EEMD-Based Speaker Automatic Emotional Recognition in Chinese Mandarin

Yuqiang Qin[1,2,*] *and Yudong Qi* [1]

[1] College of Business Administration, Capital University of Economics and Business,100070,Beijing, China
[2] College of Economics and Management, Taiyuan University of Science and Technology, 030024,Taiyuan , China

Abstract: Emotion feature extraction is the key to speech emotional recognition. And ensemble empirical mode decomposition(EEMD) is a newly developed method aimed at eliminating emotion mode mixing present in the original empirical mode decomposition(EMD). To evaluate the performance of this new method, this paper investigates the effect of a parameters pertinent to EEMD: speech emotional envelope. Firstly, a speaker emotional envelope features extraction based on EEMD is proposed in the paper. Using the piecewise power function in speech emotional envelope has a better effect in emotional identification. At the same time, the proposed technique has been utilized for classification of four kinds of emotional(angry, happy, sad and neutral) speech signals. Emotional intrinsic mode functions(IMFe) are obtained by empirical mode decomposition on emotional speech signals, the fast fourier transform(FFT) of each intrinsic mode function is extracted as the emotional feature coefficient which is used in speaker emotional identification applying by vector quantization. MATLAB is used to calculate the characteristic of emotional speech signals using empirical mode decomposition (EEMD). We obtain an emotional envelope by transforming the IMFe of emotional speech signals, and obtain a new method of emotion recognition according to different emotional envelop feature vectors. The results indicate the proposed method works well in speaker emotional identification.

Keywords: Speech emotional recognition, Ensemble empirical mode decomposition(EEMD), Emotional intrinsic mode functions(IMFe), EEMD emotional features sifting (EEFS),emotional envelope

1 Introduction

Affective human computer interaction has been the focus of artificial intelligence research for several years now, and the research has moving ahead from the simple information exchange between human and computer towards the affective communication [1]. Affective human computer interaction technology could be widely applied in virtual reality, especially in the field of entertainment and games. Besides, the virtual human and psychiatric aid are the further application prospects for affective human computer interaction [2]. Making computer recognize the emotion of human being is the foundation of affective human computer interaction. Emotion recognition of speech as a significant part has become a challenge to speech processing. The accustomed way for speech emotion recognition is to distinguish between a defined set of discrete emotions [3, 4].

Speech has been one of the major communication medium for years and will continue to do so until video communication becomes widely available and easily accessible. Although numerous technologies have been developed to improve the effectiveness of speech communication system, human interaction with machines and robots are still far from ideal. It is acknowledged that human can communicate effectively with each other through the telephony system. This situation motivates many researchers to study in depth the human communication system, with emphasis on its ability to express and infer emotion for effective social communication. Understanding the interlocutors emotion and recognizing the listeners perception is the key to boost communication effectiveness and interaction. Nonetheless, the perceived emotion is subjective and very much dependent on culture, environment and the pre-emotional state of the listener. Attempts have been made to understand the influence of culture in speech

---

* Corresponding author e-mail: qinyuqiang@126.com

emotion and researchers have reported mixed findings that lead us to believe there are some common acoustical characteristics that enable similar emotion to be discriminated universally across culture. Yet there are unique speech attributes that facilitate exclusive emotion recognition of a particular culture. Understanding culture dependency is thus important to the performance of the speech emotion recognition system [5, 6].

In this paper, an innovative algorithm of improved EMD has been presented for the analysis of frequency modulation (FM) signals [7]. First, extract the first level extrema and the second level extrema to construct one two-level extrema structure (TLES). Using the locations' relations between the first level extrema and the second level extrema, what the traditional EMD does for FM signals is determined. The propositions that can help us to judge whether the multi-component can be separated well or not by EMD in practice and then tell us what the EMD will do in practice are deduced. Four cases are presented for decomposition of FM components using EMD. According to these cases, different improved methods are presented to separate the FM components that EMD fails to separate. The least-squares method is used to estimate the phase polynomials of FM components. The orthogonal relations between different components are used to estimate the optimal amplitudes of FM components. Moreover, the performance of EMD is analyzed via the multi-level extrema distribution as well, which can be used in practice for the aforehand judgement. Finally, experiments are given to show our new improved EMD [8, 9].

Therefore, the ensemble empirical mode decomposition (EEMD) [9], a signal processing technique particularly suitable for non-linear and non-stationary series, has recently been proposed as a new tool for emotional data analysis. The EEMD method is able to decompose a complex emotional speech signal into a series of emotional intrinsic mode functions (IMFe) and a residue in accordance with different frequency bands [10]. These are signal components highlighting distinct time-scales (frequencies) of the input time-series. The IMFe represents the natural oscillatory mode embedded in the signal. EEMD is self-adaptive because the IMFe works as the basis functions determined by the signal itself rather than what is predetermined. Therefore, EEMD is highly efficient in non-stationary emotional data analysis [11].

2 EEMD-Based Emotional Feature Extraction

2.1 Emotional intrinsic mode function(IMFe)

HilbertCHuang transform (HHT) is a novel analysis method for nonlinear and non-stationary data, which was developed by Huang et al. in 1998. Its key part is the so-called empirical mode decomposition (EMD), with which any complicated data set can be decomposed into finite (often smaller) number of intrinsic mode functions (IMFs) which admit well-behaved Hilbert transforms. An IMF is defined as a function satisfying the following conditions:

(1) The number of extrema and the number of zerocrossings must be either equal or differ at most by one.

(2) At each point of the data, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

With Hilbert transform, the IMFs yield instantaneous frequencies as functions of time, that give sharp identifications of embedded structures. The final presentation of the results is a timeCfrequencyCenergy distribution, designated as the Hilbert spectrum. Being different from Fourier decomposition and wavelet decomposition, EMD has no specified basis. Its basis is adaptively produced depending on the signal itself, which brings not only high decomposition efficiency but also sharp frequency and time localization. A key point is that the signal analysis based on HHT is physically significant. Because of its excellence, HHT has been utilized and studied widely by researchers and experts in signal processing and other related fields.

A multi-resolution decomposition technique is presented, EEMD which is adaptive and appears to be suitable for non-linear and non-stationary emotional speech signal analysis. It was carried in the time domain to form the basis functions adaptively. The major advantage of EEMD is the basis functions can be directly derived from the emotional speech signal itself. The emotional intrinsic modes are not necessarily sinusoidal functions. Apparently, EEMD is empirical, intuitive, direct, and adaptive. The EEMD decomposes the original emotional signal into a definable set of adaptive basis of functions called the emotional intrinsic mode functions(IMFe). In fact, IMFe can be both amplitude and frequency modulated. Each IMFe must satisfy two basic conditions:

(1) in the whole emotional data set, the number of extrema and the number of zero crossings must either equal or differ at most by one. Note, either local minima or local maxima are extrema. Moreover, a sample S_i in a time-series is a local maximum if $S_i > S_{i-1}$ and $S_i > S_{i+1}$, and a sample S_i is a local minimum if $S_i < S_{i-1}$ and $S_i < S_{i+1}$, where i is a discrete time;

(2) at any point, the mean value of the emotional envelope, one defined by the local maxima (upper

emotional envelope) and the other by the local minima (lower emotional envelope) is zero. Since emotional sifting is a recursive process, a sifting stopping rule is required. The definition above is empirical and currently there are no definite equations for estimating IMFe.

Therefore, any arbitrary time series satisfying conditions 1 and 2 is an IMFe [12].

Early assessment of machinery health condition is of paramount importance today. A sensor network with sensors in multiple directions and locations is usually employed for monitoring the condition of rotating machinery. Extraction of health condition information from these sensors for effective fault detection and fault tracking is always challenging. EEMD is an advanced signal processing technology that has been widely used for this purpose. Standard EMD has the limitation in that it works only for a single real-valued signal. When dealing with data from multiple sensors and multiple health conditions, standard EMD faces two problems. First, because of the local and self-adaptive nature of standard EMD, the decomposition of signals from different sources may not match in either number or frequency content. Second, it may not be possible to express the joint information between different sensors. The present study proposes a method of extracting fault information by employing multivariate EEMD and full spectrum. EEMD can overcome the limitations of standard EMD when dealing with data from multiple sources. It is used to extract the intrinsic mode functions (IMFs) embedded in raw multivariate signals. A criterion based on mutual information is proposed for selecting a sensitive IMF. A full spectral feature is then extracted from the selected fault-sensitive IMF to capture the joint information between signals measured from two orthogonal directions. The proposed method is first explained using simple simulated data, and then is tested for the condition monitoring of rotating machinery applications. The effectiveness of the proposed method is demonstrated through monitoring damage on the vane trailing edge of an impeller and rotor-stator rub in an experimental rotor rig.

2.2 EEMD emotional features sifting (EEFS)

In the cause of the separate emotional components called IMFe, we perform a process called EEMD emotional features sifting (EEFS)[13]. The purpose of emotional sifting is to subtract the large-scale emotional features of the speech signal repeatedly until only the fine-scale emotional features remain. First, the original emotional speech signal, $x_e(t)$, should be enclosed by the upper and lower emotional envelope in the time domain. Using a cubic spline, the local maxima is connected forming the upper emotional envelope $u_+(t)$ and the local minima is connected forming the lower emotional envelope $l_-(t)$.

The two emotional envelopes cover all the data points. The emotional envelope mean is determined as follows, $m(t) = (u_+(t) + l_-(t))/2$. The first emotional component is described as:

$$h_1 = x_e(t) - m(t) \qquad (1)$$

The emotional component $h_1(t)$ is now examined to see if it satisfies the conditions to be an IMFe. If $h_1(t)$ doesn't satisfy the conditions, $h_1(t)$ is regarded as the original emotional data, the sifting process would repeat, obtaining the mean of the upper and lower emotional envelopes, which is designated as m_{11} ; therefore,

$$h_{11}(t) = h_1(t) - m_{11}(t) \qquad (2)$$

Then, repeat the procedure until k_{1k} is an IMFe, i.e.,

$$h_{1k}(t) = h_{1(k-1)(t) - m_{1k}(t)} \qquad (3)$$

After k emotional siftings, we explained the first emotional intrinsic mode to be

$$c_1 = h_{1k} \qquad (4)$$

Finally, c_1 revealed the higher frequency component of IMFe. To obtain enough physical definitions of IMFe, the sifting stop criteria, namely, the stop conditions, are of great importance and are found by the following equation:

$$SD = \sum[\frac{|h_{1()k-1}(t) - h_{1k}(t)|^2}{h_{1(k-1)(t)*h_{1(k-1)(t)}}}] \qquad (5)$$

The typical values of SD are 0.2 and 0.3. To obtain the second and subsequent emotional intrinsic mode functions, the residual signal can be calculated as :

$$x_e(t) - c_1(t) = r_1(t) \qquad (6)$$

r_1 considers the original emotional data, and by repeating the above procedures, $x_e(t)$ could be obtained by the second IMFe component c_2 . The procedure as described above is repeated for times, then the n-IMFe of emotional signal $x_e(t)$ could be obtained

$$r_1(t) - c_2(t) = r_2(t)...r_{n-1}(t) - c_n(t) = r_n(t) \qquad (7)$$

The decomposition procedure can be stopped when the residue, r_n , becomes a constant, a monotonic function, or a function containing only a single extrema, from which no more IMFe can be extracted. By summing Eqs. (6) and (7), the original emotional speech signal can be reconstructed as follows:

$$x_e(t) = \sum c_j(t) + r_n(t) \qquad (8)$$

where n is the number of IMFes extracted from $x_e(t)$, $r_n(t)$ denotes the final residue, and is time. The IMFes $c_1, c_2, ..., c_n$ are comprised of different frequency bands ranging from high to low. The frequency components included in each different frequency band, while r_n
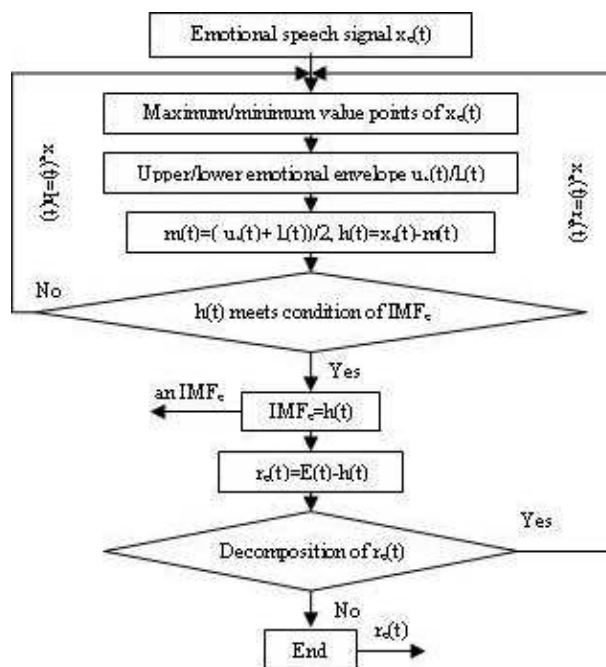
Fig. 1: The entire flow chart of EEMD-Based Emotional Envelope Computing Decomposition (EEMD emotional feature sifting, EEFS)

represents the central tendency of emotional speech signal $x_e(t)$. Figure 1 indicates the entire flow chart of EEMD-Based Emotional Envelope Computing Decomposition , call EEMD emotional feature sifting (EEFS).

3 Speech Emotional Envelope Optimal Computing in Continuous Chinese Mandarin

Computing upper and lower emotional envelope is the key of EEMD emotional envelope feature extraction [14, 15]. If emotional envelope is inaccurate, then EEMD of emotional signal is not entirely. Therefore the piecewise power function is used in speech emotional envelope as follows.

All interpolation points:$P_1(x_1, y_1)P_2(x_2, y_2)P_n(x_n, y_n)$,interpolating function: $y = f(x)$. Power function interpolate any three contiguous points: $P_{i-1}(x_{i-1}, y_{i-1})P_i(x_i, y_i)P_{i+1}(x_{i+1}, y_{i+1})$. So:

$$h_i = \frac{(x_{i+1} - x_{i-1})(y_{i-1} - y_i) - (x_{i-1} - x_i)(y_{i+1} - y_{i-1})}{x_{i+1} - x_{i-1}}$$
(9)

When $x \le x_i$, interpolating function $f_i(x)$ meet:

$$f_i = (\frac{x - x_i}{x_{i-1} - x_i})^\beta h_i + \frac{y_{i+1} - y_{i-1}}{x_{i+1} - x_{i-1}}(x - x_i) + y_i$$
(10)

When $x \ge x_i$, interpolating function $f_i(x)$ meet:

$$f_i = (\frac{x - x_i}{x_{i+1} - x_i})^\beta h_i + \frac{y_{i+1} - y_{i-1}}{x_{i+1} - x_{i-1}}(x - x_i) + y_i$$
(11)

After interpolating $f_i(x), f_{i+1}(x)$ again, the piecewise power function curve $f_{i,i+1}(x)$ between $P_i(x_i, y_i)$ to $P_{i+1}(x_{i+1}, y_{i+1})$ are attained:

$$f_{i,i+1}(x) = \frac{x_{i+1} - x}{x_{i+1} - x_i} f_i(x) + \frac{x - x_i}{x_{i+1} - x_i} f_{i+1}(x)$$
(12)

Error tolerance of the piecewise power function is analyzed as follows:

Suppose $f(x) \subseteq C_1[a, b]$, the piecewise function $E(x)$ meets Eqs.(12), and $e_i(x), e_{i+1}(x)$ meet Eqs.(10),(11) respectively,

$$M = max|f''(\xi)| : x_{i-1} \le \xi \le x_{i+1},$$
$$h = max|x_{i+1} - x_i| : 1 \le i \le n, then :$$

$$max|f(x) - E(x)| \le \frac{1}{2}Mh^2 + \frac{2\beta^\beta}{(\beta + 1)^{\beta+1}}Mh^2$$
(13)

Prove : $\sum l_j = 1 and x \subseteq [x_i, x[i+1]] \Rightarrow l_i(x) + l_{i+1}(x)$, $f(x_{i-1}) = y_{i-1} f(x_i) = y_i f(x_{i+1}) = y_{i+1} \Rightarrow$

$$|f(x) - E(x)| =$$
$$\left| \frac{x - x_{i+1}}{x_i - x_{i+1}} f(x) + \frac{x - x_i}{x_{i+1} - x_i} f(x) - [\frac{x - x_{i+1}}{x_i - x_{i+1}} e_i(x) + \frac{x - x_i}{x_{i+1} - x_i} e_{i+1}(x)] \right|$$
$$\le frac - x_{i+1}{x_i - x_{i+1}}|f(x) - e_i(x)| + \frac{x - x_i}{x_{i+1} - x_i} f(x) - e_{i+1}(x)|$$
(14)

$$|f(x) - e_i(x)| =$$
$$|f(x) - [(\frac{x - x_i}{x_{i+1} - x_i})^\beta h_i + \frac{y_{i+1} - y_{i-1}}{x_{i+1} - x_{i-1}}(x - x_i) + y_i]|$$
$$= \left| f(x) - f(x_i) - \frac{x - x_i}{x_{i+1} - x_i})^\beta [(y_{i-1} - y_i) \right.$$
$$\left. - \frac{x_{i-1} - x_i}{x_{i+1} - x_{i-1}}(y_{i+1} - y_{i-1})] - \frac{y_{i+1} - y_{i-1}}{x_{i+1} - x_{i-1}}(x - x_i) \right|$$
$$= \left| f'(\xi_1)(x - x_i) - \frac{x - x_i}{x_{i+1} - x_i})^\beta \right.$$
$$[f'(\xi_2)(x_{i-1} - x_i) - f'(\xi_3)(x_{i-1} - x_i)] - f'(\xi_3)(x - x_i)|$$
$$= |(f'(\xi_1) - f'(\xi_3))(x - x_i)$$
$$- \frac{x - x_i}{x_{i+1} - x_i})^\beta (f'(\xi_2) - f'(\xi_3))(x_{i-1} - x_i) \right|$$
$$\le |(f'(\xi_1) - f'(\xi_3))||(x - x_i)|$$
$$+ |\frac{x - x_i}{x_{i+1} - x_i}|^\beta |f'(\xi_2) - f'(\xi_3)||x_{i-1} - x_i|$$
$$(x_i \le \xi_1 \le x, x_{i-1} \le \xi_2 \le x_i, x_{i-1} \le \xi_3 \le x_{i+1})$$
(15)

$$|f(x) - e_{i+1}(x)| =$$

$$|f(x) - (\tfrac{x-x_{i+1}}{x_i-x_{i+1}})^\beta h_{i+1} - \tfrac{y_{i+2}-y_i}{x_{i+2}-x_i}(x-x_{i+1}) - y_{i+1}|$$

$$= \left| f'(\overline{\xi_1})(x-x_{i+1}) - f'(\overline{\xi_3})(x-x_{i+1}) \right.$$

$$\left. -(\tfrac{x-x_i}{x_{i+1}-x_i})^\beta)[f'(\overline{\xi_2})(x_i-x_{i+1}) - f'(\overline{\xi_3})(x_i-x_{i+1})] \right|$$

$$\leq |f'(\overline{\xi_1}) - f'(\overline{\xi_3})||x-x_{i+1}|$$

$$+ |\tfrac{x-x_{i+1}}{x_i-x_{i+1}}|^\beta |f'(\overline{\xi_2}) - f'(\overline{\xi_3})||x_i-x_{i+1}|$$

$$(x \leq \overline{\xi_1} \leq x_{i+1}, x_i \leq \overline{\xi_2} \leq x_{i+1}, x_{i-1} \leq \overline{\xi_3} \leq x_{i+1}) \quad (16)$$

Eqs. (15)(16) → Eqs. (14)

$$|f(x) - E(x)| \leq \frac{x-x_{i+1}}{x_i-x_{i+1}}|f(x) - e_i(x)|$$

$$+ \frac{x-x_i}{x_{i+1}-x_i}|f(x) - e_{i+1}(x)|$$

$$\leq 1/4|f'(\xi_1) - f'(\xi_3)||x_{i+1}-x_i|$$

$$+ 1/4|f'(\overline{\xi_1}) - f'(\overline{\xi_3})||x_{i+1}-x_i|$$

$$+ \frac{\beta^\beta}{(\beta+1)^{(\beta+1)}}|f'(\xi_2) - f'(\xi_3)||x_{i-1}-x_i|$$

$$+ \frac{\beta^\beta}{(\beta+1)^{(\beta+1)}}|f'(\overline{\xi_2}) - f'(\overline{\xi_3})||x_i-x_{i+1}|$$

$$\leq 1/2Mh^2 + 2\frac{\beta^\beta}{(\beta+1)^{(\beta+1)}}Mh^2$$

Error range of the piecewise power function increase when β reduce, and the maximum range of error is $[Mh^2/2, 5Mh^2/2]$.

4 Simulation Experiment and Analysis

4.1 Emotional corpus database

An emotion database is specifically designed and set up for text-independent emotion classification studies. The database includes short utterances covering the 4 archetypal emotions, namely angry, happy, sad and neutral. Non-professional speakers are selected to avoid exaggerated expression. A total of 12 native Chinese Mandarin language speakers (6 males and 6 females) are employed to generate 720 utterances. Sixty different utterances, ten each for each emotional mode, are recorded for each speaker. The recording is done in a quiet environment using a mouthpiece microphone.

Subjective assessment of the emotional speech corpus by human subjects was carried out. One of the objectives of the subjective classification is to determine the ability of listeners to correctly classify the emotional modes of the utterances. Another objective is to compare the accuracy of classification by the proposed system with human classification performance [16,17].
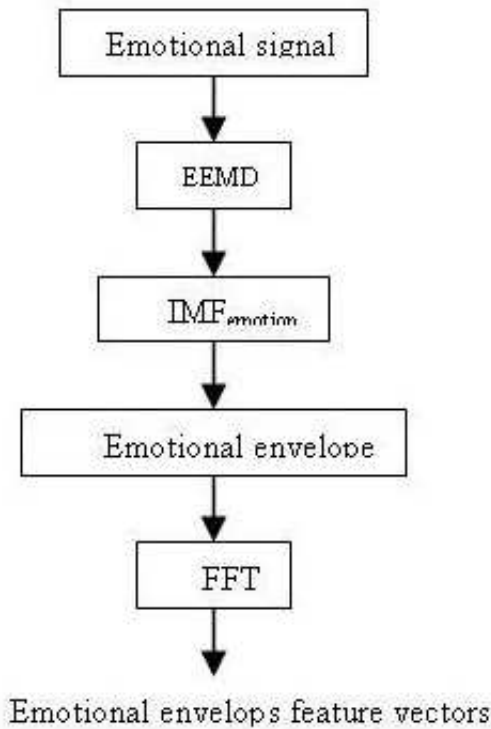

Fig. 2: EEMD-based Emotional Envelope Feature Extraction for 4 Emotional Speech Signal

Three subjects of different language background are engaged for the subjective tests. The utterances were played back in random order and the subjects were requested to indicate which one of the 4 emotional modes was portrayed. The utterances were presented via headphones and repeated two times. The language of the utterances presented to the human subject was neither his mother tongue nor any other language that he has any knowledge to perceive linguistically. Hence the judgment was made based on the perceived emotional content rather than the context of the utterances [18].

4.2 Simulation result

MATLAB [19] is used to calculate the characteristic of emotional(angryhappysad and neutral) speech signal using EEMD, we attain a lot of components of IMFemotion, and obtain an emotional envelope by transforming the IMFemotion of emotional speech signal, and obtain a new method of emotion recognition according to different emotional envelops feature vectors, as the Figure 2 suggests that.

Figure 3 show simulation result of 4 emotional(angry, happy, sad and neutral) continuous Chinese mandarin "My father bought me a car(in Chinese)", and its emotional envelops features [20].
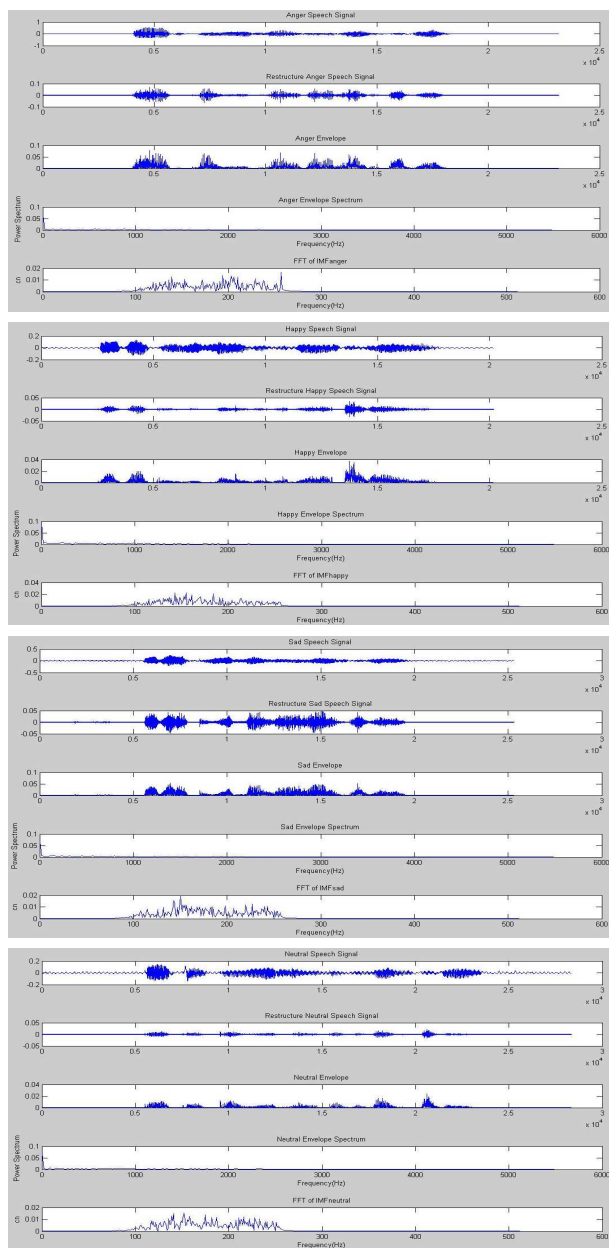
Fig. 3: 4 Emotional Envelopes and Its Emotional Envelope Feature Vectors

As depicted in the Figure.3, we can see that emotions are decomposed using the piecewise power function in continuous Chinese mandarin speech, and gain a lot of IMFemotions. It is therefore that better emotional envelopes are attained.

In term of Figure 3, characteristics of frequency distribution of $IMF_{emotion}$ are attained by FFT, as depicted Table 1. Four emotional feature vectors are T_{angry}=[120,140,155,195,210,260], T_{happy}=[140,150,160,185,205,210],

Table 1: 4 Emotional Envelop Feature Vectors in Chinese Mandarin

IMF_e	f_1(Hz)	f_2(Hz)	f_3(Hz)	f_4(Hz)	f_5(Hz)	f_6(Hz)
IMF_{angry}	120	140	155	195	210	260
IMF_{happy}	140	150	160	185	205	210
IMF_{sad}	110	140	150	200	215	220
$IMF_{neutral}$	135	150	155	180	210	215

Table 2: EEMD-Based Confusion Matrix for Automatic Emotional Recognition

Emotion	angry	happy	sad	neutral	rate(%)
angry	119	8	0	0	93.7
happy	12	52	0	3	77.6
sad	0	0	62	0	100
neutral	2	2	0	75	94.9
precision(%)	89.5	83.9	100	96.2	

T_{sad}=[110,140,150,200,215,220]　　　　　　and $T_{neutral}$=[135,150,155,180,210,215] respectively.

4.3 Evaluation

Confusion matrix are shown in Tables 2 (left-most column being the true emotions), for the best recognition performance achieved by envelop features vectors. The rate column lists per class recognition rates, and precision for a class is the number of samples correctly classified divided by the total number of samples classified to the class.

We can see from the confusion matrix that the envelope optimal algorithm contributes to improving the emotional recognition and precision rates of all emotion categories. It is also shown that most emotions can be correctly recognized with above 90% accuracy, with the exception of joy, which forms the most notable confusion pair with angry. The results also divulge that the algorithm detects sad emotion efficiently with a recognition rate of 100%.

5 Conclusions and Future Work

In this paper, a speaker emotion identification algorithm based EEMD for emotional feature extraction and classification using emotional envelop feature has been developed. The EEMD can effectively sift the riding wave from every complex emotional signal of the time series. The sifted emotional IMF_e components represented important information in the entire emotion signal set. The energy correlates closely with every emotional IMF_e component. The experimental results indicated the proposed algorithm is an effective recognition method for speaker emotion identification. Emotions are decomposed using the piecewise power function in continuous Chinese mandarin speech, and better emotional envelopes are

attained. It is also shown that most emotions can be correctly recognized with above 90% accuracy, with the exception of happy, which forms the most notable confusion pair with angry. The results also divulge that the algorithm detects sad emotion efficiently with a recognition rate of 100%.

As most of the emotional signals under consideration are chaotic in nature, so application of non-linear dynamic tools like the computation of Lyapunov exponents can also be potentially explored for feature extraction procedures.

6 Acknowledgement

References

[1] Yuqiang Qin, Xueying Zhang. Fuzzy Support Vector Machine-Based Emotional Optimal Algorithm in Spoken Chinese[J]. Journal of Computational and Theoretical Nanoscience, **9**, 1715-1719 (2012).

[2] Kimberly L. Day, Cynthia L. Smith.Understanding the role of private speech in children's emotion regulation[J]. Early Childhood Research Quarterly, **28**, 405-414 (2013).

[3] Norhaslinda Kamaruddin, Abdul Wahab, Chai Quek. Cultural dependency analysis for understanding speech emotion[J]. Expert Systems with Applications, **39**, 5115-5133 (2012).

[4] Ishi, C., Ishiguro, H., Hagita, N. Analysis of the roles and the dynamics of breathy and whispery voice qualities in dialogue speech[C]. EURASIP J. Audio Speech Music Process., article ID 528193, 12 pages (2010).

[5] Thomas F.Quatieri. Discrete-time speech signal processing : principles and practice[M]. Beijing: Publishing House of Electronics Industry, 468-488 (2004).

[6] Wu, S., Falk, T., Chan, W.-Y.. Automatic recognition of speech emotion using long-term spectro-temporal features[C]. In: Proc. Internat. Conf. on Digital Signal Processing, 1-6 (2009).

[7] Julien Fleureau, Amar Kachenoura, Laurent Albera, Jean-Claude Nunes, Lotfi Senhadji. Multivariate empirical mode decomposition and application to multichannel filtering[C]. Signal Processing, In Press, Corrected Proof, 789-797 (2011).

[8] Xu Guanlei, Wang Xiaotong, Xu Xiaogang. Time-varying frequency-shifting signal-assisted empirical mode decomposition method for AMCFM signals[C]. Mechanical Systems and Signal Processing, **23**, 2458-2469 (2009).

[9] Huang NE, Wu ML, Qu WD. Applications of Hilbert-Huang Transform to Non-stationary Financial Time Series Analysis [J]. Appl. Stochastic Models Bus. Indust, **19**, 245-268 (2003).

[10] You-Shyang Chen, Ching-Hsue Cheng. Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry [J]. Knowledge-Based Systems, **39**, 224-239 (2013).

[11] Xiaomin Zhao, Tejas H. Patel, Ming J. Zuo. Multivariate EMD and full spectrum based condition monitoring for rotating machinery[J]. Mechanical Systems and Signal Processing, **27**, 712-728 (2012).

[12] Md. Ashfanoor Kabir, Celia Shahnaz. Denoising of ECG signals based on noise reduction algorithms in EMD and wavelet domains[J]. Biomedical Signal Processing and Control, **7**, 481-489 (2012).

[13] GUO Shu-Qing. Data discussion of the EMD method in the practice[J]. Signal Processing, **26**, 277-285 (2010).

[14] QI Yu-dong, FAN He-jun. Deregulation: Reform Trend of Chinas Natural Monopoly Industries[J]. China Industrial Economics, **21**, 62-72 (2009).

[15] Quek S T, Tua P S, Wang Q. Detecting anomalies in beams and plate based on the Hilbert - Huang transform of real signals [J].Smart Materials and Structures, **34**, 447-460 (2003).

[16] ZHONG You-ming, JIN Tao, QIN Shu-ren. New Envelope Algorithm for Hilbert-Huang Transform[J]. Journal of Data Acquisition & Processing, **20**, 13-17 (2005).

[17] Xu Guanlei, Wang Xiaotong, Xu Xiaogang, Zhou Lijia. Improved EMD for the analysis of FM signals[J] Mechanical Systems and Signal Processing, **33**, 181-196 (2012).

[18] Sun, R., E., M., Torres, J.. Investigating glottal parameters for differentiating emotional categories with similar prosodics[C]. In: Proc. Internat. Conf. on Acoustics, Speech and Signal Processing, 4509-4512 (2009).

[19] FU Yao, WANG Hong-jun, WU Guo-xin. Realization of EMD signal processing method in LabVIEW and MATLAB[J]. Journal of Beijing Institute of Machinery, **23**, 16-18 (2008).

[20] Yuqiang Qin, Xueying Zhang. MSF-Based Speaker Automatic Emotional Recognition in Continuous Chinese Mandarin[J]. Procedia Engineering, **15**, 2229-2233 (2011).

Yuqiang   Qin

1976-, postdoctor, associate professor, Capital University of Economics and Business, Beijing, Taiyuan University of Science and Technology, Taiyuan,Shanxi,China.

The major field of study is business administration, information technology.

Yudong Qi

1966-,   professor, doctoral supervisor, Capital University of Economics and Business, Beijing, China. The major field of study is business administration.