A Chaotic Time Series Prediction Model for Speech Signal Encoding Based On Genetic Programming

Lei Yang³, Junxi Zhang⁴, Xiaojun Wu^{1, 2, 3*}, Yumei Zhang^{2, 3}, Jingjing Li¹

1. Key Laboratory of Modern Teaching Technology, Ministry of Education (Shaanxi Normal University), Xi'an, 710062, China
2. School of Computer Science, Shaanxi Normal University, Xi'an, 710062, China
3. School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China
4. Department of Vehicle Engineering, Xi'an Aeronautical University, Xi'an, 710077, China

Abstract

In this paper, a novel solving method for speech signal chaotic time series prediction model was proposed. A phase space was reconstructed based on speech signal's chaotic characteristics and the genetic programming (GP) algorithm was introduced for solving the speech chaotic time series prediction models on the phase space with the embedding dimension m and time delay τ . And then, the speech signal's chaotic time series models were built. By standardized processing these models and optimizing parameters, a speech signal's coding model of chaotic time series with certain generalization ability was obtained. At last, the experimental results showed that the proposed method can get the speech signal chaotic time series prediction models much more effectively, and had a better coding accuracy than linear predictive coding (LPC) algorithms and neural network model.

Keywords: Chaotic Time Series Prediction; Genetic Programming; Nonlinear Coding Model.

1. Introduction

The Linear Prediction Coding (LPC) has been widely used in speech signal processing field. It is based on the correlation between the speech samples, using a linear model, with the past p sample values to predict the sample values present or future. Then, a large number of studies on non-linear characteristics of speech signal found that speech signal and the discrete speech signal sequence showed a very complicated nonlinear process and obvious chaotic characteristics^[1-4]. Nonlinear science inspired the study of speech signal which can be analyzed by non-linear methods. The research of nonlinear models for speech signal coding became a hot spot of speech signal processing, and using chaos theory to research speech signal has made some achievements^[4-6]. It is proved that the reconstructed speech signal helped to analyze its characteristics deeply and accurately^[6]. Therefore, the introduction of chaotic time series analysis theory has very important theoretical and practical values for not only researching the chaos characteristics and the processing methods of speech signal, but also constructing nonlinear speech signal processing

* Corresponding author.
E-mail address: xjwu@snnu.edu.cn (Xiaojun Wu).

models.

The neural network model is the most commonly used method for nonlinear speech signal processing, and has some achievements.^[7-9] It has been applied to speech recognition^[10], speech transmission^[11] etc.

Thyssen et al.^[12] holds that the speech signal was the combination of linear and nonlinear function. Linear prediction method has been used for modeling the linear part. They proved that the residual signal has obvious nonlinear characteristics, and used Volterra and neural network methods to modeling the residual part which is nonlinear part of the signal. The combination of those linear models and nonlinear models has improved the accuracy of coding. Atthew K. Luka established a local ethnic languages recognition system through the neural network technology^[13]. This system uses Mel-frequency Cepstral Coefficients of the speech signal that extract from Mel filter banks as the input of the multilayer neural network. They combined the conjugate gradient backward propagation algorithm with neural network to improving the convergence rate, which shows good recognition ability. Using chaos theory to analyze speech signal, Lee and Tong^[14] and Aina et al.^[15] applied the traditional local linear and radial neural network to the analysis of chaotic time sequence, achieved two kinds of nonlinear speech coding predictors. The result showed that the predictor established in the reconstructed phase space had significantly improved compared with the linear predictor.

The neural network model has been successfully used in the speech signal coding, and also provide a structured prediction model. In recent years, it has made some achievements in the nonlinear prediction of the speech signal sequence. But the structure of the neural network model is based on the optimization of the weights; it is not conducive to the speech signal's analysis and processing as it can't provide the fixed structure for the models to describe the speech signal like LPC, and it needs to rebuild for different samples. So its practical application has been limited. To solve these problems, this paper introduces GP (Genetic Programming) as solving method for the speech signal prediction models. Comparing to neural network method, GP can obtain the explicit model structure which can facilitate its analysis and application. As early as 1998, Conrads et al^[16] used GP in speech, they found that GP could find programs to discriminate certain spoken vowels and consonants. Xie H and Zhang M^{[17][18]} did a series of research in rhythmic stress detection in spoken English using GP. Later, GP was proposed to construct nonlinear speech coding models^[19].

GP is a special optimization algorithm developing from GA (Genetic Algorithm)^[20]. It uses hierarchical structure to describe programming problems. Issues in different fields are attributed to computer program that searching for solutions to meet fixed constraint conditions. GP algorithm can take advantage of limited samples to discover parallel estimate of the model structure and its parameters. In recent years, it has been widely used in nonlinear system modeling^[21], data analysis^[22] and other areas.

Based on the following four reasons GP can be used to solve the problem of speech signal processing. Firstly, compared with LPC, GP highlight the non-linear characteristics. Secondly, compared with the Artificial Neural Network, it can obtain the explicit model structure, which is facilitated for its analysis, optimization and application. Thirdly, Compared to other non-linear models, using GP models, we can achieve the overall modeling without considering the linear or non-linear part. Forth, Compared with the Volterra model, it overcame the weakly nonlinear problem of the models, and it is more streamlined and more efficient.

Considering chaotic characteristics of the speech signal, this paper provides an improved genetic programming algorithm to establish the non-linear model of the speech signal; Through

the model analysis, a model structure which has generalization ability can be chosen. By optimizing parameters, speech signal encoding could be achieved and it provides a new model for speech signal's nonlinear analysis.

2 GP Algorithm for Nonlinear Modeling

Compared with other algorithms, the GP algorithm can better gain a nonlinear model with minimum error, and it doesn't need any prior knowledge. The evolutionary process changes the structure of the model automatically through crossover and mutation, calculates the fitness of new individuals by the constructed fitness function and keeps the optimal one, to make the evolution direction towards the better target. As the model that is produced by GP algorithm has explicit structure, it can intuitively describe the problem to be solved to a large extent, so it is widely used in the problem of nonlinear modeling, especially in time series prediction model^{[14][23]}.

Lee and Tong separated complete time series into linear and nonlinear parts according to the analysis of the time series' composition, built the prediction model of linear part using ARIMA algorithm and a nonlinear model with the GP algorithm for the residual sequence^[14]. The nonlinear model that describes the time series is obtained and it is effectively improved than linear method. It benefits the combination of these two kinds of models. Wagner et al. separated the time series environment into static environment and dynamic environment, and applied the GP algorithm to the dynamic environment (DyFor GP), because they believe that the traditional GP algorithm does not consider the dynamic factors in the environment when predicting time series^[22]. In the evolutionary process, the method changes and adjusts the order dynamically depending on the different environment to make the model be more optimized. Estevez et al.^[24] researched the identification of the speech signal and noise signal using GP, the function set contains some logical operators such as AND, OR, NOT, GT(greater-than). The composition of Terminator set is the constant R and the terminator defined in ITU-T G.729B VAD standard. Finally, the judgment method of speech signal was presented. Compared with standard G.729B, the judgment of speech signal and non-speech signal by using the method is more accuracy.

GP algorithm solves the nonlinear problems effectively, and its explicit model structure is more conducive for data analyzing and processing. This paper works on the speech signal coding model with GP for getting more valuable model structure.

3 Proposed Speech Coding Method

The speech signal is a chaotic system that produces orders and laws from disorder and complexity. The chaotic attractor is one of the chaotic characteristics indicating the regularity of a chaotic system. The inherent regularity of the chaotic system shows that it's predictable. This paper uses chaos system's internal regularity, combined with nonlinear prediction method (GP) to predict and build models of the speech signal.

In this paper, the speech signal processing flow chart, as shown in Fig.1, include five contents:

- (1) Process the speech signal, enhance high-frequency part of the speech signal through the pre-emphasis, and then use the phase space reconstruction technique to reconstruct the discrete sequence.

- (2) According to the characteristics of the nonlinear model of the speech signal to improve the GP algorithm, and solve the "best" non-linear model of each frame speech signal by using the improved GP algorithm.

- (3) Summarize and analyze the prediction model in the different samples, and get a model

structure with good generalization ability.

(4) On the basis of model structure, use DUPSO optimization algorithm to solve the prediction parameters of each frame of data.

(5) Evaluate the coding model in this article by experiments.

On the basis of the speech signal's chaotic characteristics, this paper provides a method for solving speech signal's nonlinear model by improved GP algorithm. As a result of the limitations of samples used in the modeling process, the model structure may only apply to the limited features speech samples. For a large number of speech samples with different characteristics need different models to describe their structure, this study tries to make the model for the sample has adaptive capacity (generalization ability). But more samples still need a different classification for models. Therefore, it's a problem which needs further study to build classified model for speech sample characteristics, and needs more scientists to participate in. With the development of research, better description for the nonlinear model of speech signal will be found and applied.

Fig 1. Structure diagram of this research

4 Prediction Model of Speech Signal based on Improved GP

Algorithm

As using the traditional linear prediction model for speech coding, the model structure is fixed. The encoding and decoding of speech signal can only transmit different parameters which will be decoded at the receiver.

Wu and Yang ^[19] makes improvements in two ways on the GP algorithm. First, in the initialization of the population, uses a variety of groups, in order to increase the diversity of solutions and improve the global search capability. Second, introduces the hill-climbing algorithm for further optimize model parameters which are based on structural optimization, so the GP algorithm gets more objective evaluation of the model structure in the evolutionary process.

In this paper, after the speech signal pre-processing and reconstruction phase space, we use that improved GP algorithm to modeling each frame data, and then analyze the structure of the model, select one or a group of standardized models with better generalization ability. Finally, optimization algorithm is used to optimize the parameters of the selected model in each frame of speech data. Then the coding of the speech signal is finished.

4.1 Speech Signal Pre-processing and Phase Space Reconstruction

4.1.1 Speech Signal Pre-processing

The sound produced by vocal fold vibration is called speech, otherwise is called speechless. Speech includes all vowels and some consonants, and part of the consonants included in speechless. In this paper, we only consider speech in the signal while establish the model of the speech signal by the GP algorithm.

For the average power spectrum of the speech signal is influenced by the glottal excitation

and Oronasal radiation, high frequency more than about 800 Hz drop with 6dB / octave, so it needs the pre-emphasis in the pretreatment. The purpose of pre-emphasis is to enhance the high frequency part and make the spectrum of the signal more flat. So that it is easier spectrum analysis. In this paper, we make Pre-emphasis after the A/D conversion. Realize by the first-order digital filter which has the improved frequency characteristic with 6dB / octave, as (1) shows:

$$H(z) = 1 - \mu z^{-1} \quad (1)$$

The de-emphasis works in the decode stage of the speech signal. A frequency characteristic of 6dB/octave decreased is used to restore the original characteristics.

4.1.2 Phase Space Reconstruction

The phase space is a space used to describe all the possible states of a system; each point represents the entire physical state, including all the details of the system (position and momentum coordinates of each particle of the system). For speech signal, firstly, recover its phase space using phase space reconstruction theory, and then use the improved GP algorithm to construct non-linear model of speech signals. In this paper we use the delay method to reconstructs the phase space:

$$\mathbf{X}_k = [x_k, x_{k+\tau}, \dots, x_{k+(m-1)\tau}]^T \quad (2)$$

x_k is the discrete variable of the system; τ is time delay, m is the embedding dimension. According to the Takens embedding theorem, as long as the embedding dimension m meet $m > 2n$, n is the fractal dimension of system attractor.

Traditional methods of calculating the embedding dimension mainly include false nearest neighbor method (FNN)^[25], Cao method^[26] and singular value decomposition (SVD)^[27], etc. And the methods calculating the time delay mainly include C-C method^[28], the auto correlation function method^[26], the average displacement (AD)^[29] and mutual information method^[30] etc. In this paper, Cao method and mutual information method is used to calculate embedding dimension and time delay.

λ is Lyapunov exponent which is the effective features for describing the nonlinear dynamic characteristics. When $\lambda \leq 0$, the trajectory of phase space contracts, system is not sensitive to initial conditions, and there is no chaos. When $\lambda > 0$, the trajectory of phase space is separated, long time dynamic behavior is sensitive to initial conditions, this is the chaotic state. This paper uses the small-data method to calculate Lyapunov exponent^[29]. Table 1 shows the four samples' embedding dimension, the time delay and the value of Lyapunov exponent.

Figure 2 is the original speech time series waveform graph of [ɔ:], Figure 3 is its corresponding two-dimensional phase space graph (we only made 300 samples in its trajectory).

Table 1. Reconstructed phase space's vectors

phoneme	Embedding dimension	Time delay	Lyapunov exponent
[ai]	10	2	2.48
[ɔ:]	10	2	1.88
[a:]	10	2	3.56
[əu]	11	2	3.96

Fig 2a. The waveform of [ɔ:]

Fig 2b. The 2-D phase space graph of [ɔ:]

4.2 Solving Process of The Speech Signal Prediction Model

Assume $x(t), t = 1, 2, \dots, N$ is the discrete-time sequence of the speech signal, the form of state transition of the m -dimensional phase space is:

$$\mathbf{Y}(t+1) = f(\mathbf{Y}(t)) \quad (3)$$

Where $\mathbf{Y}(t)$ is the point in phase space, τ is time delay and

$$\mathbf{Y}(t) = (x(t), \dots, x(t + (m-1)\tau)) \quad (4)$$

Expand equation (4)

$$(x(t+1), \dots, x(t+1 + (m-1)\tau)) = f(x(t), \dots, x(t + (m-1)\tau)) \quad (5)$$

Using the improved GP algorithm for solving nonlinear model of the speech signal is the process of looking for mapping function f of equation (3). The process uses the previous one or several vectors that were reconstructed as input to predict the latter vectors. In forecasting process, through continuous evolution, crossover and mutation, we make the final mapping function f more effectively in predicting the reconstructed vectors.

In the reconstructed phase space, the former $m-1$ components of vector number $n + \tau$ are the latter $m-1$ components of vector number n (τ is the time delay, m is the embedding dimension). For example, the time series $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, assuming the embedding dimension is 4, the time delay is 2, then its reconstructed phase space's vectors shows in Table 2. In Table 2, the former three components of vector $\mathbf{X}_3$ are the latter three components of vector $\mathbf{X}_1$, the first three components of vector $\mathbf{X}_4$ are the former three components of vector $\mathbf{X}_2$. In this paper, the initial vectors of the system are the τ vectors of the previous, and each components of the former vector as input, GP algorithm randomly generated individuals as the prediction model to predict the last component of vector $\tau+1$, and then used the second vector as input to predict the last component of vector $\tau+2$, and so on. Until reach the required prediction steps. Finally, according to the fitness of each prediction model, the best one is selected as the final prediction model.

Transform the equation (5) get the function y

$$x(t + m\tau) = y(x(t), x(t + \tau), \dots, x(t + (m-1)\tau)) \quad (6)$$

Table 2. The value of 3 parameters of samples

Vector number	The components of vector			
$\mathbf{X}_1$	1	3	5	7
$\mathbf{X}_2$	2	4	6	8
$\mathbf{X}_3$	3	5	7	9
$\mathbf{X}_4$	4	6	8	10

4.3 Analysis of the Model

After pre-processing the speech signal, the nonlinear prediction model for each frame of data is built based on the improved GP algorithm. For each frame of data, GP algorithm is able to obtain a model with the optimal prediction accuracy. By comprehensive analysis of these models considering their structural features, the more concise structure and similar models are selected for normalization. The (7) (8) and (9) are the best models for the selected samples (a_i is coefficient, x_i is input variables).

$$y = \sin(a_1 x_8) + \sin(a_2 x_9) + a_3 x_0 + a_4 x_1 + a_5 x_3 + a_6 x_6 + a_7 x_3 \sin(a_8 x_1) \quad (7)$$

$$y = \sin(a_1 x_1) + \sin(a_2 x_3) + \sin(a_3 x_5) + \sin(a_4 x_8) + a_5 x_2 + a_6 x_7 + a_7 x_8 + a_8 x_9 \quad (8)$$

$$y = \sin(a_1 x_3) + \sin(a_2 x_5) + \sin(a_3 x_7) + \sin(a_4 x_7 + a_5 x_1) + a_6 x_4 + a_7 x_8 + a_8 x_9 \quad (9)$$

Observed above three formulas, the basic form of the prediction model is composed of $\sin(a_i x_i)$ and $a_i x_i$, but also includes heterogeneous structures such as the formula(7) and formula(10): $a_j x_j \sin(a_i x_i)$ and $\sin(a_i x_i + a_j x_j)$. In this paper, the embedding dimension is m , the complex structures of the prediction model are approximately equivalent to the reference model structure as follows by replacing with $\sin(a_i x_i)$ and $a_i x_i$:

$$y = \sum_{i=1}^m (\sin(a_i x_i) + a_{m+i} x_i) \quad (10)$$

After testing the sample data and optimizing the model coefficients, the results show that compared with (7), (8) and (9), the Equation (10) has similar or even better description ability, by selecting the optimal model coefficients a_i , the better speech signal's nonlinear encoding model could be obtained.

4.4 Optimization of the Model

In this paper, as the model structure has been fixed, the PSO (Particle Swarm Optimization) algorithm can be used to optimize the parameters for different speech samples. The PSO algorithm is derived from the simulation of the preying behavior of birds, and potential solutions in the D-dimensional space can be described as massless particles with no volume, and "flight" in the solution space has a certain speed. During the iterative process, each particle remembers the optimal position that themselves and the particle swarm or the limited neighbors found, then track

these two extreme by the position to update the formula by searching space in order to obtain a better solution.

In the iterative process of the traditional PSO algorithm, the particles gather together for the rapid flow of information in particles, resulting in decreased diversity of the population and got into a local optimal solution. While the population evolve to equilibrium, the evolutionary process will be stagnant. Xie et al. proposed DPSO (Dissipative Particle Swarm Optimization), which is used to exclusion process, repositioned the particle position and velocity information in accordance with a certain probability. It improved the diversity of particles, and has good results in the inhibition of precocious particle group^[30]. For the optimal parameters of this model is non-uniform distribution, in this paper we use UPSO (Uniform Particle Swarm Optimization) to optimize the coefficients of the model^[31]. By defining the searching center of the particle, and calculating the uniformity value in the process of local optimum and global optimum, the UPSO get the optimization process of uniform search particle swarm algorithm,, in (11), r is a random number between 0 and 1, c is the learning coefficient, p_i is the optimal position that is found by the memory of particle i , p_g is the optimal position that is found by the limited neighbors of particle i .

$$\begin{aligned} v_i(t+1) &= w * v_i(t) + c * [r * p_i(t) + (1-r) * p_g(t) - x_i(t)] \\ x_i(t+1) &= x_i(t) + v_i(t+1) \end{aligned} \quad (11)$$

Formula (13) is used to update the speed of particles and location information. Space dimension D is equal to the number of parameters need to be optimized in the model structure, and the location $x_i(t)$. In this paper, (9) is used to update the speed of particles and location information. We use DUPSO algorithm to globally optimize the prediction parameters. Space dimension D is equal to the number of parameters which are need to be optimized in the model structure. In the mean while, the location fitness of improved GP algorithm is equal to the minimum mean square error.

When use DUPSO to optimize the model, coding all the parameters of the model, and constitute a D -dimensional space vector. Such as formula (10), the number of input variables is m , so we need to optimize parameters of $2m$, to get D dimension space. It's clearly that $D = 2m$. The particle of DUPSO is described as a D dimension vector and a_i is described as the position information as Fig 3.

a_0	a_1	...	a_{D-1}
-------	-------	-----	-----------

Fig 3. The position information of UPSO elementary particle

Suppose there are N reference samples. There are several steps to evaluate the fitness ν of the population. Firstly, put the current position information of the particles into the formula (12) as the model coefficients, to get predicted value $\hat{x}$, where y is determined by formula (10).

$$\begin{cases} \hat{x}_i = x_i & i < D \\ \hat{x}_i = y(\hat{x}_i) & i \geq N \end{cases} \quad (12)$$

$$v = \sum_{i=0}^{i < N} (x_i - \hat{x}_i)^2 \quad (13)$$

Then, calculate the predicted value by formula (12), according to formula (13), the fitness v of the particle at the present location can be obtained.

After searching, select the particle which had the minimum fitness value as the local optimization value. Particles continually update their location information according to formula (13), thus the algorithm search in the global extent. At last, the optimal solutions are the optimal parameters of the model.

The determination of model parameters is one of the most important conditions for the evaluation of model. The algorithm used in this paper is only used for model validation. The solution for specific applications needs further study.

5 Experiments

In our experiment, the samples consist of different recording pronunciations of the given English phonemes and words with the sampling rate 8 kHz, and the data linear quantizing by the 8bits PCM. The environmental factors are strictly controlled during the recording process. So the collected speech signal has a higher signal to noise ratio.

During the experimental process, according to different speech samples, the improved GP algorithm is used to establish the nonlinear model of speech signal. The structure model is analyzed and normalized. The verification of the model structure has been found. The coding model of the data frame is obtained by solving the optimal parameters for different speech data frame based on optimization algorithm. Firstly, preprocess the speech data; secondly, calculate the embedding dimension m and time delay τ by using the phase space reconstruction techniques; and finally, we use 30ms for one frame as segment forecast the speech samples. So each sample is divided into several frames, and then reconstruct the phase space of sub-frame data with m and τ . This experiment selected four speech elements and samples data of a speech clipped for modeling, and encoded the speech data using the model structure. In this paper, we compared the proposed method with the LPC model through 3 aspects:

- 1) Compared with the original signal waveform;
- 2) Compared with the original signal error;
- 3) Compared with SNR (Signal to Noise Ratio) and MOS (Mean Opinion Score) and other methods for model evaluation.

5.1 The Solving of Predicting Model

According to the selected speech sample, using the method that was described in the fourth part, we calculate each frame data and get a group model that has the smallest error. Then analyze prediction models structures and standardize them. Finally we get a standardized nonlinear prediction model with good generalization ability.

In the process of modeling with improved GP algorithm, there are 40 populations, 100 individuals in each population, 2000 iterations for the algorithm, and 20 iterations in the hill-climbing algorithm optimization process. The crossover probability is 0.1, mutation probability is 0.2, and the crossover probability between the populations is 0.2. Evaluation

function in the evolutionary process use the mean square error, the mathematical expression is:

$$F = \frac{1}{n} \sum_{i=1}^n (y(i) - \hat{y}(i))^2 \quad (14)$$

$y(i)$ is the original value, $\hat{y}(i)$ is prediction value, n is the number of prediction data, Table 3 is the typical prediction model structure.

Table 3. The typical prediction model structure

Model Number	Model Structure
1	$y = \sin(a_1 x_8) + \sin(a_2 x_9) + a_3 x_0 + a_4 x_1 + a_5 x_3 + a_6 x_6 + a_7 x_3 \sin(a_8 x_1)$
2	$y = \sin(a_1 x_1) + \sin(a_2 x_3) + \sin(a_3 x_5) + \sin(a_4 x_8) + a_5 x_2 + a_6 x_7 + a_7 x_8 + a_8 x_9$
3	$y = \sin(a_1 x_3) + \sin(a_2 x_5) + \sin(a_3 x_7) + \sin(a_4 x_7 + a_5 x_1) + a_6 x_4 + a_7 x_8 + a_8 x_9$
4	$y = \sin(a_1 x_3) + \sin(a_2 x_4) + \sin(a_3 x_8 + a_4 x_9) * (a_5 x_8 + a_6 x_9) + a_7 x_0 + a_8 x_1 + a_9 x_5 + a_{10} x_6 + a_{11} x_7$
5	$y = \sin(a_1 x_0) + \sin(a_2 x_3) + \sin(a_3 x_9) + \sin(a_4 x_2 * a_5 x_8) * a_6 x_2 + a_7 x_0 + a_8 x_2 + a_9 x_3 + a_{10} x_6 + a_{11} x_8 + a_{12} x_9$
...	...

The model structures in Table 3 are chosen from a large number of prediction models with high similarity. As shown in the Table 4, the model structure is quite similar and be comprised of $\sin(a_i x_i)$ and $a_i x_i$ mostly. The model-based analysis that is introduced in the fourth part, replaced the heterogeneous structure with basic structure in the predicted model and completed the default variable to get standardized model structures such as (10), m is the embedding dimension of prediction data.

This paper uses DUPSO to optimize the model coefficients based on the determined model structure. In the optimization process, the size of population is 100, the space dimension is determined by the number of model coefficients. The coefficients of formula (11) chose $C = 1.3$, $w = 0.78$ from the reference [24], and DPSO select the default parameters in reference [25]. Table 4 shows the mean square error of forecast data both for the original model and the standardized model.

Table 4. The mean square error comparisons

Speech signal	model	Frame number				
		1	2	3	4	5
[ai]	original model	37.69	115.51	73.88	57.94	16.67
	standardized model	36.73	53.67	71.14	38.5	11.5
[ɔ:]	original model	175.01	385.19	763.29	480.85	276.79
	standardized model	218.2	556.29	453.19	339.06	91.14
[a:]	original model	19.47	84.8	72.21	121.99	85.42
	standardized model	25.93	128.02	68.73	90.47	78.27
[əu]	original model	46.99	90.32	136.47	237.73	

standardized model	29.47	53.94	82.72	217.2
--------------------	-------	-------	-------	-------

There are 19 samples in Table 4, where 15 of them have less mean square error which is generated by the standardized model than that of the original model. Only 4 samples have increasing mean square error. The experimental results showed that in most conditions, the standardized model will improve the accuracy of the model.

5.2 Evaluation of the Model

In this paper, a novel method of a non-linear molding was proposed, and a standardized model had been built for the collected samples. The comparative experimental results of the collected speech samples were obtained by the LPC model, radical basis function (RBF) and the established model in this paper. Waveform comparison and square error comparisons are used. Two kinds of objective evaluation methods: MOS subjective evaluation method and signal to noise ratio comparisons are also used between LPC model and the novel method.

5.2.1 Waveform Comparison

Fig 4 [ai] waveform comparison

Fig 5 [ɔ:] waveform comparison

The prediction order is set as 20 when LPC is used in linear predict coding for the speech signals. The input data were determined by the dimension of the phase space of speech signal and time delay while using specific formula (10) to predict the speech signal. The embedding dimension of [ai] and [ɔ:] is 10, the time delay is 2, then the input data is first two sets of vector after the phase space reconstruction. RBF neural network is a kind of three-layer feedforward neural network. In this experiment, input layer is the phase space reconstruction of speech signal vector, hidden layer selected by default, the output is the predicted speech signals. [ai] and [ɔ:] respectively take 1200 in these experiments. It can be obtained from Fig.4 and Fig.5 that: the waveform of the speech signal produced by the standardized model is close to the waveform of the original sample, and the other samples have the similar results, so compared with the LPC model and RBF, the standardized model has better prediction accuracy.

5.2.2 Square Error Comparisons

Fig 6 [ai] mean square error comparisons

Fig 7 [ɔ:] mean square error comparisons

Fig. 6 and Fig. 7 are comparisons of the mean square error of speech data [ai] and [ɔ:] using standardized model, LPC and RBF.

By calculating the mean square error, [ai] gets $3.2e-6$ using LPC, $1.1e-2$ using RBF and $2.1e-7$ using standardized model. The mean square error of [ɔ:] is 0.0785 by LPC, $3.2e-1$ by RBF, and 0.0173 by standardized models. The results show that the mean square error of speech signals generated by standardized model is less than by LPC model and RBF.

5.2.3 Signal to Noise Ratio Comparisons

Signal to noise ratio (SNR) is a time domain method for evaluating the distortion measure. It can effectively evaluate the two methods by analyzing the signal to noise ratio of the data generated by the two prediction methods. And in this paper, we use the short-term signal to noise ratio (Short-Time Signal-Noise Ratio, STSNR) which realized by block SNR, i.e., divide a speech into L blocks, each block is M frames:

$$SEGSNR = \frac{10}{L} * \sum_{i=0}^{L-1} \log \left\{ \frac{\sum_{n=0}^{M-1} s(i * (M-1) + n)^2}{\sum_{n=0}^{M-1} (s(i * (M-1) + n) - \hat{s}(i * (M-1) + n))^2} \right\} \quad (15)$$

$s(n)$ is the original speech signal, $\hat{s}(n)$ is the encoded signal. Table 5 is the signal to noise ratio of the forecast data using the LPC and standardized model (10) (which [ainə] intercepted from a word, and then separately predict [ai] [nə], and finally merge).

In Table 5, the standardized model (10) predict the signal with a higher noise ratio of speech signal than that the LPC predict, which is more conducive to the recovery of the original signal.

Table 5 Comparison of the samples' SNR

Prediction methods	The samples' signal to noise ratio comparisons				
	[ai]	[ɔ:]	[a:]	[əu]	[ainə]
LPC	0.69	3.08	0.14	1.22	0.34
Standardized model	19.09	18.5	23.07	25.04	11.65

5.2.4 Evaluation of Tone Quality

This paper presents the evaluation of the tone quality between the two prediction models' speech signal by the mean opinion score (Mean Opinion Score, referred to as the MOS score). MOS scores is based on the absolute level evaluation for the systematic evaluation of speech overall satisfaction or quality of speech communication. In this paper, there is five scoring criteria that the people participated the test would choose one of the scoring criteria as the MOS scores of the testing speech quality after they listened the testing speech.

In this paper, MOS method is used to evaluate the speech signal which is tested by these two methods, there are 10 people participated in the test and calculated the averaging of the result, the results are shown in Table 8.

The Table 8 shows that the MOS scores of speech signal are higher in standardized model prediction than in the LPC, so the speech signal predicted by standardized model (10) is better.

The Table 8 shows that the MOS scores of speech signal are higher in standardized model prediction than in the LPC, so the speech signal predicted by standardized model (10) is better.

Table6. The scoring criteria of MOS scores

score	Level of quality(MOS)	Level of distortion(DMOS)
5	excellent	Undetectable
4	good	A little detectable
3	fair	A little detectable and a little intolerable
2	poor	Detectable, repulsive but tolerable
1	bad	intolerable

Table 7. Evaluation of tone quality

phoneme	MOS scores with LPC	MOS scores with
---------	---------------------	-----------------

	prediction	standardized model prediction
[ai]	3.25	4.15
[ɔ:]	3.45	4.30
[a:]	3.02	4.14
[əu]	3.15	3.95
[ainə]	3.39	3.66

5.3 Speech Coding

The transmission rate is one of the indicators for evaluation of speech coding model, which refers to the number of bits need to transport per second of speech. Using the chaotic time series of the speech signal coding model (8) to encode the speech signal, we need to transfer P initial value and Q prediction parameters to the decoder. As the same value for the same speech P, Q is determined by the number of parameters for each frame of the prediction model. Set the number of speech samples as N_s , the number of prediction parameters is N_d , and data of each frame is 30ms, so the data transfer bit rate is:

$$I = (P * N_s + Q * N_d) / (30ms) \quad (16)$$

While using the standardized model to predict [ai], the number of the initial value is 20, the number of bits of prediction parameters is 10, the number of prediction parameters is $m * 2 = 20$ (m is the embedding dimension). The transfer bit rate is: 12kbit/s.

In the standardized model used in this paper, the transfer bit rate of the sample affected by the order number of the model and the number of initial values. The number of the initial value and order number of the model are determined by the embedding dimension and time delay of sample, i.e. the order number of model is the embedding dimension of samples and the number of initial values is the product of embedding dimension and time delay. In the practical application, we choose the order number of model and the number of initial values for the different samples to reduce the coding rate. In addition, coding for different characteristics of speech samples with different types of model is also able to further reduce the coding rate.

6 Conclusion

Speech signal is a chaotic system which can find order and regularity from the disorder and complexity. In this paper, we reconstructed the phase space of speech signal time series. The construction is based on the chaotic nature of speech signal. In order to find solutions for the model of such speech signal chaotic time series, we introduced the GP algorithm with hill-climbing algorithm. That also can improve the model structure optimization and parameter optimization capabilities. Then we use the improved algorithm on 5 groups of samples to calculate a set of optimal model structure. After further selection and standardization of the model, the speech signal chaotic time series forecasting model with certain generalization ability was obtained. We also evaluated the model by empirically re- searching; the results showed that the coding accuracy of chaotic time series nonlinear model obtained by the proposed method in this paper is better than LPC algorithm and RBF model.

As different speech signals have different chaos characteristics, using the chaos characteristics for classification modeling the speech signal is necessary. This study proposed an effective model structure for the sample selected, but the more important is that this paper proposed a method for solving the nonlinear model of chaotic time series. According to the present results, establishment

for effective coding models to classify diverse speech samples with various chaotic characteristics still need to be investigated in future.

Acknowledgements

This work reported in this paper was supported by the NSF of China (Grant no. 11172342, 11372167), and the Program of Shaanxi Science and Technology Innovation Team (Grant no. 2014KTC-18). And the Shaanxi Natural Science Foundation Project (2014JM8353). The authors thank the referees for their valuable suggestions and comments

References

- [1] O. Kocal, E. Yuruklu, I. Avcibas, Chaotic-type features for speech steganalysis, *IEEE Transactions on Information Forensics and Security*. 3(4) (2008) 651–661.
- [2] M. Dajer, J. Pereira, C. Maciel, Nonlinear dynamical analysis of normal voices, in: *Seventh IEEE International Symposium on Multimedia*, 2005, pp. 765–771.
- [3] L. Chen, X. Zhang, The nonlinear characteristics of the speech signal, *Journal of PLA University of Science and Technology*. 1(2) (2000) 11–17.
- [4] I. Kokkinos, P. Maragos, Nonlinear speech analysis using models for chaotic systems, *IEEE Transactions on Speech and Audio Processing*. 13(6) (2005) 1098–1109.
- [5] J. Sun, N. Zheng, X. Wang, Enhancement of Chinese speech based on nonlinear dynamics, *Signal Processing*. 87(10) (2007) 2431–2445.
- [6] M. Ogorzalek, Signal coding and compression based on discrete-time chaos: statistical approaches, in: *IEEE International Symposium on Circuits and Systems*, 2002, PP. 564–567.
- [7] M. L. Seltzer, D. Yu, Y. Wang, An investigation of deep neural networks for noise robust speech recognition, in: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp.7398–7402.
- [8] A. Miranian, M. Abdollahzade, Developing a local least-squares support vector machines-based neuro-fuzzy model for nonlinear and chaotic time series prediction, *IEEE Transactions on Neural Networks and Learning Systems*. 24(2) (2013) 207–218.
- [9] Y. Naniwa, T. Kondo, K. Kamiyama, H. Kamata, Study on the Artificial Synthesis of Human Voice using Radial Basis Function Networks, in: *Advanced Methods, Techniques, and Applications in Modeling and Simulation*. Springer Japan, 2012, pp. 291–300.
- [10] S. Scanzio, S. Cumani, R. Gemello, F. Mana, P. aface, Parallel implementation of artificial neural network training, in: *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2010, pp. 4902–4905.
- [11] H. Pulakka, P. Alku, Bandwidth extension of telephone speech using a neural network and a filter bank implementation for highband mel spectrum, *IEEE Transactions on Audio, Speech, and Language Processing*. 19(7) (2011) 2170–2183
- [12] J. Thyssen, H. Nielsen, S. Hansen, Non-linear short-term prediction in speech coding, in: *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1994, pp. 185–188

- [13] M. K. Luka, I. A. Frank, G. Onwodi, Neural network based Hausa language speech recognition, *International Journal of Advanced Research in Artificial Intelligence*. 1(2) (2012). 39–44.
- [14] Y. S. Lee, L. I. Tong, Forecasting time series using a methodology based on autoregressive integrated moving average and genetic programming, *Know.-Based Syst.* 24(1) (2011) 66–72.
- [15] Q. Aina, H. Zhong, G. Weihua, Nonlinear speech predictor using models for chaotic systems, *Computer Engineering and Applications*. 44(18) (2008) 141–143.
- [16] M. Conrads, P. Nordin, W. Banzhaf, Speech sound discrimination with genetic programming, *Genetic Programming*. Springer Berlin Heidelberg. (1998) 113-129.
- [17] H. Xie, M. Zhang, P. Andreae, Genetic programming for automatic stress detection in spoken English, In *Applications of Evolutionary Computing*. (2006) 460-471.
- [18] P. Andreae, H. Xie, M. Zhang, Genetic programming for detecting rhythmic stress in spoken english. *International Journal of Knowledge-based and Intelligent Engineering Systems*. 12(1) (2008) 15-28.
- [19] X. Wu, Z. Yang, Nonlinear speech coding model based on genetic programming, *Applied Soft Computing*. 13(7) (2013) 3314 – 3323.
- [20] J. R. Koza, J. P. Rice, *Genetic programming II: automatic discovery of reusable programs*, MIT Press, Cambridge, MA, USA, 1994.
- [21] K. Y. Chan, C. K. Kwong, T. C. Wong, Modelling customer satisfaction for product development using genetic programming, *Journal of Engineering Design*. 22(1) (2011) 55–68.
- [22] A. J. Parkes, E. Ozcan, M. R. Hyde, Matrix analysis of genetic programming mutation, in: *Proceedings of the 15th European conference on Genetic Programming, EuroGP'12, 2012*, pp.158–169.
- [23] N. Wagner, Z. Michalewicz, M. Khouja, R. McGregor, Time series forecasting for dynamic environments: The dyfor genetic program model, *IEEE Transactions on Evolutionary Computation*. 11(4) (2007) 433–452.
- [24] P. A. Estevez, N. Becerra-Yoma, N. Boric, J. A. Ramirez, Genetic programming-based voice activity detection, *Electronics Letters*. 41(20) (2005) 1141–1143.
- [25] H. D. I. Abarbanel, R. Brown, J. J. Sidorowich, L. S. Tsimring, The analysis of observed chaotic data in physical systems, *Reviews of Modern Physics*. 65(4) (1993) 1331–1392.
- [26] L. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, *Physica D: Nonlinear Phenomena*. 110(12) (1997) 43 – 50.
- [27] M. Qingfang, Z. Qiang, P. Jinfeng, New method for determining the minimum embedding dimension based on four-order cumulate, *System Engineering Theory and Practice*. (2005). 83–88.
- [28] H. S. Kim, R. Eykholt, J. D. Salas, Nonlinear dynamics, delay times, and embedding windows. *Physica D: Nonlinear Phenomena*. 127(1) (1999) 48 – 60.
- [29] M. T. Rosenstein, J. J. Collins, C. J. D. Luca, A practical method for calculating largest

- Lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena*. 65(1) (1993) 117–134.
- [30] X. F Xie, W. J. Zhang, Z. L. Yang, A dissipative particle swarm optimization, in: Congress on Evolutionary Computation (CEC), Hawaii, USA, 2002, pp. 1456-1461
- [31] W. Xiaojun, Y. Zhanzhong, Z. Ming, A uniform searching particle swarm optimization, *Acta Electronica Sinica*. 39(6) (2011) 1261–1266.

Graphical abstract