

# Emotion Recognition from Speech Signals by Using Evolutionary Algorithm and Empirical Mode Decomposition

Shing-Tai Pan  
Department of Computer  
Science and Information  
Engineering, National  
University of Kaohsiung,  
Kaohsiung 811,  
Taiwan, R.O.C.  
span@nuk.edu.tw

Chih-Hung Wu  
Department of Electrical  
Engineering, National  
University of Kaohsiung,  
Kaohsiung 811,  
Taiwan, R.O.C.  
johnw@nuk.edu.tw

Chen-Sen Ouyang  
Department of Information  
Engineering,  
I-Shou University,  
No.1, Sec. 1, Syuecheng  
Rd., Dashu District,  
Kaohsiung City 84001,  
Taiwan, R.O.C.  
ouyangcs@isu.edu.tw

Ying-Wei Lee  
Department of Computer  
Science and Information  
Engineering, National  
University of Kaohsiung,  
Kaohsiung 811,  
Taiwan, R.O.C.  
chasewind\_5@hotmail.com

**Emotion status impacts massively the human's health and the job performance of people. A system that can continuously and automatically monitor people's emotion is worthy of development. Besides, the speech signals always contain some emotion features and are the most commonly used for communication between humans. The exploration of emotion recognition from speeches then becomes more important. In this paper, we propose a strategy for emotion recognition from speech by combining evolutionary algorithm (EA) with Empirical Mode Decomposition (EMD) to improve the emotion recognition rate. First, some emotional speeches were decomposed into several Intrinsic Mode Functions (IMFs) by using EMD process. The emotional part of a speech is then extracted by using these IMFs. In this paper, some weighted IMFs obtained from EMD are combined for the following recognition process. Hence, it is one of the goals of this paper to find the optimal weights corresponding to each IMF and to combine these weighted IMFs to make the recognition results as accurate as possible. The weights for each IMF are trained by evolutionary algorithm to find an optimal combination of IMFs. The reason why evolutionary algorithm is used here is that evolutionary algorithm always obtains some outstanding performances in many research concerning optimal design. The Mel-Frequency Cepstral Coefficients (MFCCs) are then computed and are used as the features for emotion recognition. An open database, eNTERFACE 2005 emotion database, is adopted in this paper as training and testing data for the experiments.**

*Emotion recognition. Speech signals. Empirical mode decomposition. Genetic algorithm.*

## 1. INTRODUCTION

It's important to detect the people's emotion in their daily lives. Emotion status impacts massively the human's health as well as the quality of life. Emotion of people also affects the job performance. Hence, a system that can continuously and automatically monitor people's emotion is worthy of development. Besides, the speech signals always contain some emotion features and are the mostly used for communication between humans. The exploration of emotion recognition from speeches then becomes more important. Although this topic is a new research area, there are already some

literatures about this topic, see for example the papers (Schuller, Rigoll & Lang 2003, Fragopanagos & Taylor 2005, Cen et al. 2010, Wu, Falk & Chan 2011, Petrusihin 2000, Tato et al. 2002, Yacoub et al. 2003, Schuller, Rigoll & Lang 2004, New, Foo & Silva 2003) and the references therein. However, in the existing literatures, the emotion recognition rate from speech is low and is far away from practical applications. Hence, how to improve the recognition rate is an open problem. The studies on the topic of emotional speech recognition are consequently innovative.

On the other hand, EMD was first proposed by Prof. Huang in combination with the Hilbert Transform (HT) to analyse nonlinear and non-stationary time series. The combination of EMD and HT is therefore referred to as the Hilbert Huang Transform (HHT) (Huang 1996). The advantage of EMD over other frequency-domain transformations is that the components decomposed from a mixed signal are related to specific physical sources. This allows us to examine the physical phenomena of a signal through the components obtained by EMD. Initially, EMD was applied to signal analysis in the field of geoscience, strength analysis of material structures, trend analysis of the stock market, etc. More recently, EMD has been applied in the measurement and enhancement of speech signals (Gales & Young 1995, Tsao & Lee 2009), biomedical signal analysis (Windmann & Haeb-Umbach 2009), earthquake signal analysis (Huang 1996) and speech feature capture (Li et al. 2008), etc. For speech feature capture, the paper (Wang, Li & Zhang 2006) uses first EMD to decompose the speech signal into several IMFs, and then uses Renyi Entropy to calculate each IMF's energy. The features are then obtained from these energies. Due to the impressed performance of EMD on features extraction, this paper will apply EMD on the derivation of features for emotional components from some speeches.

Besides, the strategy of emotion features extraction in this paper is to combine some weighted IMFs obtained from EMD process and then compute the emotion features for the resulting signals by using MFCC. So, the design of the weights for each IMF becomes the key for feature extraction of emotion. It is one of the goals of this paper to find the optimal weights corresponding to each IMF. Recently, since the computation ability of computers has been enormously enhanced, evolutionary algorithm computation is widely applied in many areas of engineering; for a recent review of these applications, please see e.g. (Uesaka & Kawamata 2003). The most popular evolutionary algorithm is genetic algorithm (GA). GA is based on the Darwinian "Survival of the Fittest" strategy. Each individual in the population represents a potential solution to the problem at hand. They compete and mate with each other in order to produce stronger individuals. Much recent research has pointed out that GA is efficient in solving optimization problems (Haupt & Haupt 1998, Karamalis, Kanatas & Constantinou 2009, Doong, Lai & Wu 2007, Pan & Lai 2008, Lai & Chang 2009, Tsai, Huang & Chan 2011). Hence, in this paper, weights for each IMF are trained by GA to find an optimal weighted combination of IMFs. This will increase the recognition precision. The proposed method is then shown in Figure 1.

## 2. EMPIRICAL MODE DECOMPOSITION (EMD)

This study applies EMD to decompose an emotional speech signal into emotional parts and non-emotional parts. In this section, the procedure for performing EMD is introduced. The purpose of EMD is to shift the original non-stationary data series until the final data series is stationary. The main step to perform EMD operation is to divide a speech signal into several intrinsic mode functions (IMFs). The condition for IMFs on a data series can be described as follows (Huang 1996).

1. In the whole data series, the number of local extremes and the number of zero crossings must be equal or differ at most by one.
2. At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

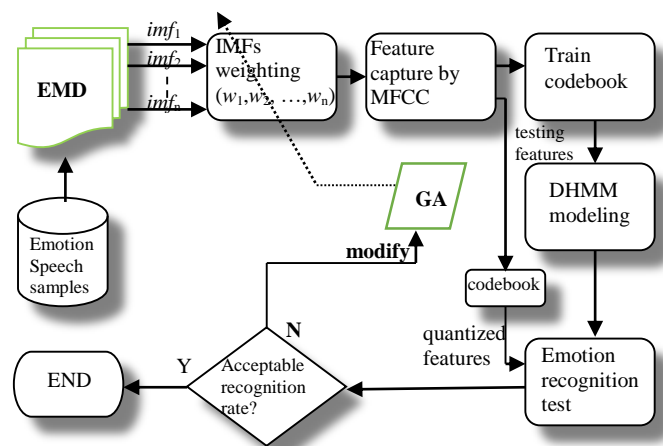


Figure 1: Proposed strategy for emotion recognition from speech.

The procedure of EMD for a data series or a signal is then introduced as follows. The Cubic Spline (Mathews & Fink 2004) is used to generate the upper envelop and lower envelop of the signals during the process of finding IMFs. Suppose that the original signal is  $X(t)$  and  $Temp(t) = X(t)$ .

### Step 1:

Find the upper envelop  $U(t)$  and lower envelop  $L(t)$  of the signal  $Temp(t)$ . Calculate the mean of the two envelops  $m(t) = [U(t) + L(t)]/2$ . The component of  $Temp(t)$  is obtained by the equation:

$$h(t) = Temp(t) - m(t). \quad (2.1)$$

### Step 2:

Check whether the signal  $h(t)$  satisfies the conditions of IMF or not. If it does, then we obtain the first IMF as  $imf_1(t) = h(t)$  and go to next step, else we assign the signal  $h(t)$  to  $Temp(t)$  and go to Step 1.

**Step 3:**

Calculate the residue  $r_1(t)$  as

$$r_1(t) = Temp(t) - imf_1(t). \quad (2.2)$$

Assign the signal  $r_1(t)$  to  $X(t)$  and repeat Step 1 and Step 2 to find  $imf_2(t)$ .

**Step 4:**

Repeat Step 3 to find the subsequent IMFs as follows.

$$r_n(t) = r_{n-1}(t) - imf_n(t), \quad n = 2, 3, 4, \dots \quad (2.3)$$

This step comes to an end when the signal  $r_n(t)$  is constant or is a monotone function.

After that the EMD procedure Step 1~ Step 4 is finished, the following decomposition of  $X(t)$  is obtained.

$$X(t) = \sum_{i=1}^n imf_i(t) + r_n(t) \quad (2.4)$$

However, it is hard to satisfy the second condition of IMF in practical application, since zero mean value of the envelopes for all time  $t$  is almost impossible. Hence, a looser condition is lunched to replace the second condition of IMF. To construct this looser condition, an index for the mean value of the envelopes is computed and a threshold for this index is assigned. In general, the threshold is assigned in the range 0.2 to 0.3. Moreover, the index can be calculated through the following equation (Huang 1996).

$$SD_{ik} = \frac{\sum_{t=0}^T |h_{i(k-1)}(t) - h_{i(k)}(t)|^2}{\sum_{t=0}^T h_{i(k-1)}^2(t)} \quad (2.5)$$

in which  $h_{i(k)}(t)$  is the  $k$ th iteration for  $i$ th IMF. It is noted that the function  $h_{i(k-1)}(t) - h_{i(k)}(t)$  in numerator of eq. (2.5) is equal to the mean  $m_{i(k)}(t)$  i.e.,  $h_{i(k)}(t) = h_{i(k-1)}(t) - m_{i(k)}(t)$ . This means that  $SD_{ik}$  is the ratio of the energy of  $m_{i(k)}(t)$  to that of  $h_{i(k-1)}(t)$ . In practical applications, the second condition of IMF is then replaced by the looser condition: the  $SD_{ik}$  should be smaller than the assigned threshold. That is, if  $SD_{ik} < 0.3$ , for example, and the second condition of the IMF is satisfied, then the iteration for  $i$ th IMF ends and we get a new IMF.

**3. DISCRETE HIDDEN MARKOV MODEL**

The HMM model is used for emotion recognition from speeches in this paper. According to the type of the probability distributions used in HMMs, HMMs can be categorised into Continuous Hidden Markov Model (CHMM) and Discrete Hidden Markov Model (DHMM). The DHMM provides more stable recognition results and faster training with

almost the same level of recognition accuracy as CHMM. Therefore, the DHMM is adopted for emotion recognition in this paper. Some useful features of the emotion recognition are selected to train the DHMM. The built HMM model can become a reliable computer-assisted tool for monitoring peoples emotion in the future.

Figure 2 illustrates the training process of DHMM model. First, the matrices A, B and  $\pi$  which describe the DHMM model are randomized at the initial step. Then, the features obtained from the EMD on emotional speeches are quantized through a trained codebook. The quantized features are used for the observation data of the DHMM model. The corresponding probability for every observation can be found from the elements in the matrices A, B and  $\pi$ . By using these probabilities, we can run the Viterbi Algorithm (Blunsom 2004) to update the matrix A, B and  $\pi$  until the elements in these matrix converge.

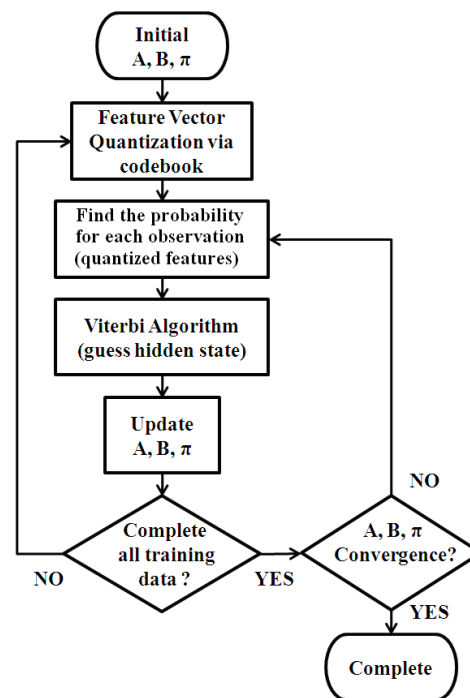


Figure 2: The DHMM model training process.

In a DHMM, the hidden states are unobservable while the outputs (i.e., the observation) of each state are observable. Each hidden state has a probability distribution over the possible output tokens. Therefore, the sequence of output tokens generated by the DHMM gives some information about the sequence of states. For the purpose of clarification, a figure for the relations between the features, the observations and the hidden states of DHMM is depicted in Figure 3. In the following paragraph, the definition and the detail training method of the parameters in DHMM are introduced

(Blunsom 2004). First, the definition of the parameters in DHMM is introduced as follows.

$\lambda$  : DHMM model,  $\lambda = (A, B, \pi)$

$A$ :  $A = [a_{ij}]$ ,  $a_{ij}$  is the probability of that the state  $x_i$  transfers to the state  $x_j$ ,  $a_{ij} = P(q_t = x_j | q_{t-1} = x_i)$

$B$ :  $B = [b_j(k)]$ ,  $b_j(k)$  is the probability of  $k$ th observation which is observed at the state  $x_j$ , i.e.,  $b_j(k) = P(o_t = v_k | q_t = x_j)$

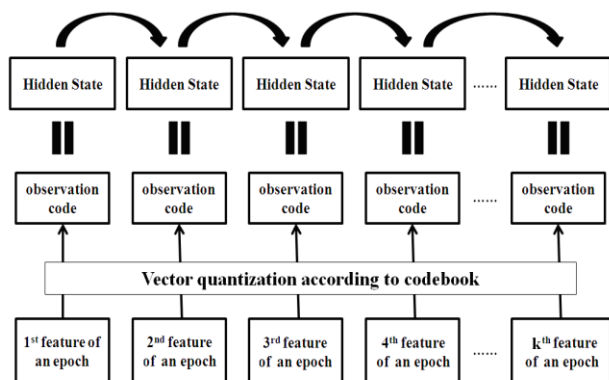
$\pi$ :  $\pi = [\pi_i]$ ,  $\pi_i$  is the probability of the case that the initial state is  $x_i$ , i.e.  $\pi_i = P(q_1 = x_i)$

$X$ : the state vectors of DHMM,  $X = (x_1, x_2, \dots, x_N)$

$V$ : the observation event vector of DHMM,  $V = (v_1, v_2, \dots, v_M)$

$O$ : the observation results of DHMM,  $O = o_1, o_2, \dots, o_T$

$Q$ : the inferred hidden states of DHMM,  $Q = q_1, q_2, \dots, q_T$



**Figure 3:** Relation between the features, the observation and the hidden states of DHMM.

Before introducing the training method for the DHMM model  $\lambda = (A, B, \pi)$ , some notations are defined for convenience as follows:

$E_{ij}$  : the events of the transition from state  $x_i$  to state  $x_j$

$E_{i\bullet}$  : the events of the transition from state  $x_i$  to other states

$E_{\bullet j}$  : the events of the transition from other states to state  $x_j$

$E_{hi}$  : the events of state  $x_i$  appears at initial state

$n(E_{ij})$  : the number of the events of the transition from state  $x_i$  to state  $x_j$

$n(E_{i\bullet})$  : the number of the events of the transition from state  $x_i$  to other states

$n(E_{\bullet j})$  : the number of the transition from other states to state  $x_j$

$n(E_{\bullet j, o = v_k})$  : the number of the events of the transition from other states to state  $x_j$  at which the observation code is  $v_k$

$n(E_{hi})$  : the number of the events of state  $x_i$  appears at initial state

In the process of training the matrix  $A, B$ , and  $\pi$  of DHMM, the hidden states for each observation are estimated first through the initial  $A, B$ , and  $\pi$  by using Viterbi Algorithm. Then the values  $n(E_{ij})$ ,  $n(E_{i\bullet})$ ,  $n(E_{\bullet j})$ ,  $n(E_{\bullet j, o = v_k})$ ,  $n(E_{hi})$  are computed according to the estimated hidden states for the whole training data. Subsequently, the elements in matrices  $A, B$ , and  $\pi$  are updated as follows.

$$\bar{a}_{ij} = \frac{n(E_{ij})}{n(E_{i\bullet})}, \quad (2.1)$$

$$\bar{b}_j(k) = \frac{n(E_{\bullet j, o = v_k})}{n(E_{\bullet j})}, \quad (2.2)$$

$$\bar{\pi}_i = \frac{n(E_{hi})}{n_{TD}}, \quad (2.3)$$

where  $n_{TD}$  is the number of training data. By using these updated  $A, B$ , and  $\pi$ , we run the Viterbi Algorithm again. The above steps are repeated until the matrices  $A, B$ , and  $\pi$  converge. The training process for a DHMM is then completed. During the recognition process, the probability of the observations is calculated based on the model  $\lambda = (A, B, \pi)$  by using the following equation (2.4) (Tsao & Lee 2009):

$$P(O | \lambda) = \sum_Q P(O | Q, \lambda) P(Q | \lambda) \\ = \sum_{q_1 \dots q_T} \pi_{q_1} b_{q_1}(o_1) \cdot a_{q_1 q_2} b_{q_2}(o_2) \dots a_{q_{T-1} q_T} b_{q_T}(o_T). \quad (2.4)$$

This equation enables us to evaluate the probability of the observations  $O$  based on the DHMM model  $\lambda = (A, B, \pi)$ . It is notable that the time taken to evaluate  $P(O | \lambda)$  directly would be an exponential function of the observation number  $T$ . For this reason, the Forward Algorithm (Tsao & Lee 2009) is applied to reduce the computation time.

#### 4. GENETIC ALGORITHM

In order to explain clearly the strategy proposed in this paper, this section introduces briefly the genetic algorithm (GA). It is well known that GA solves optimization problems through the evolution processes for chromosomes in each generation. The evolution processes are selection, crossover, and mutation. It is hence the key point to define the chromosomes in GA for the problems in hand. A chromosome is organised by multiple genes. The number of genes in a chromosome is determined

by the number of the parameters, which are going to be designed in a problem. In general, a GA with longer chromosomes needs fewer generations for evolution to obtain an optimal solution. However, it will take more computing time for each evolution. While a GA with shorter chromosomes will take less computing time for each evolution. But, the disadvantage is that a local optimum may be obtained. Consequently, the number of genes in a chromosome should be carefully determined. Moreover, there are three ways for the encoding of genes: binary encoding, real number encoding, and symbol encoding. The way that the genes are encoded is decided according to the parameters type in a problem. Figure 4 shows the three different ways for encoding genes in a chromosome.

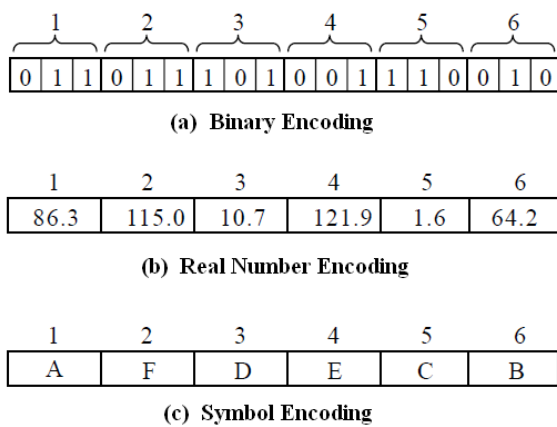


Figure 4: Three encoding ways: (a) binary encoding (b) real number encoding (c) symbol encoding (Lin 2005).

In this paper, a chromosome is made up of twelve features which are calculated from MFCC for emotional speeches after EMD process. And real number encoding is used here, since all the features calculated by MFCC are real.

A GA is applied in this paper to find the optimal weights for the IMF obtained from the EMD process on emotional speeches. Each chromosome is encoded here as a vector of weights  $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]$ , in which the index  $i$  means the  $i$ th generation and  $n$  means the number of genes in a chromosome. The process of GA starts from randomizing the chromosomes in the first generation. Then, the fitness function is calculated for each chromosome. In this paper, the fitness function is defined simply as the emotion recognition rate. According to "Survival of the Fittest", some chromosomes with better fitness values are selected based on Roulette Wheel Selection for evolving the next generation. Then, a linear crossover is used for crossover of the selected chromosomes. The mutation process is performed after the crossover process. In the mutation process, an offset for mutation probability is used. A random variable is applied to generate a

random value for determining whether the mutation will be performed or not. If the random value for a gene is larger than the offset, the gene mutates. The flowchart in Figure 5 reveals the procedure of GA in this paper.

## 5. ALGORITHM FOR PROPOSED METHOD

In order to make the illustration of the proposed strategy clearer, an algorithm is introduced as follows.

1. Find the IMFs for each training emotion speech by using EMD method.
2. Find the optimal weight for each IMF by using GA.
3. Add every weighted IMF to obtain a new speech signals.
4. Perform MFCC on the speech signals obtained in Step 3 to get the emotional features (vector) of speech.
5. Use the feature vectors got from Step 4 to train HMM.
6. Calculate the recognition rates. If the recognition rate satisfies the target, then we stop the training stage, or go back to step 2.
7. Go to the testing stage.

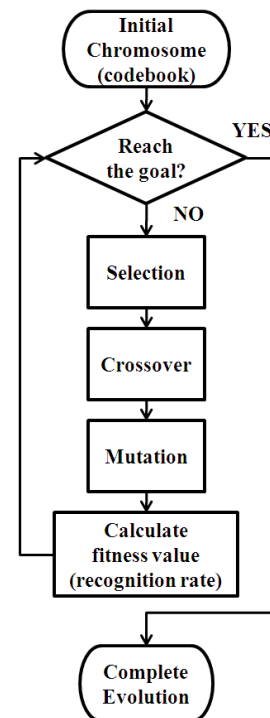


Figure 5: Flow chart of GA.

## 6. EXPERIMENTAL RESULTS

In the experiments, the speeches in the emotional speech database, eINTERFACE 2005, are used as the training and testing data for the proposed system. The speeches in the database are collected from 35 males and 9 females and are all

in English. The database involves speeches with 6 different types of emotions: Anger, Disgust, Fear, Happiness, Sadness, and Surprise. And there are totally 1295 recordings in AVI format in the database. The recordings are transformed into WAV format before they are processed. Then, the recordings in WAV format are sampled with sampling rate of 256 samples, length of 16 bits, and monotone. In this experiment, a 10-fold cross-validation is adopted. This experiment performs ten different runs. There are ten different combinations of the recordings for the ten runs. Each run of experiment is performed for six emotions with 40 recordings for training and 20 recordings for testing.

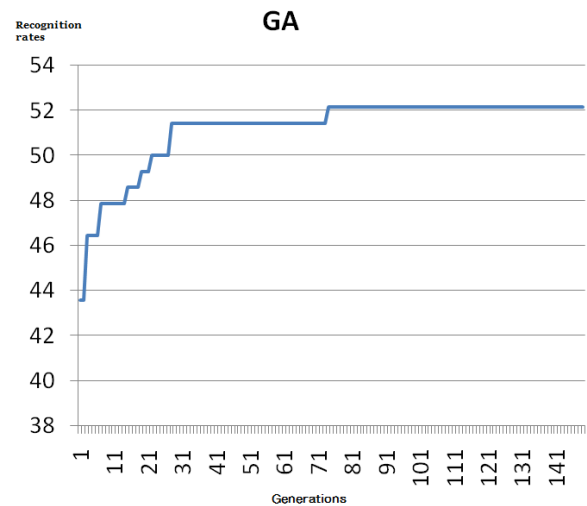
For the feature extraction from the emotional speeches, the EMD process is first applied to the speeches. Then, the weighted IMFs obtained from EMD process are added to form a new signal. Thereafter, MFCC features are calculated from the new signal. In this experiment, the number of IMF in EMD process is set to be 5 and the dimension of the MFCC feature vectors is set to be 12. Moreover, the parameters in GA used in this experiment for finding the optimal weights of IMF are set in Table 1. Concerning the training of HMM, a codebook is first trained by the k-mean method. The matrices A, B and  $\pi$  in HMM are then calculated from the codebook. Each HMM is trained for an emotion in the six considered emotions. That is, an HMM is trained by the all training data for a specific emotion.

In Figure 6, the evolution of GA in the proposed method for one of the ten experiments is revealed. The emotion recognition rate in the evolution converges about at the 75th generation. The average recognition rates for the ten experiments are shown in the Table 2. In order to illustrate the advantage of the proposed method on the feature extraction, a table (Table 3) in which the average recognition rates for various combination of IMFs is displayed. The case that the MFCC is calculated without the pre-process of EMD is also included in the table. It is obvious that the proposed method in which the weighted IMFs with the weights trained by GA are used for feature extraction reaches better recognition rate compared to the other combinations in Table 3. This confirms the performance of the proposed method. However, from Table 4 in which the average recognition rates of the ten experiments for each emotion are revealed, the average recognition rates for the three emotions: disgust, fear and surprise are lower than those of the other emotions. This is because the features of these three emotions are similar to each other. This causes wrong classification between these emotions.

Besides, in order to verify the performance of the proposed method, some comparisons of recognition rates for the ten experiments among the proposed method, the method by using MFCC without EMD and the methods in the papers (Lenzi, Mendes & Da Silva 2000, He et al. 2011) are revealed in Figure 7. It can be seen that the proposed method outperforms the other three methods in the figure.

**Table 1:** Parameters in GA for this experiment.

<b>Number of chromosomes in a generation</b>	32
<b>Number of genes in each chromosome</b>	5
<b>Method of selection</b>	Roulette Wheel Selection
<b>Method of crossover</b>	Linear crossover
<b>Mutation rate</b>	0.03
<b>Survival rate</b>	0.5



**Figure 6:** Evolutions of GA for emotion recognition in the proposed method.

**Table 2:** Average recognition rates for the ten experiments.

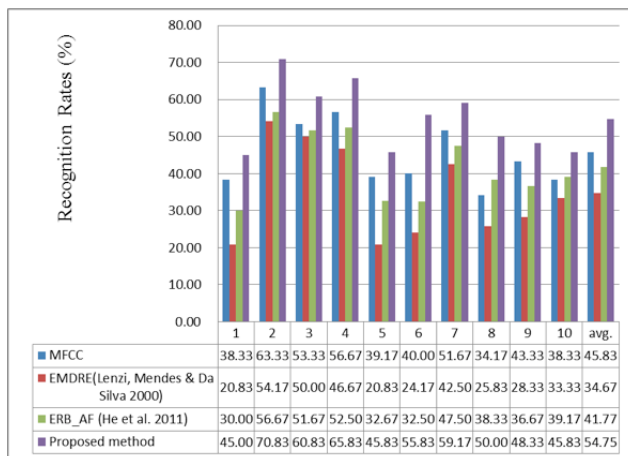
No. of the runs in ten experiments	Avg. Recognition Rates for Each Run (%)
1	45.00
2	70.83
3	60.83
4	65.83
5	45.83
6	55.83
7	59.17
8	50.00
9	48.33
10	45.83
<b>Avg. of the ten experiments</b>	<b>54.75</b>

**Table 3:** Recognition rates by using different combination of IMFs on feature extraction.

	MFCC	IMF1+2+3	IMF2+3+4	IMF1+2	IMF2+3+4+5
1	38.33	40.00	32.50	35.83	26.67
2	63.33	65.00	69.17	54.17	71.67
3	53.33	55.00	52.50	51.67	52.50
4	56.67	53.54	54.17	53.33	56.67
5	39.17	42.50	37.50	32.50	29.17
6	40.00	42.50	35.00	41.67	32.69
7	51.67	51.67	50.83	47.50	50.00
8	34.17	38.33	41.67	33.33	38.33
9	43.33	42.50	47.50	46.79	41.67
10	38.33	38.33	30.00	34.17	34.17
avg.	45.83	46.94	45.08	43.10	43.35

**Table 4:** Recognition rates for six various emotions.

Exper. #	1	2	3	4	5	6	7	8	9	10	Avg.
Anger	0.5	0.75	0.8	0.7	0.75	0.8	0.75	0.7	0.65	0.7	0.71
Disgust	0.6	0.7	0.65	0.7	0.2	0.4	0.4	0.4	0.5	0.35	0.49
Fear	0.1	0.7	0.35	0.8	0.4	0.5	0.45	0.15	0.35	0.35	0.415
Happiness	0.55	0.5	0.65	0.55	0.5	0.65	0.65	0.65	0.45	0.45	0.56
Sadness	0.65	0.9	0.8	0.9	0.6	0.7	0.8	0.6	0.65	0.45	0.705
Surprise	0.3	0.7	0.4	0.3	0.3	0.3	0.5	0.5	0.3	0.45	0.405
Avg.	0.45	0.71	0.61	0.66	0.46	0.56	0.59	0.50	0.48	0.46	0.548



**Figure 7:** Comparison among different methods.

## 7. CONCLUSIONS

This paper proposes a strategy by using EMD, GA and HMM to recognise six types of emotions in the database eNTERFACE 2005. The experiments in this paper demonstrate the good performance of the proposed method compared to the other methods in some existing papers. However, the recognition rates in the experiments are still low, three emotions: disgust, fear and surprise, especially. This is very likely due to the fact that the features of these three emotions are similar to

each other. Hence, some methods, for example, SVM, Fuzzy Logic, Neural Networks, may be integrated into the proposed method to distinct the features more precisely.

## 8. ACKNOWLEDGMENTS

This research work was supported by the Ministry of Science and Technology of the Republic of China under contract MOST 106-2221-E-390 -020 -MY2.

## 9. REFERENCES

- Blunsom, P. (2004) *Hidden Markov Model*. The University of Melbourne, Department of Computer Science and Software Engineering.
- Cen, L., Ser, W., Yu, Z. L., and Cen W. (2010) Automatic Recognition of Emotional States From Human Speeches. *Pattern recognition Recent Advances*, Adam Herout Edition, Chap. 22.
- Doong, S. H., Lai, C. C., and Wu, C. H. (2007) Genetic subgradient method for solving location allocation problems. *Applied Soft Computing*, 7, pp. 373–386.
- Fragopanagos N. and Taylor, J. G. (2005) Emotion Recognition in Human-Computer Interaction. *Neural Networks*, 18(4), pp. 389–405.
- Gales, M. J. F., and Young, S. J. (1995) Robust Speech Recognition in Additive and Convolutional Noise Using Parallel Model Combination. *Comput. Speech Lang.*, 9, pp. 289–307.
- Haupt, R. L., and Haupt, S. E. (1998) *Practical Genetic Algorithms*. John Wiley & Sons.
- He, L., Lech, M., Maddage, N. C., and Allen, N. B. (2011) Study of Empirical Mode Decomposition and Spectral Analysis for Stress and Emotion Classification in Natural Speech. *Biomedical Signal Processing and Control*, 6(2), pp. 139–146.
- Huang, N. E. (1996) The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-Stationary Time Series Analysis. *Proc. R. Soc. London*, 903–995. London.
- Karamalis, P. D., Kanatas, A. G., and Constantinou, P. (2009) A Genetic Algorithm Applied for Optimization of Antenna Arrays Used in Mobile Radio Channel Characterization Devices. *IEEE Trans. Instrumentation and Measurement*, 58, pp. 2475–2487.
- Lai, C. C., and Chang, C. Y. (2009) A hierarchical evolutionary algorithm for automatic medical image segmentation. *Expert Systems with Applications*, 36(1), pp. 248–259.

- Lenzi, E. K., Mendes, R. S., and Da Silva, L. R. (2000) Statistical Mechanics Based on Renyi Entropy. *Physica A: Statistical Mechanics and its Applications*, 280(3-4), pp.337–345.
- Li, X., Zou, X., Zhang, R., and Liu, G. (2008) Method of speech enhancement based on Hilbert-Huang Transform. *7th World Congress on Intelligent Control and Automation*, Chongqing, China, 25–27 June 2008, 8419–8424.
- Lin, F. T. (2005) Evolutionary Computation Part 2: Genetic Algorithms and Their Three Applications. *Journal of Taiwan Intelligent Technologies and Applied Statistics*, 3(1), pp.29–56.
- Mathews, J. H., and Fink, K. D. (2004) *Numerical Methods Using MATLAB, 4th Edition*, Prentice-Hall.
- New, T. L., Foo, S. W., and De Silva, L. C. (2003) Speech Emotion Recognition Using Hidden Markov Models. *Speech Communication*, 41(4), pp.603–623.
- Pan, S. T., and Lai, C. C. (2008) Identification of Chaotic systems by Neural Network with Hybrid Learning Algorithm. *Chaos, Solitons and Fractals*, 37(1), pp.233–244.
- Petrusihin, V. A. (2000) Emotion Recognition in Speech Signal: Experimental Study, Development, and Application. *International Conference on Spoken Language Processing (ICSLP 2000)*, Beijing, China, 16–20 October 2000.
- Schuller, B., Rigoll, G., and Lang, M. (2003) Hidden Markov Model-Based Speech Emotion Recognition. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2003)*, Baltimore, MD, USA, 6–9 July 2003, 2(1), pp.1–4.
- Schuller, B., Rigoll, G., and Lang, M. (2004) Speech Emotion Recognition Combining Acoustic Feature and Linguistic Information in a Hybrid Support Vector Machine-Belief Network Architecture. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004)*, Montreal, Que., Canada, 17–21 May 2004, 1, 577–580.
- Tato, R., Santos, R., Kompe, R., and Pardo, J. M. (2002) Emotion Space Improves Emotion Recognition. *International Conference on Spoken Language Processing (ICSLP 2002)*, Denver, Colorado, USA, 16–20 September 2002.
- Tsai, C. C., Huang, H. C., and Chan, C. K. (2011) Parallel elite genetic algorithm and its application to global path planning for autonomous robot navigation. *IEEE Trans. on Industrial Electronics*, 58(10), pp.4813–4821.
- Tsao, Y., and Lee, C. H. (2009) An Ensemble Speaker and Speaking Environment Modeling Approach to Robust Speech Recognition. *IEEE Trans. on Audio, Speech, and Language Processing*, 17(5), pp.1025–1037.
- Uesaka, K., and Kawamata, M. (2003) Evolutionary synthesis of digital filter structures using genetic algorithm. *IEEE Trans. Circuits Syst. II*, 50, pp.977–983.
- Wang, W., Li, X., and Zhang, R. (2006) Speech Detection Based on Hilbert-Huang Transform. *First International Multi-Symposiums on Computer and Computational Sciences*, Hanzhou, Zhejiang, China, 20–24 June 2006, 290–293.
- Windmann S., and Haeb-Umbach, R. (2009) Parameter Estimation of a State-Space Model of Noise for Robust Speech Recognition. *IEEE Trans. on Audio, Speech, and Language Processing*, 17(8), pp.1577–1590.
- Wu, S., Falk, T. H., and Chan, W. Y. (2011) “Automatic Speech Emotion Recognition Using Modulation Spectral Features. *Speech communication*, 53(5), pp.768–785.
- Yacoub, S., Simske, S., Lin, X., and Burns, J. (2003) Recognition of Emotion in Interactive Voice Response Systems. *8th European Conference on Speech Communication and Technology*, Geneva, Switzerland, September 2003, 1–4.