It can simply be found by calculating from the sound signal. To calculate this, the input speech signals are divided into *f* number of frames. Then, the energy for each frame will be normalized before finding the Entropy for each sub-block using the following equation:

$$E_e = -\sum_{k=0}^{f-1} \mu^2 * \log_2(\mu^2) \tag{2}$$

$$\mu^2 = \sum_{b=1}^{N} \frac{(\mathrm{N}*\frac{W_1}{S_b})}{F} \tag{3}$$

Where μ^2 - is the normalized energy
N - is the total number of blocks
W_1 - is the window length
S_b - is the number of short blocks
F - is the frequency

By exploiting the energy entropy equation that is given in Eq.1, the energy entropy feature or E_e will be derived.

2) Short-time Energy *(S_e)*

Short-time energy is defined as the sum of squares of the samples in a frame. To calculate S_e, the input signal will be divided into w number of windows and then windowing function for each window will be calculated. S_e can be help distinguished between voiced and unvoiced speech segments. This is due to the fact that unvoiced speech has significantly small value of short-time energy. S_e is calculated by the following equation:

$$S_e = \sum_{i=-\infty}^{\infty} x(i)^2 * h(w-i) \tag{4}$$

Where $x(i)$ - is the input signal
$h(w)$ - is the impulse response

3) Zero-crossing Rate (zcr)

Zero-crossing Rate represents the rate of sign-changes along a signal line. That is the rate of the signal that changes from positive to negative or vice versa. This feature is often used in speech recognition and it is also very suitable for emotion recognition because it can help to express those hidden information. The hidden information may include speech rate, speed of changing tempo that vary from emotion to emotion. zcr is defined in the following form:

$$zcr = \frac{1}{T-1} \sum_{t=1}^{T-1} \Pi\{s_t s_{t-1} < 0\} \tag{5}$$

Where s is a signal of length T and the indicator function Π is 1 if its argument is true and 0 otherwise.

4) Spectral-Roll-Off

Spectral-Roll-Off represents the amount of the right-skewedness of the power spectrum. Right-skewedness or higher frequencies will have higher spectral-roll-off values. It is defined as the second frequency bin $M_c^R(J)$ below which the c percent (e.g. $c = 90$) of the magnitude distribution of the DFT X_R coefficient is concentrated for frame

$$\sum_{k=0}^{M_c^R(j)} |X_{jk}| = \frac{c}{100} \sum_{k=0}^{s-1} |X_{jk}| \tag{6}$$

5) Spectral Centroid

Spectral Centroid represents the characteristic of a spectrum. It is obtained by evaluating the “center of gravity” using the Fourier Transform’s Frequency, with their magnitudes as the weights. The individual centroid of a spectral frame is defined as the average frequency weighted by amplitudes and divided by the sum of the amplitudes as shown in the following equation

$$SpectralCentroid = \frac{\sum_{k=1}^{N} kF[k]}{\sum_{k=1}^{N} F[k]} \quad (7)$$

where $F[k]$ is the amplitude corresponding to bin k in DFT spectrum.

6) Fundamental Frequency (F_0)

Fundamental Frequency represents the frequency of the vocal cords vibration of individual voice signal. It has been used in a variety of areas. This is due to its nature of high computational time complexity and resource consumption, it has been abandoned. However, with the state of art of numerical method, we can estimate F_0 by deploying Zero-crossing method.

7) Mel Frequency Cepstral Coefficient *($MFCCs$)*

Mel Frequency Cepstral Coefficients ($MFCCs$) represents the short-term power spectrum of a sound. It is often used in various applications of speech processing. $MFCCs$ is based on human hearing capability which cannot easily perceive frequencies that over 1 KHz. The calculation of $MFCCs$ can be divided into the following steps:

1. Compute window function.
2. Compute power spectrum by Fast Fourier Transform (FFT)
3. Compute Mel-filter bank.
4. Compute Discrete Cosine Transform (DCT).
5. The $MFCCs$ are the amplitudes of the resulting spectrum.

$$MFCC_i = \sum_{k=1}^{N} X_k \cos\left[i(k - \frac{1}{2}) \frac{\pi}{N} \right], i = 1, 2, ..., M \quad (8)$$

where M is the number of cepstrum coefficients, X_k, $k = 1, 2, ..., N$ represents the log-energy output of the k^{th} filter, and N is the number of triangular band pass filters (with standard value of around 20)

8) Linear Predictive Coding (LPC)

Linear Predictive Coding (LPC) is a source-filter analysis-synthesis methodology. Which the vocal tract transfer function is modeled by an all-pole filter with a transfer function shown in Eq. (9)

$$H(z) = \frac{1}{1 - \sum_{i}^{p} a_i z^{-i}} \quad (9)$$

LPC is based on the source-filter model, where a_i is the filter coefficients. The speech signal s_n assumed to be fixed over the analysis frame which is approximated as a linear combination of the past p samples shown in Eq. (10)

$$\hat{s}_n = \sum_{i=1}^{p} a_i s_{n-i} \quad (10)$$

In Eq. (10) a_i can be found by minimizing the mean square filter prediction error between $\hat{s}_n$ and s_n

9) Formant Frequencies (F)

Formant Frequencies (F) represent the spectral peaks of the sound spectrum. Formant Frequencies are also used to mean an acoustic resonance of the human vocal tract. It shows the differentiation of human speech frequencies. A single speech consists of more than one formant frequencies: $F1, F2, F3$ and etc. Most often the first two formants, $F1$ and $F2$, are enough to disambiguate the vowel in speech processing. Formant Frequencies (F) are calculated by following equation:

$$F = \frac{F_s}{2\pi} \arctan \frac{im(s)}{re(s)} \quad (11)$$

where F_s, $im(s)$ and $re(s)$ represent the sampling frequency, imaginary and the real part of sound signal s, respectively.

10) Perceptual Linear Predictive (PLP)

Perceptual Linear Prediction (PLP) model represents the human speech based on the three concepts of psychophysics of hearing. They are 1) the critical-band spectral resolution, 2) the equal-loudness curve and 3) the intensity-loudness power law. PLP is used to find the all-pole model of the auditory-like short-time spectrum of speech. The PLP coefficients can be derived from the filter bank coefficients by applying the following steps

1. Pre-emphasis with the simulated equal-loudness curve & amplitude compression
2. Equal loudness curve corresponding to the frequency f_k for the k^{th} filter bank given by

$$L_k = \left(\frac{f_k^2}{f_k^2 + 1.6e^5} \right)^2 \left(\frac{f_k^2 + 1.44e^6}{f_k^2 + 9.61e^6} \right) \quad (12)$$

3. Apply equal-loudness curve & amplitude compression: $\hat{m}_k = (L_k M_k)^{\beta}, (\beta = \text{comp} \cdot \text{factor})$
   Perform inverse DFT and Linear Prediction to obtain LP coefficient
4. Applying inverse DFT to filter pre-emphasized bank yields auto-correlation coefficients
5. Apply Durbin algorithm to obtain LP coefficients
6. Convert LP coeff.(a_i) to cepstral coef (c_n)

$$c_n = -(\mathrm{a}_n + \frac{1}{n} \sum_{i=1}^{n-1} (n-i) \mathrm{a}_i \mathrm{c}_{n-i}) \quad (13)$$

11) Harmonic Product Spectrum (HPS)

Harmonic Product Spectrum represents the geometric mean of amplitudes of the harmonics associated with a particular frequency in the spectrum. The amplitude spectrum of voiced sounds consists of a series of peaks and has sharp peaks that occur at integer multiples of the fundamental frequency, where pitch represents the perceived fundamental frequency of a sound. When the spectrum is compressed, the distance between those peaks becomes shorter. The Harmonic

Product Spectrum $P(n)$ is the product of R frequency-shrunken replicas of the amplitude spectrum $\left| X(e^{j\frac{2\pi}{N}n}) \right|$

$$P(n) = \sqrt[R]{\prod_{r=1}^{R} \left| X(\mathrm{e}^{j\frac{2\pi}{N}nr}) \right|} \tag{14}$$

where N is the number of FFT points and $R = \left[N/2n \right]$ is the maximum shrinkage of amplitude spectrum which can still provide an amplitude value on discrete frequency n .

12) Autocorrelation (AC)

Autocorrelation is popular for its pole retaining and noise separation properties. $R(t)$ is used to find periodicity of a time frame of length T in the time domain. It expresses the similarity between the signal $x(t)$ and its copy shifted by t

$$R(t) = \frac{1}{T-t} \sum_{\tau=0}^{T-t-1} x(\tau) x(\tau + t) \tag{15}$$

AC achieves its maximum value at $t = 0$ (i.e. $|R(t)| \le R(0)$ for all t)

13) Spectral Flux (SF)

Spectral Flux or Spectral Difference Analysis represents the spectral change of two successive frames. *SF* is calculated by finding the difference between the current value of each magnitude spectrum bin in the current window from the corresponding value of the magnitude spectrum of the previous window. It is also known as the Euclidean distance between the two normalized spectra as shown in the following equation.

$$SF_i = \sum_{k=1}^{N/2} \left(|X_i(k)| - |X_i(k-1)| \right)^2 \tag{16}$$

14) Harmonic Ratio (HR)

Harmonic Ratio represents the proportion of harmonic components in the power spectrum. The extraction for every frame is standardized in the following way. That is the maximum value of the normalized autocorrelation function is computed for all frame. If the signal is purely periodic, its peak values will be lags M (m is the lag index of the autocorrelation). The maximum lag M corresponds to the minimum fundamental frequency that can be estimated

$$M = \frac{F_s}{f_0^{\min}} \tag{17}$$

harmonic ratio is calculated by following

$$HR = \max_{M_0 \le m \le M} \{\mathrm{T_i(m)}\} \tag{18}$$

HR is close to one for harmonic signals and zero for white noise.

C. Feature Selection: Fisher Score (*F-Score*)

The *F-Score* is a method for determining the most relevant features for classification. It uses discriminative methods, and generative statistical models to accomplish this. *F-Score* is under class labels control and it pursues best distribute ability features.

Let n_i denote the number of samples in class i . Let μ_r^i and $(\sigma_r^i)^2$ be the mean and variance of class i , $i = 1,...,c$, corresponding to the $r-th$ feature. The Fisher score of the $r-th$ feature F_r , which should be maximized, is computed as follows

$$F_r = \frac{\sum_{i=1}^{c} n_i (\mu_r^i - \mu_r)^2}{\sum_{i=1}^{c} n_i (\sigma_r^i)^2} \tag{19}$$

III. Proposed Methods

Thai Speech Emotion Recognition System proposed in this paper consists of three main steps: 1) pre-processing, 2) feature extraction and selection, and 3) classification. Dataset uses in this research is taken from Speech Corpus can be found in [8]. There are four main emotions: Sad, Angry, Happy and Fear. Overall Thai emotion samples used in this research consist of 800 samples. Figure I illustrates 4 emotions of sample “โดยเฉพาะอย่างยิ่ง” or “especially”. It can be seen that they all 4 signals are similar. Hence, it is difficult to classify one from the others.

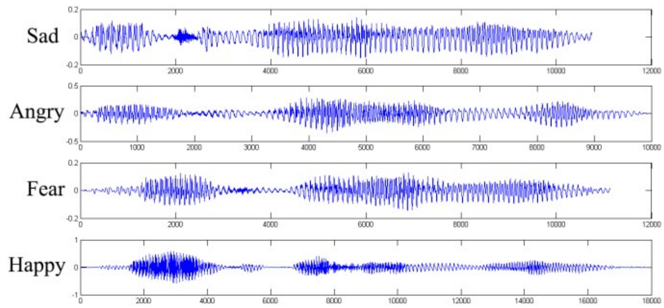

Figure I: Sound wave of 4 emotions of word “โดยเฉพาะอย่างยิ่ง”

A. Preprocessing

We start the proposed method by applying Pre-emphasis Filtering Algorithm with coefficient = 0.9375. Each source sample is separated into windows or frame which have the frame size of 30 msec with 15 msec overlapping. Then, for each window, we apply hamming window to reduce signal discontinuity in order to avoid spectral leakage.

B. Feature Extraction and Selection

In this research, we propose Fisher Feature to select features that will provide the optimal result. In our experiment, we focus only on the recognition of Thai emotion by using sound wave as input objects. The basic assumption is that once the emotion or temper is changed, the characteristic of sound wave will be altered. Under this assumption, we start our experiment by extracting sound features into 14 dimensions as mentioned in Section II. For each dimension 5 statistical values consisting of *Mean, Median, Max, Min,* and *Variance* were calculated. This led to the total number 70 data points or 70 input vectors for each sample. We then use Fisher Score Feature Selection or *F-Score* to select optimal features in the learning process. As described in Section II, *F-Score* is statistical tools for evaluating and discriminating one data point from others. The main characteristic of *F-Score* is that it can differentiate member of each individual group by trying to avoid the overlapping member as much as possible. From our experiment, 22 data points of 7 features had been seleted as an optimal features for classifying process.

C. Classificaion

In the classification stage, we compare the performance metrics with two classification methods namely Back-propagation Neural Networks (BPNN) and Radial-Basis Function (RBF) by WEKA machine learning tools. 10 folds cross validation had been used for training and testing. Root mean square errors (RMSE) were used to evaluate the performance of Thai speech emotion recognition. Optimal learning parameters of each classification method are defined as follows:

- BPNN (Back-propagation Neural Networks)
  Learning rate 0.1
  Momentum 0.1
  Number of epoch 500
  Number of Hidden layers 12

- RBF (Radial-Basis Function)
  Gaussian function
  clusteringDees 1
  maxIts -1
  minStdDev 0.1
  numCluster 2
  ridge 1.00E-08

IV. Experimental Results

The experimental results using Thai Speech dataset from [8]. We applied Pre-emphasis Filtering Algorithm with coefficient = 0.9375 as our pre-processing process. Table I and II show the classification performance matrix or accuracy rate of those 70 original data points of 14 features based on BPNN and RBF leaning algorithm respectively. The average accuracy rate are 94.50 for BPNN and 87.63 for RBF. Table III and IV illustrate the results of performance matrix of 22 data points of 7 features which were selected by *F-Score* feature selection algorithms based on BPNN and RBF learning algorithm respectively. Those selected data points are consisted of data points from following features: Energy Entropy Block, Short-time energy, Zero-crossing rate, Spectral-roll-off, Spectral Centroid, Formant Frequencies, and MFCC. The average accuracy rate are 95.13 for BPNN and 92.38 for RBF.

From experiment results shown in Table I – V clearly demonstrate that Fisher Feature selection is an optimal solution for Thai emotional feature selection. In addition, it can be seen that the results from Table III BPNN performance matrix of Thai Emotional Dataset based on 22 data points of 7 features selected by *F-Score* provides the best result. It means that 22 data points from 7 features are the optimal input vector for Thai Emotional learning process.

TABLE I. BPNN PERFORMANCE MATRIX OF THAI EMOTIONAL DATASET BASED ON 70 DATA POINTS OF 14 FEATURES

Emotion (%)	Sad	Angry	Fear	Happy
Sad	92.50	2.50	4.50	0.50
Angry	3.00	96.00	0.00	1.00
Fear	6.00	0.00	93.50	0.50
Happy	0.50	2.50	1.00	96.00

TABLE II. RBF PERFORMANCE MATRIX OF THAI EMOTIONAL DATASET BASED ON 70 DATA POINTS OF 14 FEATURES

Emotion (%)	Sad	Angry	Fear	Happy
Sad	84.50	1.00	14.50	0.00
Angry	8.00	89.50	0.50	2.00
Fear	9.00	1.00	88.50	1.50
Happy	1.50	5.00	5.50	88.00

TABLE III. BPNN PERFORMANCE MATRIX OF THAI EMOTIONAL DATASET BASED ON 22 DATA POINTS OF 7 FEATURES SELECTED BY *F-SCORE*

Emotion (%)	Sad	Angry	Fear	Happy
Sad	**92.50**	1.50	6.00	0.00
Angry	3.50	**96.00**	0.50	0.00
Fear	5.00	0.00	**95.00**	0.00
Happy	0.00	2.50	0.50	**97.00**

TABLE IV. RBF PERFORMANCE MATRIX OF THAI EMOTIONAL DATASET BASED ON 22 DATA POINTS OF 7 FEATURES SELECTED BY *F-SCORE*

Emotion (%)	Sad	Angry	Fear	Happy
Sad	**91.00**	0.50	8.50	0.00
Angry	5.50	**93.50**	0.50	0.50
Fear	7.00	0.00	**91.50**	1.50
Happy	0.50	3.00	3.00	**93.50**

Table V shows the performance evaluation based on Root Mean Square Error (RMSE) of the BPNN of value 0.1413 for selected data points reconfirms that the proposed method of *F-Score* feature selection can offer the optimal features for Thai speech emotion classification.

TABLE V. PERFORMANCE EVALUATION (RMSE)

Accuracy Rate (%)	**14 Features (70 data points)**	**7 Features (22 data points) (based on F-score)**
BPNN	0.1525	**0.1413**
RBF	0.2317	0.1854

V. DISCUSSIONS AND CONCLUSIONS

This research presents *F-Score* as an algorithm for selecting optimal features of Thai Speech Emotion Recognition. This is due to the characteristic of *F-Score* method that can distinctly distinguish the different among groups of speech. Hence it can help to reduce those redundant features. It clearly demonstrated that almost 70% of the unnecessary data points were omitted from the original training dataset. This figure enormously makes a huge reduction of computing time. From the experiment results, there are only 7 features that were selected from 14 features. They are: Energy Entropy Block, Short-time energy, Zero-crossing rate, Spectral-roll-off, Spectral Centroid, Formant Frequencies, and MFCC. These features are proofed to be suitable for distinguish the difference of Thai Speech Emotion with a high recognition rate. That is for the proposed method has RMSE rate of 0.1413 as compare to 0.1525 of unselected features learning with BPNN and has RMSE of 0.1854 as compare to 0.2317 of unselected features learning with RBF. And the most suitable learning algorithm for Thai Speech Emotion Recognition is BPNN.

ACKNOWLEDGEMENT

This research was supported by the National Research Council of Thailand grant to the Burapha University, year 2013-2014.

REFERENCES

[1] Yun Jin, Peng Song, Wenming Zheng and Li Zhao, “A Feature Selection and Extract Fusion Combination Method for Speaker-Independent Speech Emotion Recognition”, in *the Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, Italy, May 4-9, 2014, pp. 4841-4845.

[2] Jun-Seok Park and Soo-Hong Kim, "Emotion Recognition from Speech Signals using Fractal Features", *International Journal of Software Engineering and Its Applications*, Vol.8, No.5 (2014), pp.15-22.

[3] Dipti D. Joshi and M.B. Zalte, “Recognition of Emotion from Marathi Speech using MFCC and DWT algorithms”, *International Journal of Advanced Computer Engineering and Communication Technology (IJACECT)*, Vol.2, Issue 2013.

[4] Muhammad Sanaullah and Masud H. Chowdhury, “Neural network based classification of stressed speech using nonlinear spectral and cepstral features”, in *the proceedings of 2014 IEEE 12th International New Circuits and Systems Conference (NEWCAS)*, pp.33-36.

[5] S. Sultana, C. Shahnaz, S.A. Fattah, I. Ahmmed and W.-P. Zhu and M.O. Ahmad, “Speech Emotion Recognition Based on Entropy of Enhanced Wavelet Coefficients”, in *the proceedings of 2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp.137-140.

[6] Namrata Dave, “Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition”. *International journal for advance research in engineering and technology*, July 2013, Volume 1 Issue 6 pp.1-4.

[7] Amita Dev and Poonam Bansal, “Robust Features for Noisy Speech Recognition using MFCC Computation from Magnitude Spectrum of Higher Order Autocorrelation Coefficients”, *International Journal of Computer Applications*, Vol. 10, No. 8, Nov. 2010, pp.36-38.

[8] Thapanee Seehapoch and Sartra Wongthanavasu, "Speech emotion recognition using Support Vector Machines", in *the Proceedings of 2013 5th International Conference on Knowledge and Smart Technology (KST)*, Jan 31 - Feb 1, 2013. pp. 86-91.