ATLANTIS PRESS

Advances in Engineering Research, volume 154

International Conference on Mechanical, Electrical, Electronic Engineering & Science (MEEES 2018)

Design of the Nonlinear Prediction Model for Chinese Speech Signal Based on RBF Neural Network

Xiaohong Gao

College of electrical engineering, Longdong University, Qingyang 745000, China.

Keywords: Nonlinear prediction model, Chinese speech signal, Radical Basis Function Neural Network.

Abstract. The nonlinear characteristic of Chinese speech signal is further studied, combined with radical basis function neural network, a nonlinear prediction model is designed. Firstly, delay time, embed dimension and maximum lyapunov exponent of Chinese speech phoneme are calculated by using C-C algorithm, false neatest neighbor algorithm and wolf algorithm, it is found out that Chinese speech signal has nonlinear characteristic. Secondly, combined with delay time and embed dimension, radical basis function neural network analysis method is applied successfully to design nonlinear prediction model. Lastly, compared with adaptive differential pulse code modulation linear prediction model and back propagation neural network nonlinear prediction model, prediction error of the nonlinear prediction model is significantly reduced, and the prediction performance gets much better.

1. Introduction

Recently, Analysis theories and processing methods of nonlinear dynamic for speech signal processing have been studied extensively [1, 2]. The researchers have applied chaos, fractal and other theories to research of speech signal nonlinear characteristic modeling [3-5], and then processing methods, such as wavelet transform and neural network and so on, have been applied to speech prediction, speech codec, speech recognition, etc [6-8]. Different neural networks have been widely investigated and applied in speech signal nonlinear modeling, thus different nonlinear prediction model have been constructed, and have gained certain achievements [9-12].

Lin et al. [9] applied recurrent neural network (RNN) to the prediction of speech. The result indicates that the predictor based on RNN not only has better prediction performance for long-range correlation, but also has better robustness for embedding dimension. Al-Jumeily et al. [10] hold that self-organized neural network inspired by the immune algorithm (SMIA) is presented, which applied to women and man voices counting from one to ten in Arabic. The simulation results indicated that SMIA is superior in prediction performance to the multilayer perceptions neural network (MLPN). Using RBF neural network to speech signal nonlinear modeling, Lin and Liu studied various training methods of modeling for speech signal-based RBF neural network [11]. In Ref. [12], nonlinear predictor based on RBF neural network is constructed, compared with LPC linear predictor, many aspects have been significantly improved.

However, an interesting question arises: whether Chinese speech signal has nonlinear? If nonlinear exists in Chinese speech signal, how to construct prediction model combined with Chinese speech signal nonlinear characteristic. This will be an interesting topic.

In this paper, analysis of the Chinese speech signal is based on phoneme, because phoneme is the most basic, the smallest and indecomposable unit from the sound quality and natural attribute, and which is also the foundation for analyzing speech signal. Firstly, using C-C algorithm to calculate delay time and embed dimension of Chinese speech signal at the same time, and FNN algorithm is adopted to calculate embed dimension, the values of these two algorithms is researched and analysis at once, the algorithms used for calculating delay time and embed dimension are eventually ascertained. Then, maximum lyapunov exponent (MLE) is calculated by wolf algorithm. The nonlinear characteristic of Chinese speech signal is proved by these nonlinear characteristic parameters. Finally, nonlinear prediction model is constructed based on RBF neural network

combined with characteristic parameters of Chinese speech phonemes. To be special, the number of the three layers neurons for RBF neural network are determined by the mean of delay time and the mean of embed dimension for 38 Chinese speech phonemes. The ultimate aim is to design codec system for Chinese speech which can better restore the original signal.

2. Nonlinear Characteristics of Chinese Speech Signal

2.1 Phase Space Reconstruction of Chinese Speech Signal.

Decision, analysis and prediction for arbitrary chaotic time series are always carried out in reconstructed phase spaces, so phase space reconstruction is the key to analysis and research nonlinear time series by using dynamical methods. As for as speech signal concerned, phase space reconstruction is not only the most important method for extracting dynamical information from speech signal time series, but also the first step for analyzing speech signal dynamic system. One-dimension time series of speech signal can be written as

$$X(t) = \{x1, x2, x3, \cdots, all\} \quad (1)$$

According to Taken embedding theorem [13], delay coordinate reconstitution theory is used to reconstruction the phase space:

$$Xi = [xi, xi + \tau, xi + 2*\tau, \cdots, xi + (m-1)*\tau], (1 \le i \le N) \quad (2)$$

Where N is the total number of embedded points in m-dimensional space ($N=L-(m-1)*\tau$), L is the size of the data set, τ denotes time delay, m denotes the embedding dimension.

Accurate calculation methods for reconstruction parameters (τ and m) from times series determine strongly the quality of phase space reconstruction. At present, there are two important views: one is that the two parameters are uncorrelated, the methods for calculating the time delay mainly include the autocorrelation function [14], mutual information method [15], etc. And methods for calculating the embedding dimension mainly include false nearest neighbors method (FNN) [16], Cao method [14], etc. The other is that the two parameters are correlated, a typical method is C-C method [17, 18]. However, each one has its own merits in these methods, for instance, although autocorrelation method is suitable for small data sets and easy to calculate, nonlinear factors of the system are not considered, and mutual information method is useful for large data sets, and the nonlinear factors of the system are considered, but the computation of the method is very large and the calculating speed is lower, nevertheless, C-C method can maintain nonlinear characteristics of the system, and it is not only easier to implement but also has less demanding computationally. In addition, the most important advantage of the method is that it can calculate τ and m simultaneously.

In this paper, the delay time and the embedding dimension of Chinese speech phoneme at 8 kHz sampling rate and 16 Quantification bits are calculated by using C-C method at the same time, Table 1 shows the partial results.

2.2 Lyapunov Exponent of Chinese Speech Signal.

Lyapunov exponent (LE) is an important index for describing the system dynamic characteristics, which is also key parameter to determine whether there exits dynamic chaos in the system and measure the average exponential rate of divergence or convergence of adjacent orbits in phase space. When LE>0, the motion trajectory of phase space is very unstable and long-term unpredictability, long time dynamic behavior is sensitive dependence on initial conditions, and this is the chaos state. When LE<0, the trajectory of phase space is contracted and eventually merged into point, this is a stable fixed point and periodic motion. When LE=0, the initial error is neither magnified nor reduced, which corresponds to the stable boundary, it indicates that the system time delay reconstruction graph is close to limit cycle.

Time delay reconstruction graphs of the partial Chinese speech phonemes is depicted in Fig.1, τ is the delay time, x[n] is original speech signal, x [n+τ] is the signal after time delay. For a high dimensional system, at least one positive lyapunov index calculated shows that the system does chaotic motion. Thus, the maximum lyapunov exponent of Chinese speech signal calculated is greater than 0, the signal has chaotic.

At present, besides definition calculation, many numerical methods are researched for calculating lyapunov exponent which can approximately be assorted to two main classes: Wolf algorithm [19] and Jocobian algorithm [20]. Wolf algorithm needs a lot of data and fractal dimension can not be too high, which is suitable for the time series without noise and the evolution of space vectors with highly nonlinear. Jocobian algorithm is suitable for the time series with high noise and the evolution of space vectors close to linear. Thus, to better verify the signal nonlinear characteristic, Wolf algorithm is used for calculating maximum MLE of Chinese speech, and Table 1 shows that MLE of the partial Chinese speech phonemes.

Table 1. Delay time τ, embedding dimension m and the MLE of the partial phonemes

Phoneme	τ (C-C method)	m (C-C method)	m (FNN method)	MLE	D (GP method)
b	7	3	3	0.0683	1.12±0.05
g	4	3	4	0.0123	1.15±0.05
h	4	3	4	1.2610	no convergence
c	7	1	4	2.5403	no convergence
m	4	3	4	0.0339	no convergence
n	6	4	4	0.0015	1.28±0.02
l	4	4	3	0.0965	1.55±0.01
a	4	4	5	0.1707	2.55±0.03
o	3	4	3	0.0116	1.67±0.02
u	6	3	3	0.0055	1.05±0.03
an	4	3	4	0.3262	no convergence
in	3	4	3	0.0024	1.10±0.01
ing	5	3	3	0.0119	1.06±0.02

Annotation: voiceless: b, g; voiceless fricative: h; voiceless affricate: c; voiced nasal: m, n; voiced lateral: l; simple finals: a, o, u; nasal finals: anointing.

Table 1 and Fig.1 are observed and some conclusions have been made.

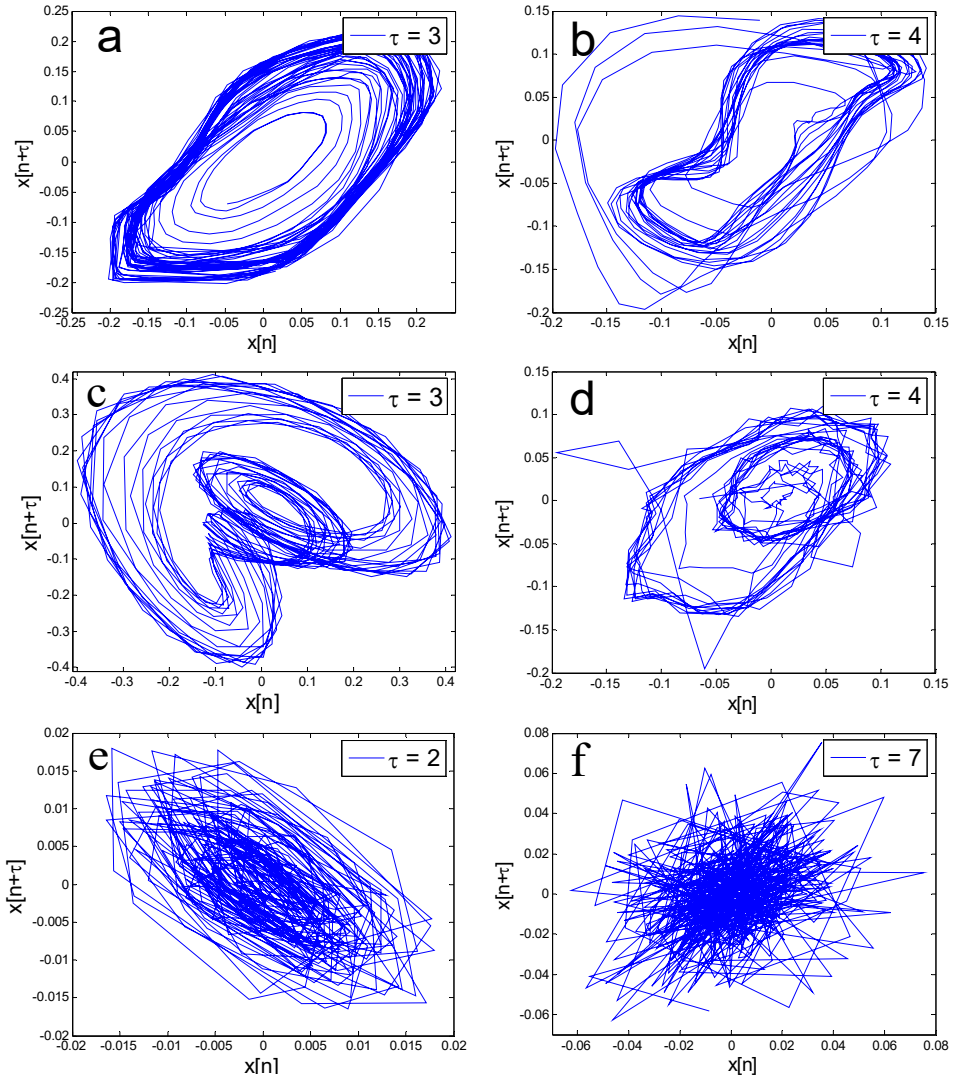
(1) MLE of the partial Chinese speech phonemes is greater than 0, which indicates that the signal has chaotic property.

(2) The shape for the time delay reconstruction graph of every Chinese phoneme is related to its delay time and embedding dimension, thus, different Chinese phoneme has different time delay reconstruction graph.

(3) Voiced has periodic, time delay reconstruction graphs have manifested as closed torus, on the contrary, unvoiced has no periodic, which is quite different from voiced, time delay reconstruction graphs have manifested as irregular curves.

(4) The MLE values of phoneme g m n o u in and ing are positive, and they are close to 0, It is shown that their time delay reconstruction graph is closer to limit cycle. Looking closely at Fig.1, we find that time delay reconstruction graphs of phoneme m and in are much approximate to limit cycle, and with the increase of MLE value, limit cycle will be more complex, such as phoneme o.

(5) Compared with consonants, MLE values of vowels are smaller, this also shows that chaotic degree of consonants is relatively high, in addition, by observing tine delay reconstruction graphs of phoneme l h and c in Fig.1, it can be found that the lager MLE value, the higher the chaotic degree is, the more complex the time delay reconstruction graph is.



(a) time delay reconstruction graph of phoneme in (b) time delay reconstruction graph of phoneme m (c) time delay reconstruction graph of phoneme o (d) time delay reconstruction graph of phoneme l (e) time delay reconstruction graph of phoneme h (f) time delay reconstruction graph of phoneme c

Fig 1. Time delay reconstruction graph of the partial Chinese speech phonemes.

3. Design of Chinese Speech Signal Prediction Model Based on RBF Neural Network

3.1 RBF Neural Network.

Neural network is the most widely used in nonlinear prediction of time series because of its better nonlinear modeling ability. Many methods are used for analyzing neural network, such as RBF neural network, BP neural network, multilayer perceptron neural network, etc. RBF neural network is a good tool for speech signal nonlinear processing, and it is not only a universal function approximator, but also it has better local performance and stability.

The basic RBF neural network is shown in Fig.2, which consists of three parts: input layer, hidden layer and output layer. Every layer plays different role: the input layer connects the input signal directly; hidden layer realizes nonlinear mapping from input layer to hidden layer, and the higher its dimension, the higher the precision of the function approximation; the output layer realizes the linear combination of the hidden layer output signal.

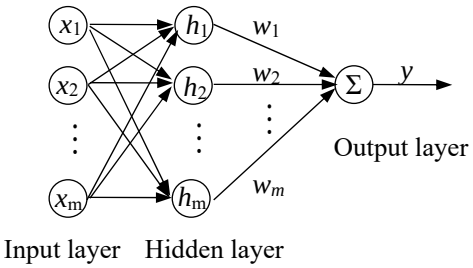


Fig 2. RBF neural network model.

RBF neural network can be mapped to f: R m →R.

$$f(x) = \sum_{j=0}^{m-1} w_j \exp(-\frac{\|X - c_j\|^2}{\delta_j^2}) \tag{3}$$

X is input vector, m is the number of node in hidden layer (the number of neurons), and its dimension is m; and the three important parameters of RBF neural network: cj is the basic function center of hidden layer node, and its dimension is m, δj is the basic function width, wj is the output weight; $\|\cdot\|$ is European norm. The closer the input vector is from the center, the larger the function value; the wider the width, and the farther the input vector is from the center, the attenuation amplitude of the function is greater. Gaussian function can directly reflect the characteristic of local response, and which has the advantages, such as radial symmetry, good smoothness and arbitrary order derivable, etc. Therefore, Gaussian function is usually chosen as the basic function of the hidden layer.

K-means clustering algorithm is used to training center cj and width δj, where mj is the number of training samples in the cluster j, Xj is the input vector in the cluster j.

$$c_j = \frac{1}{m_j} \sum X_j \tag{4}$$

$$\delta_j^2 = \frac{1}{m_j} \sum \|X_j - c_j\|^2 \tag{5}$$

The learning process of RBF network is divided into two stages: (1) Training stage: firstly, providing input samples and output samples to neural network, then numerical calculation and parameters optimization of input samples are done, the parameters are constantly adjusted until the given input can produce the desired output, through this process, the center cj and width of hidden layer δj, and the weights of output layer wj are determined. (2) Prediction stage: the unknown samples are predicted according to the trained network.

3.2 Design of Chinese Speech Signal Prediction Model.

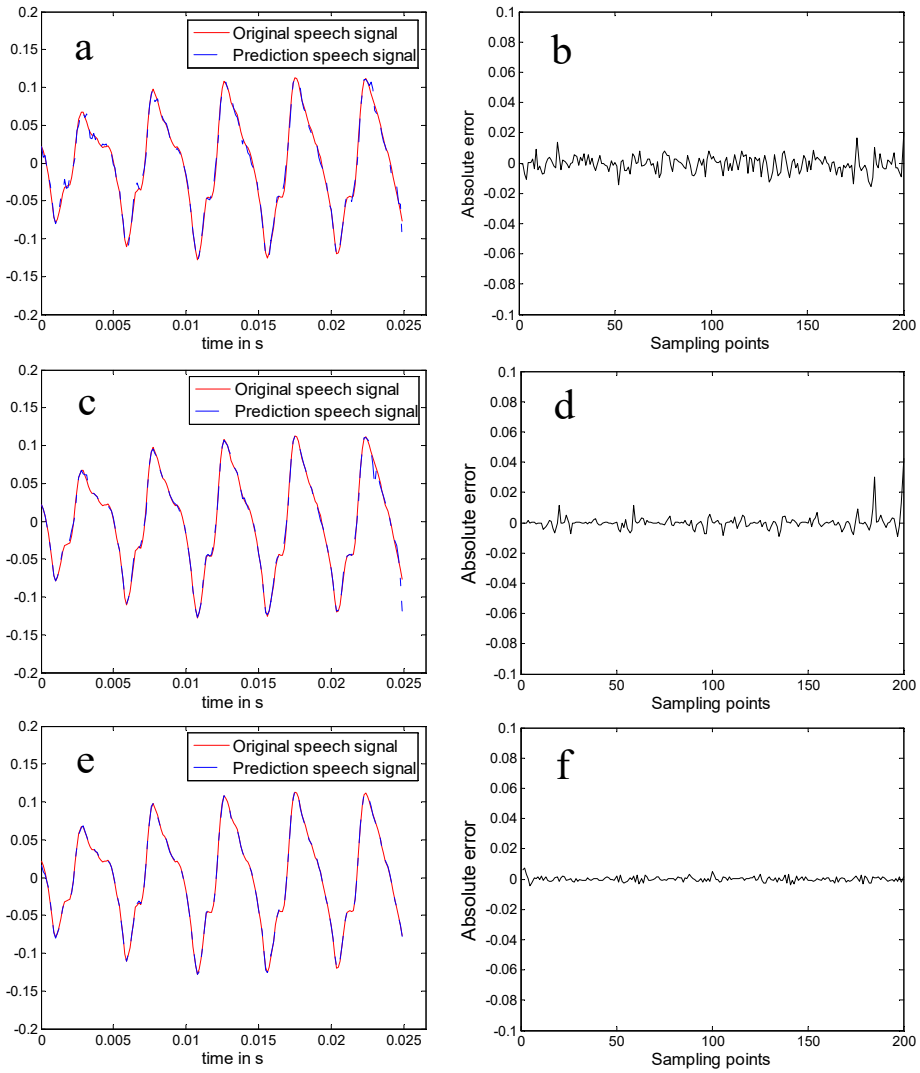
Delay time and embedding dimension of every phoneme are different, so the mean value of delay time and the mean value of embedding dimension for 38 Chinese speech phonemes are calculated as the delay time and embedding dimension of all Chinese speech signal. The design idea of the prediction model is that delay time as the number of input layer neuron and output layer neuron and embedding dimension as the number of hidden layer neuron. Then the specific reasons are as follows:

(1) In the linear prediction of speech signal, the prediction accuracy reaches a saturation value with the increase of prediction order. The phenomenon can be interpreted that the larger the prediction order and embedding dimension, and the lower the proportion of the false nearest neighbor points in phase space, but the proportion of the false nearest neighbors will become a fixed value when embedding dimension increases to be a certain value [12]. Only the hidden layer has nonlinear characteristics in RBF network, so we chose embedding dimension as the number of the hidden layer neurons.

(2) Time delay τ plays a key role in phase space reconstruction, which makes the adjacent data in phase space reconstruction uncorrelated as much as possible, so that phase point in the embedding space contains information of the original attractors as large as possible, therefore, τ is used as the number of the input layer neuron and the out-layer neuron.

3.3 Comparison and Analysis of Simulation Results for Models.

The length of Chinese speech phoneme is usually about 2~4 frames, in order to better test prediction performance of prediction model, one frame speech signal (25ms, 200 sampling points) is selected for predicting using constructed RBF neural network nonlinear prediction model, BP neural network nonlinear prediction model and adaptive differential pulse code modulation (ADPCM) linear model. Fig.3 is absolute error comparison of Chinese speech signal linear prediction and nonlinear prediction.



(a) Prediction Results based on the ADPCM linear prediction model; (b) Absolute error based on the ADPCM linear prediction model; (c) Prediction Results based on the BP nonlinear prediction model; (d) Absolute error based on the BP nonlinear prediction model; (e) Prediction Results based on the RBF nonlinear prediction model; (f) Absolute error based on the RBF nonlinear prediction model.

Fig 3. Absolute error comparison between the original Chinese speech signal and prediction Chinese speech signal.

Three kinds of evaluation indexes are introduced in this paper in order to further test the prediction performance of three prediction models:

Relate mean square root error (RMSE).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - x_i'}{x_i}\right)^2} \tag{6}$$

Xi is the actual value of sample point i, xi' is the prediction value, n is the total number of sample points (where n is 200).

Signal to noise ratio (SNR)

$$SNR = 10 * \log(\frac{P_s}{P_n}) \tag{7}$$

Ps is the effective power of the original speech signal, and Pn is the effective power of the predictive signal.

MOS score

In this paper, MOS method is used evaluate the speech signal which is tested by the method, there are twenty people who participated in the test and calculated the mean value of the results. The results are shown in Table 2.

Table 2. Comparison of prediction performance for the three prediction models.

Evaluation index	*RMSE*	*SNR*	MOS
ADPCM linear prediction model	0.2949	21.3638	3.12
BP nonlinear prediction model	0.2061	22.1633	3.65
RBF nonlinear prediction model RBF	0.1526	26.4085	3.97

One finds from Table 2 and Fig.3, RMSE of RBF nonlinear prediction model are relatively small, and SNR and MOS score are very high, which indicate that the prediction performance of the prediction model is better than two other models, and the prediction value is very close to the actual value.

4. Conclusion

The nonlinear characteristics of Chinese speech signal have deeply researched in this paper, the methods of calculating the nonlinear characteristics parameters are determined, including delay time, embedding dimension and maximum lyapunov exponent. Then the number of the three layers neurons for RBF neural network is determined by the mean value of calculating delay time and the mean value of calculating embedding dimension for 38 Chinese speech phonemes, 5 neurons in input layer and output layer, 4 neurons in the hidden layer, thereby a nonlinear prediction model based on RBF neural network is constructed successfully. The simulation results show that compared with the BP nonlinear prediction model and ADPCM linear prediction model, nonlinear prediction model based on RBF network obviously reduces prediction error, increases signal to noise ratio, and improves prediction performance. Thus, the prediction accuracy of Chinese speech signal nonlinear model obtained by the proposed method in this paper is better than ADPCM linear model and BP nonlinear model.

References

[1]. A. Kumar, S.K. Mullick. Nonlinear dynamical analysis of speech. J.Acoust.Soc.Am. Vol. 100 (1996) No. 1, p. 615-629.

[2]. S.S. Narayanan, A.A. Alwen. A nonlinear dynamic system analysis of fricative consonants. J. Acoust. Soc. Am. Vol. 97 (1995) No. 4, p. 2511-2524.

[3]. N. Chaitra, D. Murali Mohan, D. Narayana Dutt. Nonlinear Dynamical Analysis of Speech Signals. Proceedings of international conference on VLSI, Communication, Advancsd Devices, Signals & Systems and Networking. 2013, p. 258, 343-351.

[4]. S.Q. Hu, Y. Zhang, Y.M. Huang. Nonlinear dynamic characteristic analysis of speech for Chinese. Acustica. Vol. 25 (2000) No. 4, p. 329-334.

[5]. O.H. Kocal, E. Yuruklu, I. Avcibas, Chaotic-type features for speech steganalysis. IEEE Trans. Inf. Forensics Security. Vol. 3 (2008) No. 4, p. 651–661.

[6]. X. Wu, Z. Yang, Nonlinear speech coding model based on genetic programming, Appl. Soft Comput. Vol. 13 (2013) No. 7, p. 3314–3323.

[7]. Y. Wang, D. Yu, M.L. Seltzer. An investigation of deep neural networks for noise robust speech recognition, in: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. 2013, p. 7398–7402.

[8]. M.K. Luka, I.A. Frank, G. Onwodi, Neural network-based Hausa language speech recognition, Int. J. Adv. Res. Artif. Intell. Vol. 1 (2012) No. 2, p. 39–44.

[9]. T. Lin, B.G. Horne, P. Tino, et al. Learning long-term dependencies in NARX recurrent neural networks. IEEE Transactions on neural networks. Vol. 7 (1996) No. 6, p. 1329-1338.

[10]. D. Al-Jumeily, A.J. Hussain, P. Fergus, N. Radi. Self-organized neural network inspired by the immune algorithm for the prediction of speech signals. Intelligent computing theories and methodologies. Vol. 9226(2015), p. 654-664.

[11]. J. Y. Lin, Y. Liu. Training methods and the performances of RBF neural networks for nonlinear modeling of speech signals. Signal processing. Vol. 17 (2001) No. 4, p. 322-328.

[12]. A.N. Qin, Z. Huang, W.H. Gui. Nonlinear speech predictor using models for chaotic systems, Comput. Eng. Appl. Vol. 44 (2008) No. 18, p.141–143.

[13]. F. Taken. Detecting strange attractors in turbulences. Dynamical System and Turbulence, Warwick, Springer Verilog, Berlin. 1980, p. 898,366-381.

[14]. Loco. Practical method for determining the minimum embedding dimension of a scalar time series. Physical D: Nonlinear Phenomena. Vol. 110 (1997) No. 1-2, p. 43-50.

[15]. X.F. Xin, W.J. Zhang, A.L. Yang. A dissipative particle swarm optimization, in: Congress on Evolutionary Computation (CEC), Hawaii, and USA. 2002, pp.1456-1461.

[16]. M.B. Kennel, R. Brown, H.D.I. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phy Rev A. Vol. 45(1992) No. 6, p. 3403-3411.

[17]. D. Kugiumtzis. State space reconstruction parameters in the analysis of chaotic time series-the role of the time window length, Physica D: Nonlinear Phenomena. Vol. 95 (1996) No. 1, p. 13-28.

[18]. H.S. Kim, R. Eykholt, J.D. Salas. Nonlinear dynamics, delay times, and embedding windows. Physica D: Nonlinear Phenomena. Vol. 127 (1999) No. 1-2, p. 48-60.

[19]. A. Wolf, J.B. Swift, H.L. Swinney, et al. Determining Lyapunov exponents from a time series. Physica D. Vol. 16 (1985) No. 3, p. 285-317.

[20]. G. Barana, I. Tsuda. A new method for computing Lyapunov exponents. Phys. Lett,A. Vol. 175 (1993) No. 6, p. 421-427.