Proceeding of the IEEE
International Conference on Robotics and Biomimetics (ROBIO)
Shenzhen, China, December 2013

Mandarin Emotion Recognition Based on Multifractal Theory Towards Human-Robot Interaction

Hong Liu*, Wenjuan Zhang

Abstract— **Emotion recognition is crucially related with friendly and humanistic human-robot interaction. Our paper aims at developing a new kind of features to mandarin emotional speech signal based on multifractal theory. Firstly, phase space structure differentiate with respect of initials and finals indicate the fractal phenomenon during speech produce process. Further, positive largest Lyapunov exponent proved existing chaos. To quantitatively measure the chaos, extension of fractal concept–multifractal is calculated by multifractal detrended fluctuation analysis (MFDFA) and Legendre transformation. Besides, the underlying fractal characteristics during calculation process is analyzed, which further verifies the emotional speech is multifractal rather than monofractal. Multifractal spectrum visually shows that various emotion is differentiated with each other. After extracting parameters of mulfractal spectrum, several comparative experiments are established, which is implemented with BP neural network and support vector machine (SVM) respectively to hence the comparison between our approach and conventional one. At last, improvement in recognition accuracies demonstrates that our method is available and effective.**

Index Terms: mandarin speech recognition, Lyapunov exponent, MFDFA, multifractal spectrum

I. INTRODUCTION

In human-human interaction, emotional information conveyed through speech is an important factor for communication, for it carries extra information including one's attitude, viewpoint and underlying intent. The intelligential gap between human and robot lies on that robot is inflexible and task-oriented, while human originally possess the properties to decision-making[1] by himself/herself. Emotion recognition for human speech increases significantly with the urge to improve both the naturalness and efficiency of human-machine interaction[2].

However, recognizing emotions from spoken language for robot is extremely challenging due to several aspects such as defining various emotion states with precise categories. What's more, judging emotion states to some extent is not only about the emotion state of speaker but about the perception and experience of receiver, which means a unified standard is not proper for robots to interact with various persons changing over time, etc.. Hence, series of assumptions are invariable for a practical approach to emotion recognition so as to bound the problem. In this paper, we focus on analyzing fractal characteristics of emotional speech and verifying its efficiency for recognition rather than pursuing high recognition accuracies towards a large variety of emotions, though it seems attractive. Therefore, taking the diversity and representability of emotion into consideration, we select several basic emotion classes according to previous study, that is neutrality, happiness, sadness and anger. Russell[20] addressed an emotion dimension with words of emotional states located in angles and distances within coordinates, where emotions mentioned above is located in four quadrants within, distinguishable and normally present in human daily communication.

Most previous studies normally used pattern recognition methods for speech emotion recognition. For example, popular classifiers are linear discriminate classifiers[3] and k-nearest neighbor(KNN)[4] for their easy implementation and insensitiveness to unbalanced classes. As an extension of LDC, support vector machine(SVM) achieves promising generalization properties[5]. Non-linear discriminative classifiers like artificial neural networks(ANN) and decision trees are also employed due to their robust performance in some cases[6][7]. Ensembles of above classifiers combine their advantages and avoid weakness. Even the same feature vector can yield totally different classification results using different algorithms[8]. In addition, feature extraction and selection play a crucial role in emotion modeling and to some extent determine what an emotional speech is perceived for a robot. A variety of features dealing with spoken mechanism, signal processing, linguistics, listening mechanism and many other realms have been explored and studied. Each realm provides a snapshot of emotional speech from different perspectives, and snapshots are merged and evaluated so as to obtain the panorama of emotional speech.

An important trend in the emotion recognition area is to investigate new methods for feature generation, such as deep learning[8]. In this paper, we provide another snapshot of emotional speech based on nonlinear dynamical system and fractal theory which is distinguishable in previous work's assumption and fundamentals. Conventional methods are basically relying on the preliminary approximation that speech is temporal stationary within a millisecond scale neglecting the non-stationary behavior of speech signals.

The process of speech production is essentially a highly involving aerodynamical phenomena, such as bifurcation,

This work is supported by National Natural Science Foundation of China (NSFC, No.60875050, 60675025), National High Technology Research and Development Program of China (863 Program, No.2006AA04Z247), Scientific and Technical Innovation Commission of Shenzhen Municipality (No.JCYJ20120614152234873, CXC201104210010A, JCYJ20130331144631730, No.JCYJ20130331144716089).

*Corresponding Author: Hong Liu is with the Engineering Lab on Intelligent Perception for Internet of Things (ELIP), Shenzhen Graduate School, Peking University, Shenzhen, China hongliu@pku.edu.cn

Wenjuan Zhang is with the Engineering Lab on Intelligent Perception for Internet of Things (ELIP), Shenzhen Graduate School, Peking University, Shenzhen, China zh_wj@foxmail.com

978-1-4799-2744-9/13/$31.00 ©2013 IEEE

turbulence and oscillation occurring in vocal tract as air flowing through; vibration as air prompting through the crack formed by vocal cords; lips and teeth causing obstacles as air jetting outside, which has been addressed by the experimental and theoretical evidence in [9][10][11]. This phenomena leads to the wide application of fractal theory and dynamical systems with irregular and unpredictable, noise-like nature to speech signal processing[12][13][14]. Besides, multifractal nature is among the invariant quantities that describe the geometrical structures in turbulence and complexity of a dynamical system[15], and has been employed to automatic speech recognition[16], speaker recognition[17], speech decomposition, representation and characterization[18][19].

In this paper, we process emotional speech based on fractal theory, which inspired by [21] and we further discuss and obtain some fundamental conclusions with theoretical discipline, and apply it to mandarin speech. This paper is organized as follows: Firstly, we verify the nonlinear dynamical characteristics by construct mandarin initials and finals in phase space, and calculate the largest Lyapunov exponent of these phonemes. Then, we extract underlying multifractal spectrum of emotional speech, which are differentiated with respect to various emotion classes visually. At last, we utilize boosting method merging support vector machine(SVM) and back propagation(BP) neural network to validate the effectiveness of the fractal features.

II. Nonlinear Dynamicalal System Analysis

Compared to conventional processing of time series in time domain and frequency domain, nonlinear time series process like chaotic invariant estimation and nonlinear modeling, relies on phase space. A system or time series in our case is chaotic or not evolving analysis of these time series in phase space, where all possible states of a system are represented with one unique point in it corresponding to each possible state of the system[1]. As Eq.(1) shows, the reconstruction of time series needs choose proper embedding dimension m and time-delayed τ. Projecting time series to 3-dimensional space provides intuitive and visual understanding to the problem. Thus, we choose embedding dimension m=3 and time-delayed τ=50. The phase plots corresponding to the finals $/a/$,$/u/$ and initials $/b/$ and$/sh/$ are shown in Fig.1, the corresponding waveforms are plotted in Fig.2.

$$\mathbf{x}(i) = \{x(i), x(i+\tau), ..., x(i+(m-1)\tau)\} \qquad (1)$$

It is clear that finals' phase plots reveal folded limit cycle structure while initials' reveal irregular random string. There are several parameters describing the phase space structure, such as correlation dimension, Lyapunov exponents and kolmogorov entropy. Lyapunov exponents show the dynamics of the (temporal) evolution of the trajectories, indicating the exponential divergence or convergence of trajectories toward an attractor in a multidimensional flow and reflecting the properties of the underlying attractor by their sign and magnitude[13].

[1] http://en.wikipedia.org/wiki/Phase_space

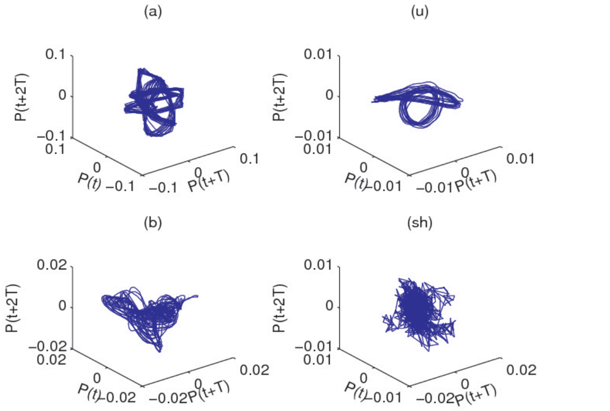

Fig. 1. Phase space plot for /a/,/u/,/b/,/sh/. Embedding dimension $d_e = 3$ with τ=50.

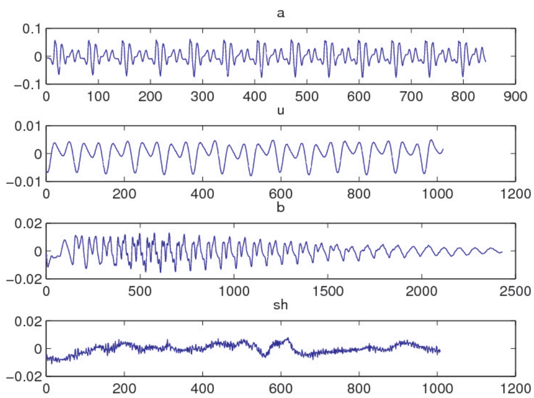

Fig. 2. Time waveforms for natural utterances of finals /a/,/u/ and initials /b/,/sh/.

There are two methods about calculating Lyapunov exponents(LEs) observing time series when dynamical equation is not known to us: Wolf's method[22] tracing the trajectories of attractor and concluding largest LE; Sano-Sawada algorithm[23] calculating the local Jacobian matrix and yielding all LEs. For simplicity and requirements of verifying chaos existing, which demand largest LE, we choose Wolf's method. The calculating steps are:

Step 1: Choose a initial point in trajectory in phase space noted as $\mathbf{Y(t_0)}$, set the closest point as $\mathbf{Y_0(t_0)}$. The distance between them is noted as $L(t_0)$. Tracing the evolution of these two points from time node t_0 to time node t_1 the distance between these two points goes beyond a given setting ε, that is

$$L'(t_1) = |\mathbf{Y(t_1)} - \mathbf{Y_0(t_1)}| > \varepsilon, \varepsilon > 0 \qquad (2)$$

Step 2: Maintain point $\mathbf{Y(t_1)}$ and find the closest point $\mathbf{Y_1(t_1)}$. Ensuring not only the distance between these two points under the given setting ε as below but also the intersection angle θ of $L(t_1)$ and $L'(t_1)$.

$$L(t_1) = |\mathbf{Y(t_1)} - \mathbf{Y_1(t_1)}| < \varepsilon, \varepsilon > 0 \qquad (3)$$

TABLE I

The largest Lyapunov exponent of 16 finals

[a]	[o]	[e]	[i]	[u]	[ü]
0.0209	0.0310	0.0250	0.019	0.024	0.028
[ai]	[an]	[ang]	[ao]	[ei]	[en]
0.26912	0.13181	0.37439	0.5415	0.3172	0.5614
[eng]	[mg]	[ong]	[ou]		
0.11179	0.2673	0.3941	0.6813		

TABLE II

The largest Lyapunov exponent of 23 initials

[b]	[d]	[g]	[q]	[s]	[r]
0.583	0.577	0.618	0.819	0.866	0.110
[p]	[t]	[k]	[x]	[zh]	[w]
0.327	0.416	0.394	0.723	0.378	0.085
[m]	[n]	[h]	[z]	[ch]	[y]
0.141	0.126	0.543	0.383	0.741	0.073
[f]	[l]	[j]	[c]	[sh]	

Step 3: Repeat step 1 and 2 and mark the repeating count as M until $\mathbf{Y(t)}$ reaches the end of trajectory. The Largest Lyapunov exponent is

$$\lambda_1 = \frac{1}{t_M - t_0} \sum_{k=1}^{M} ln \frac{L'(t_k)}{L(t_k - 1)} \tag{4}$$

Theoretically, Wolf's method can conclude all LE exponents respecting noiseless and infinite data. But the reality is that time series can't be infinite and noise can't be eliminated leading to alone reliable LE-the largest one.

Mandarin consists of 16 finals and 23 initials. In this paper, we calculate the largest Lyapunov exponents of these 39 phonemes using Wolf method.

The results of all the mandarin phonemes are positive as shown in Table I and II, which show that the chaotic mechanism does exist in mandarin. Chaos and fractal is internally correlated with each other. The phase space structure, which is often named as strange attractor, is fractal characteristics. As we will be addressed below, fractal theory provides various measurements to describe the structure which Euclidean geometry fails to.

III. Multifractal Definition

Various perspectives and summaries that introduce multifractal are provided in[25]. To understand the essence of the multifractal, fundamentals about measures theory and metric topology are necessary which cause difficulties for those who are not familiar with related knowledge. The concept of fractals defines a structure mostly formed by nature with the property of self-similar and scale invariant. The definition of multifractals can be viewed as a generalization of monofractals[23]. Thus, a multifractal process $X(t)$ satisfies the following rule $X(ct) = c^H X(t)$, where c^H represents the scaling factor which can be viewed as fractal dimension in general sense and indicates the scale invariant property. For multifractal, H is a variable appearing as spatial and temporal variation in scale invariant structure, while H is a constant in monofractal and defines a single power law exponent and assumes that the scale invariance is independent on time and space.

Fractal dimension derives many methods to adjust various issues, such as Hausdorff dimension, correlation dimension, box-counting dimension, information dimension, etc.. Previous works have proved the fractal characteristics existing in speech through evaluation of above mentioned dimension.

However, to deal with multifractal cases, those methods are straight-forward but require high computation. Another choice is the standard partition function multifractal formalism, but it fails at correctly estimating for nonstationary time series that are affected by trends or that cannot be normalized[23]. In the following section, a method named MFDFA is performed and multifractal characteristics of speech varying among these four emotions are analyzed.

IV. Estimation and Analysis of Multifractal Characteristics

Multifractal DFA was derivative of DFA-detrended fluctuation analysis, which is developed to address the problem of accurately quantifying long-range correlations in nonstationary fluctuating signals[26]. According to[24], the generalized multifractal DFA procedure consists of five steps. The first three steps are essentially identical to the conventional DFA procedure. We will conduct the five steps as following:

Firstly, convert a time series into a random walk like time series:

$$y(i) = \sum_{i=1}^{i} (x(t) - \bar{x}), i = 1, 2, ..., n \tag{5}$$

Where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x(t)$, n refers to the length of time series.

Secondly, to calculate the local root-mean-square variation in step 3, clap the time series into segments. Divide $y(i)$ into M non-overlapping segments with equal length of s. In order to cover all the data of $y(i)$, divide $y(i)$ from beginning to the end and from the end to beginning respectively. At last, we obtain $2M$ segments.

Thirdly, slow varying trends exist in speech time series and detrending of the signal is necessary to quantify the scale invariant structure of the variation around these trends.

$$F^2(s, v) \equiv \frac{1}{s} \sum_{i=1}^{s} \{y[(v-1)s + i] - y_v(i)\}^2 \tag{6}$$

$$F^2(s, v) \equiv \frac{1}{s} \sum_{i=1}^{s} \{y[n - (v - M)s + i] - y_v(i)\}^2 \tag{7}$$

Linear, quadratic, cubic, or higher order polynomials can be used in step 3 and differ in their capabilities of eliminating trends in the series.

Fourthly, in the multifractal time series, local fluctuation $y_v(i)$ will be extremely large for segments within the time periods of large fluctuations and extremely small for segments within the time periods of small fluctuations[28].

Consequently, the multifractal time series are not normal distributed and all q-order statistical moments should be considered.

$$F_q(s) \equiv \{\frac{1}{2M}\sum_{v=1}^{2M}[F^2(s,v)]^{q/2}\}^{1/q} \quad (8)$$

Fifthly, the Hurst exponent mentioned in definition of multifractal can be determined by analyzing log-log plots $F(s)$ versus s for each value of q.

$$F_q(s) \sim s^{h(q)} \quad (9)$$

Note that the size of segments influences F in step 4, since fast changing fluctuations in the time series will influence F for segments with small sample sizes whereas slow changing fluctuations will influence F for segments with large sample size. Thus, the scaling function F, should be computed for multiple segments sizes to emphasize both fast and slow evolving fluctuations that influence the structure of the time series[24].

$F(s)$ vs. s is plotted to present the corresponding relation between overall $F^2(s,v)$ and s as figure 3, which obviously is approximately linear relationship. Hurst exponent-the s-lopes of the regression line is constant as scale s increasing in the term of q, which illustrates the scale invariance structure of emotional speech time series. Besides, it is visually apparent that the difference between $F(s)$ for positive and negative q with the small segment sizes is larger than $F(s)$ with large segment sizes, for large segment covers several local periods with both small and large fluctuation and therefore average out their difference while small segment are embedded within these periods[28].

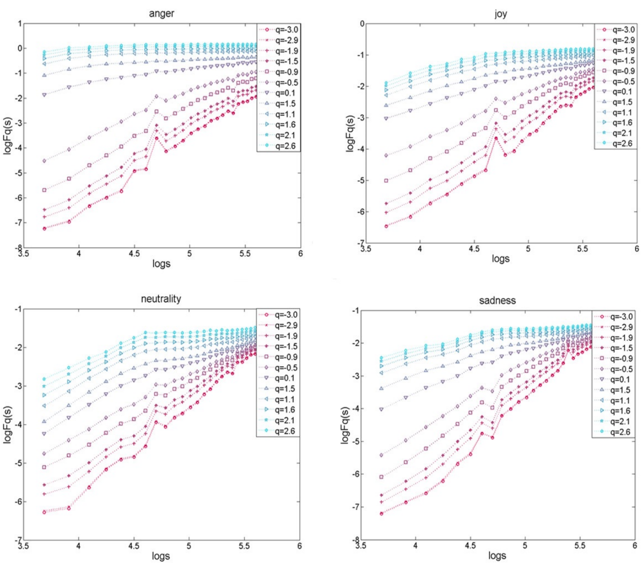

Fig. 3. $\log F_q(s) \sim \log s$ of four emotions spoken by the same utterance

Further, we explore the difference between q-order H for positive and negative q, as figure 4 shows. For monofractal time series, $h(q)$ is independent of q, since the scaling behavior of the variances $F^2(s,v)$ is identical for all segments v, and the averaging procedure in Eq.(8) will give just this identical scaling behavior for all values of q. Only if small and large fluctuations scale differently, there will be a significant dependence of $h(q)$ on q[24]. All of these is consistent with the statement that Hurst exponent here is a variable, proving that emotion speech is a multifractal process.

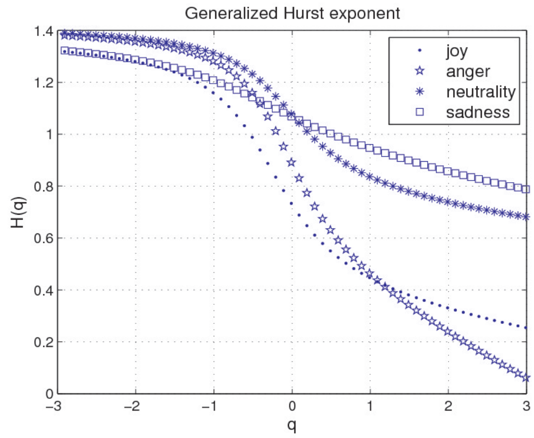

Fig. 4. Generalized Hurst Exponent $H(q)$ vs.q of four emotions

The transformation from Hurst exponent $H(q)$ to singular spectrum $f(a)$ can be obtained from some techniques such as coarse graining spectrum, Hausdorff spectrum, and Legendre spectrum[28]. Due to its simplicity, we choose Legendre spectrum, the transformation shows as Eq.(10). Multifractal, which is called fractal measures from the perspective of measure theory, describes various local probability distribution $f(a)$ under singular exponent a. The multifractal spectrum $f(a)$ vs. a is shown in figure 5.

$$\begin{cases} a(q) = \frac{d\tau(q)}{dq} \\ f(a) = q \cdot a(q) - \tau(q) \end{cases} \quad (10)$$

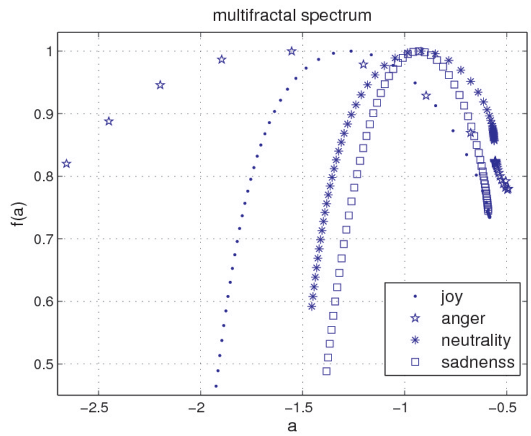

Fig. 5. Multifractal spectrum curve of four emotions spoken by the same utterance

The differentiation of the width and height of multifractal spectrum curve with respect to these four emotions distinguishes these four emotions and demonstrates the inner fractal structure of speech time series varies as emotion state changes.

An easy and comprehensive way to understand a and $f(a)$ is that a represents the singular characteristics of an asymmetry area within the underlying fractal structure of speech time series, $f(a)$ is the fractal dimension of the corresponding area. It is straight-forward that curve crossing larger range of a means the fractal structure has much more 'layer', i.e. anger's fractal structure is more 'complex' than the other three, which is identical to our daily life experience-anger arises from more 'emotional' speaking state, and fluctuates both in loudness and speaking rate which leads to turbulence occuring easily along vocal tract which contributes the inner complexity of fractal structure.

All of the four emotions' multifractal curves reach the same peak at different a, which indicates two points:

1. statistics self-similarity, which is defined to describe non-strict mathematical fractal phenomenon but widely distributed in various contexts, such as geophysicist, medical science, target detection, signal processing, of these four emotional speech obeys the speech signal statistics discipline which results in the almost same peak value. To illustrate this, we can plot the speech signal waveform of these four emotional speech together to compare, as we zoom in and obtain subsequent plot from the original one by increasing the time resolution by a factor of 10 and by concentrating on a randomly chosen subinterval, the waveform in a subtle scale is periodic and repeats itself as we divide and observe, which is caused by the speech production mechanism, for example, periodic vibrating of the vocal folds, the similar structure of vocal tract ranging from person to person.

2. the spectrum curve reaches peak at different a indicates that the local characteristics compared to the aligned integral self-similarity is distinguishable varying from one emotion to another. a is called singularity exponent, which describes the probability distribution of the fractal structure. Although self-similarity of emotional speech over all these four emotions is the same, the argument a of fractal dimension $f(a)$ means what kind of singularity contributing and dominating the fractal structure finest part.

Note that the singular spectrums of these four emotions are not intensively related with the our physical perception of emotional speech, like speaking rate, loudness, intensity, even the speaking content. Just like the wikipedia notes[2], the technique entails distorting dataset extracted from patterns to generate multifractal spectra that illustrates how scaling varies over the dataset. It seems abstract and hardly related to the intuitive perceived way of emotion speech, but the fractal theory provides another perspective to view emotion.

V. Experimental Investigation

The corpora source we choose is Beihang University Database of Emotional Speech[3], which consists of 6,300 samples recorded by 15 speakers.20 short sentences are repeated 3 times for each of the 7 kinds of emotions, namely joy, anger, disgust, fear, sadness, surprise, anger and neutrality. 3456 mandarin utterances which are accurately recognized by at least 70% of the strange listeners are collected into the experimental corpora. For copyright-related reason, we only receive three person's subsets, that is 720 utterances. 480 utterances are used to train while the rest are utilized for testing.

Conventional feature extraction methods generally can be divided into two classes-acoustics and linguistics. As linguistics is more about natural language understanding and gaining considerably importance towards spontaneous real-life speech rather than acted data. As we aim at verifying the efficiency of fractal features, we choose several commonly used acoustics, i.e. pitch, format, intensity, speaking rate, and MFCC. Multifractal features is extracted from multifractal spectrum, i.e. the maximum, minimum, range, average, variance, upper quartiles and lower quartiles of curves, respectively. Parts of the parameters describing the curves are shown in Table III.

Our experiments are implemented in MATLAB. LIBSVM[29] from C. Lin is used and linear kernel function is chosen. SVM is based on structure risk minimization and considers the sample error and the model complexity while BP network is based on empirical risk minimization and easy to trap in local optimum. Previous work has been utilized ensembles of different classifier to compensate weakness and strengthen the advantages. Thus, Adaboost algorithm is chose to combine these two.

BP neural network used in our experiment consists of three layers, with the same feature number as the number of input layer, ten nodes as the middle layer, four nodes with the same number of emotion classified as the output layer.

The experiments compare classical features and multifractal features using BP neural network and SVM respectively with cross validation. The result can be found in Table IV.

At last, the recognition accuracy utilizing Adaboost turns out to be anger 83.33%, joy 83.33%, sadness 80.00%, neutrality 86.67%

From Table IV, it can be concluded that classification using multifractal features improves the average accuracy 1.5% with BP neural network, and 5% with SVM. When combining this two method using Adaboost utilizing both traditional and multifractal features, the accuracy improves with 3.5% averagely.

VI. Dicussion/Conclusion

We present a novel approach that analyzing and calculating fractal underlying emotional speech and providing a effective feature extracted from multifractal spectrum for emotion recognition. Different from previous emotion recognition method, we regard emotional speech in another way rather than physically related speech production and listening mechanism. The proposed approach brings a significant contribution to the typical emotional speech model, and outperforms the conventional features because the distinction of multifractal spectrum among various emotions without

[2] http://en.wikipedia.org/wiki/Multifractal_system#techniques

[3] http://www.ee.buaa.edu.cn/oldeeweb/html/zykj/teachers/mx/news/22.html

TABLE III

Parameters of multifractal spectrum curve of four emotions

	a_{max}	a_{min}	Δa	$f(a_{max})$	$f(a_{min})$	Δf
joy	-0.582	-1.910	1.328	1.000	0.476	0.524
anger	-0.491	-2.623	2.132	1.000	0.802	0.198
neutrality	-0.561	-1.454	0.894	1.000	0.592	0.408
sadness	-0.590	-1.383	0.792	1.000	0.489	0.511

TABLE IV

Recognition accuracy comparing with conventional features vs. multifractal features

	anger	joy	sadness	neutrality	average
BP+conventional	80.00%	78.33%	75.00%	80.00%	78.33%
SVM+conventional	80.00%	78.33%	75.00%	76.67%	77.50%
BP+multifractal	80.00%	85.00%	76.67%	78.33%	80.00%
SVM+multifractal	83.33%	85.00%	78.33%	83.33%	82.50%
Adaboost+both	83.33%	83.33%	80.00%	86.67 %	83.33%

the influence of gender and speech content.It also proves that the emotional speech signal is multifractal instead of monofractal. Excepting computational cost is higher in Matlab platform, this method is worth further studying in its availability applied in other language and deeper functioning mechanism.

REFERENCES

[1] A. Damasio, Descartes Error: Emotion, Reason, and the Hitman Brain. London, U.K.: Putman, 1994.

[2] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor, Emotion recognition in human-computer interaction, IEEE Signal Process. Mag., vol. 18, no.1, pp.32-80, Jan. 2001.

[3] D. Roy and A. Pentland, Automatic spoken affect analysis and classification, in Proc. Int. Conf. Automatic Face Gesture Recognition, Killington, VT, pp. 363-367, 1996.

[4] J. Averill, "In the eyes of the beholder", The Nature of Emotion: Fundamental Questions, pp. 7-14, 1994.

[5] M. You, C. Chen, J. Bu, J. Liu, J. Tao, "Emotion recognition from noisy speech", In: Proceedings of ICME, Toronto, ON, July 9-12, pp 1653-1656.

[6] C.A. Martinez, A. Cruz, "Emotion recognition in non-structured utterances for human-robot interaction", In: IEEE international workshop on robot and human interactive communication, August 13-15, pp 19-23, 2005.

[7] B. Schuller, A. Batliner, D. Seppi, S. Steidl, T. Vogt, J. Wagner, L. Devillers, L. Vidrascu, N. Amir, L. Kessous, V. Aharonson, "The relevance of feature type for the automatic classification of emotional user states: low level descriptors and functional", In: Proceedings of the Interspeech, Antwerp, Belgium, August 27-31, pp 2253-2256,2007.

[8] Y. Kim, H. Lee, E. M. Provost, "Deep learning for robust feature generation in audiovisual emotion recognition", In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP'13). Vancouver, British Columbia, Canada. May, 2013.

[9] J. F. Kaiser, "Some observations on vocal tract operation from a fluid flow point of view", In: I. R. Titze, R. C. Scherer, (Eds.), Vocal Fold Physiology: Biomechanics, Acoustics and Phonatory Control, Denver Center for Performing Arts, Denver, CO, pp.358-386.

[10] H.M. Teager, S.M. Teager, "Evidence for nonlinear sound production mechanisms in the vocal tract", In: Hardcatle, W.J., Marchal(Eds.), Speech Production and Speech Modelling NATO ASI Series D, Vol. 55.

[11] T.J. Thomas, "A finite element model of fluid flow in the vocal tract", Comput. Speech Language 1, 131-151, 1986.

[12] P. Maragos, J.F. Kaiser, T.F. Quatieri, "Energy separation in signal modulations with application to speech analysis", IEEE Trans, Signal Process. 41(10), 3024-3051, 1993.

[13] S. Narayanan, A. Alwan, "A nonlinear dynamical system analysis of fricative consonants", J. Acoust. Soc. Am. 97(4),2511-2524,1995.

[14] A. Kumar, S. K. Mullick, "Nonlinear dynamical analysis of speech", J. Acoust. Soc. Am. 100(1), 615-629,1996. [15] P. Marogos, "Fractal Aspects of speech signal: dimension and interpolation", In: Proc. IEEE ICASSP, vol.1, pp. 417-420,1991.

[15] V. Pitsikalis, P. Maragos, "Analysis and classification of speech signals by generalized fractal dimension features". Speech Communication 51(2009) 1206-1223,2009.

[16] P. Maragos, "Fractal dimension of speech sounds: Computation and application to automatic speech recognition", J. Acoust. Soc. Am. 105(3), March 1999.

[17] D. C. Gonzalez, L. L. Lee, F. Violario, "Use of multifractal parameter for speaker recognition", M. Eng. Thesis, FEEC/UNCAMP, Campinas, Brazil, 2011.

[18] A. Langitt, K. Soeminatapurat, W. Kinsners, "Multifractal processing of speech signals information, communication and signal processing", In: Han, Y., Quing, S.(eds.) ICICS 1997. LNCS, vol.1334, pp. 527-531. Springer, Heidelberg, 1997.

[19] W. Kinser, W. Grieder, "Speech segmentation using multifractal measure and amplification of signal feature", In: Proc. 7th IEEE Int. Conf. on Cognitive Informatics(ICCI2008), PP.351-357,2008.

[20] J.A. Russell, "A circumplex model of affect", Journal of Personality and Social Psychology, Vol.39, No.6, 1161-1178,1980.

[21] YE Jixiang, WANG Conghui,"Speech emotion recognition based on multifractal[J]". CEA, 48(13): 186-189.2012.

[22] A. Wolf, J. Swift, H.Swinney, J. Vastano, "Determining Lyapunov exponents from a time series," Physica D 16, 285-317.

[23] M. Sano, Y. Sawada,"Measurement of the Lyapunov spectrum from a chaotic time series," Phys. Rev. Lett.55, 1082-1085.

[24] B. B. Mandelbrot, "Multifractal measures, especially for the geophysicist", Pageoph, 131, 1/2, 5 42, 1989.

[25] R. H. Riedi, M. S. Crouse, V. J. Ribeiro, R. G. Baraniuk, "A multifractal wavelet model with application to network traffic," IEEE Trans. on Information theory 45(3), 992-1018, 1999.

[26] J. W. Kantelhardt, S. A. Zschiegner, E. K. Bunde, et al., "Multifractal detrended fluctuation analysis of nonstationary time series", Physica A:Statistical Mechanics and its Application, 316(1/4): 87-114.2002.

[27] P. A. Varotos, N. V. Sarlis, E. S. Skordas, "Detrended fluctuation analysis of the magnetic and electric field variations that precede rupture", chaos 19,023114(2009).

[28] Espen A. F. Ihlen, "Introduction to multifractal detrended fluctuation analysis in Matlab", Volume 3, Article 141, frontiers in physiology, 2012.

[29] C. Chang and C. Lin. LIBSVM: a Library for Support Vector Machine. Available: http://www.csie.ntu.edu.tw/ cjlin/