

# UC Berkeley

## UC Berkeley Previously Published Works

**Title**

Application of novel chaotic neural network on Mandarin digital speech recognition

**Permalink**

<https://escholarship.org/uc/item/9k25p3sq>

**Journal**

Fuzzy Systems Conference, 1

**ISSN**

1098-7584

**Authors**

Li, Guang  
Zhang, Jin  
Freeman, Walter J, III

**Publication Date**

2007

**DOI**

10.1109/FUZZY.2007.4295337

Peer reviewed

# Mandarin Digital Speech Recognition Based on a Chaotic Neural Network and Fuzzy C-means Clustering

Guang Li, Jin Zhang, and Walter J. Freeman

**Abstract**—Modeling olfactory neural systems, the KIII model proposed by Freeman exhibits chaotic dynamic characteristics and has potential for pattern recognition. Fuzzy c-means clustering can classify an object to several classes at the same time but with different degrees based on fuzzy sets theory. Based on the KIII model, mandarin digital speech is recognized utilizing the features extracted by the fuzzy c-means clustering. Experimental results show that the KIII model can perform digital speech recognition efficiently and the fuzzy c-means clustering has better performance than the hard k-means clustering.

## I. INTRODUCTION

Digital speech recognition can be widely used to many information services, for example, inquiry system of telephone number, identification of identity card number and so on. Mandarin digit pronunciations are all monosyllables and include some ambiguous and easily confusable syllables, such as “1” and “7”, making it difficult to be recognized correctly. Furthermore, if there is a digital string, the task will be more difficult.

As a typical pattern recognition issue, many recognition methods have been proposed, such as Gaussian mixture model [2], hidden Markov model (HMM) [3] and artificial neural networks (ANN) [4, 5], however many problems are still remaining unresolved and the accuracy of recognition needs to be improved [1]. Conventional ANNs simply simulate some basic properties of neural systems, such as parallel distributed structure, non-linearity and plasticity. According to the architecture and the dynamic characteristics of olfactory systems, a chaotic neural network entitled KIII model has been constructed [6]. The KIII model can simulate some advanced functions of brains, such as EEG waveform observed in biological experiments, and exhibits strong nonlinear characteristics. It can be used as a robust classifier with high rates of training and very rapid convergence to category assignments and has been studied to perform some pattern recognition tasks, such as face recognition [7], text classification [8] and so on.

This work is supported in part by the National Creative Research Groups Science Foundation of China (NCRGSFC: 60421002) and the National Basic Research Program of China (973 Program: 2004CB720302).

G. Li is with the National Laboratory of Industrial Control Technology, Institute of Advanced Process Control, Zhejiang University, Hangzhou, 310027, China (Tel/Fax: +86-571-87952268x8228; e-mail: guangli@zju.edu.cn).

J. Zhang is with the Department of Biomedical Engineering, Zhejiang University, Hangzhou, 310027, China and Software School of Hunan University, Changsha, 410082, China (e-mail: jinzhang@hnu.cn).

W. J. Freeman is with University of California at Berkeley, LSA 142, Berkeley, CA, 94720-3206, USA (e-mail: drwjfiii@berkeley.edu).

In order to combine with an ANN to construct a speech recognition system, an effective feature extraction from digital speech is essential. And the dimension of the feature vector should be equal. Though MFCC [9] can extract features effectively, the dimension of the feature vectors varies along with the lasting time of different digital speech. So, the k-means clustering algorithm has been used to normalize extracted features to make the dimension equal [10]. As an unsupervised clustering technique, the fuzzy c-means (FCM) clustering algorithm [11, 12] is more suitable to normalize the extracted features of ambiguous Mandarin digital speech in comparison with the k-means clustering algorithm. The main idea is that scattered groups of data are divided into several groups at the same time with different degrees according to fuzzy set theory.

In this paper, based on the features extracted by the FCM clustering algorithm, the novel chaotic neural network, the KIII model, is used as a classifier for Mandarin digital speech recognition. The experimental results show that the KIII model performs well in Mandarin digital speech recognition. Compared with the hard k-means (HKM) clustering algorithm, the FCM clustering algorithm is more suitable to extract features from speech signals to be recognized by the KIII model.

## II. KIII MODEL

KIII model proposed by Freeman [6, 13] mimics the olfactory neural network from the structure to function. The topological diagram of the KIII model, in which K0, KI and KII are included, is shown in Fig. 1.

In KIII model, each node represents a neural population or cell ensemble. The dynamical behavior of each ensemble of the olfactory system is governed by Equation (1):

$$\begin{aligned} \frac{1}{a \cdot b} [x_i''(t) + (a + b)x_i'(t) + a \cdot b \cdot x_i(t)] \\ = \sum_{j \neq i}^N [W_{ij} \cdot Q(x_j(t), q_j)] + I_i(t) \end{aligned} \quad (1)$$

$$Q(x_i(t), q) = \begin{cases} q(1 - e^{-(e^{x_i(t)} - 1)/q}) & x(t) > x_0 \\ -1 & x(t) < x_0 \end{cases} \quad (2)$$

$$x_0 = \ln(1 - q \ln(1 + 1/q))$$

where  $i=1 \dots N$  ( $N$  is the number of channels).  $x_i(t)$  and  $x_j(t)$  represents the state variable of  $i$ th and  $j$ th neural population respectively, while  $W_{ij}$  indicates the connection strength between them.  $I_i(t)$  is an input function, which stands for external stimulus. The parameter  $a = 0.220 \text{ msec}^{-1}$ ,  $b = 0.720$

Corresponding to biological olfactory systems, R represents the olfactory receptor, which is sensitive to the odor molecules, and provides the input to the KIII network in the topology of the KIII network. The periglomerular (PG) layer and the olfactory bulb (OB) layer are distributed. The anterior nucleus (AON) and the prepyriform cortex (PC) are only composed of single KII network. KIII model describes the whole olfactory neural system. It includes populations of neurons, local synaptic connection, and long forward and distributed time-delayed feedback loops.

and KII's by feedforward and time-delayed feedback loops, which are either excitatory or inhibitory, forms a five-layer KIII network to model the entire olfactory system.

In the KIII model, independent rectified Gaussian noise was introduced to every olfactory receptor to model the peripheral excitatory noises and single channel of Gaussian noise with excitatory bias was introduced to model the central biological noise sources. The additive noise eliminated numerical instability of the KIII model, and made the system trajectory stable and reliable under statistical measures, which meant that under perturbation of the initial conditions or parameters, the system trajectories were robustly stable [14]. Because of this stochastic chaos, the KIII network not only simulated the chaotic EEG waveforms, but also acquired the capability for pattern recognition, which simulated an aspect of the biological intelligence, as demonstrated by previous applications of the KIII network to recognition of one-dimensional sequences, industrial data and spatiotemporal EEG patterns [15], which deterministic chaos can't do, owing to its infinite sensitivity to initial conditions.

### III. FEATURE EXTRACTION BASED ON FUZZY C-MEANS CLUSTERING

Fuzzy  $c$ -means (FCM) clustering [17] is one of unsupervised clustering techniques and is based on the minimization of an objective function called  $c$ -means functional. The FCM employs fuzzy partitioning such that a data point can belong to all groups with different membership grades between 0 and 1. FCM is an iterative algorithm to find cluster centers (centroids) that minimize a dissimilarity function, which is different from HKM that employs hard partitioning.

$$\sum_{i=1}^c u_{ij} = 1, \forall j = 1, \dots, n \quad (3)$$

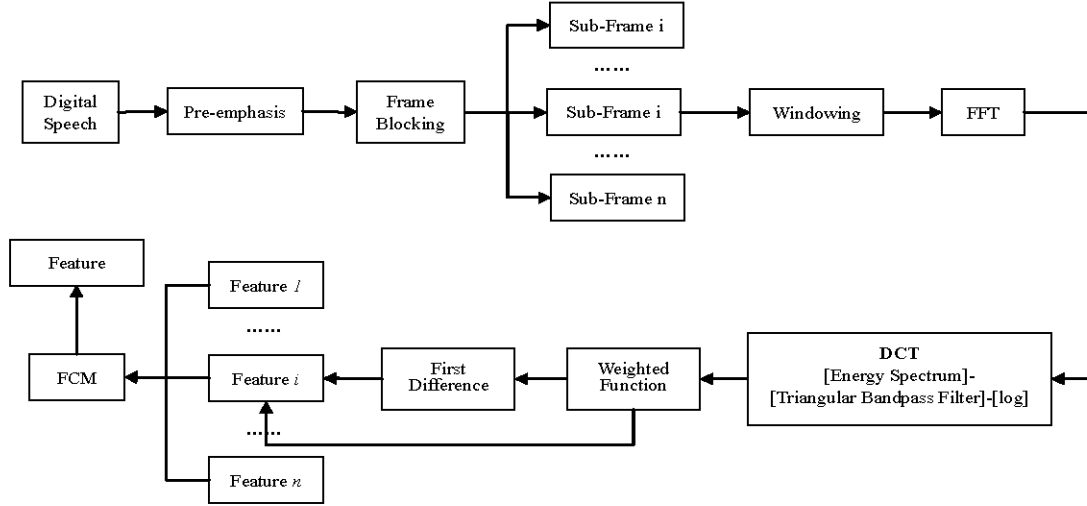


Fig.2. The whole process of feature extraction

where  $u_{ij}$  is between 0 and 1 and denotes membership degree of  $j_{th}$  data point belonging to  $i_{th}$  centroids ( $c_i$ ) when there are  $n$  data points and  $c$  centroids. The dissimilarity function used in FCM is given by Equation (4):

$$J(U, c_1, c_2, \dots, c_c) = \sum_{i=1}^c J_i = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m d_{ij}^2 \quad (4)$$

where  $c_i$  denotes the centroid of cluster  $I$ ;  $d_{ij}$  is the Euclidian distance between  $i_{th}$  centroid ( $c_i$ ) and  $j_{th}$  data point;  $m \in [1, \infty]$  is a weighting coefficient.

There are two conditions to reach a minimum of dissimilarity function described as in Equation (5) and (6).

$$c_i = \frac{\sum_{j=1}^n u_{ij}^m x_j}{\sum_{j=1}^n u_{ij}^m} \quad (5)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^c \left( \frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad (6)$$

This algorithm determines the following steps [18]:

**Step 1.** Randomly initialize the membership matrix ( $U$ ) that has constraints in Equation (3).

**Step 2.** Calculate centroids ( $c_i$ ) by using Equation (5).

**Step 3.** Compute dissimilarity between centroids and data points using Equation (4). Stop if its improvement over previous iteration is below a threshold.

**Step 4.** Compute a new  $U$  using Eq.6. Go to Step 2.

By iteratively updating the cluster centers and the membership grades for each data point, FCM iteratively moves the cluster centers to the "right" location. It can not be sure that FCM can converge to an optimal solution because  $U$  is randomly initialized and used to calculate cluster centers (centroids). So, the performance of FCM depends on initial centroids to some extent.

### B. Feature Extraction

The whole process of feature extraction is based on melscale frequency cepstral coefficients (MFCC), as shown

in Fig.2. The whole process of feature extraction includes several steps as follows [10]:

- Step 1: Pre-emphasis;
- Step 2: Frame Blocking;
- Step 3: Windowing;
- Step 4: Fast Fourier Transform (FFT);
- Step 5: Triangular Band-pass Filter;
- Step 6: Discrete Cosine Transform (DCT);
- Step 7: Delta Cepstrum;
- Step 8: Fuzzy C-means Clustering.

After Step 7, the feature of any frame of digital speech can be extracted, so the whole feature of one digital speech can be expressed by one feature matrix. Each row of matrix denotes the feature of one frame. The row number of the feature matrix of different digital speech isn't equal, but the column number is equal. So, after FCM, any feature matrix of digital speech can be transformed a new feature matrix, whose dimension is equal to each other. Combining each row of feature matrix can form the feature vector, which is the input of KIII model.

## IV. MANDARIN DIGITAL SPEECH RECOGNITION BASED ON KIII MODEL AND FCM

### A. Recognition Process

In the experiments, the KIII model is used as classifier to recognize Mandarin digital speech according to the extracted features. The parameters used in this study are optimized to fulfill the classification task. The stimulus-maintained period is from 100ms~300ms, while the initialization period is 0~100ms for the KIII to reach the basal non-convergent "chaotic" state. The simulation time is 400ms for each learning or classification process.

Similar to other ANN, the KIII model must be trained to classify the desired patterns. During training stage, each digital speech is learned only 10 times, much less than conventional ANNs, such as BP neural networks. After training, each stimulus converges to a stable limit cycle attractor, which denotes one spoken digit.

TABLE I: COMPARISON WITH BP NEURAL NETWORKS

|                | BP Network     | HKM+KIII     | FCM+KIII     |
|----------------|----------------|--------------|--------------|
| <b>Average</b> | <b>0.88333</b> | <b>0.900</b> | <b>0.984</b> |

TABLE II: THE RECOGNITION RESULTS WITH FCM

|                | 24*2          | 24*3          | 24*4          | 24*5          | 24*6          | 24*7          | 24*8          | 24*9          | 24*10         |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1              | 0.8222        | 0.5556        | 0.9333        | 1.0000        | 0.9333        | 0.9778        | 0.9889        | 1.0000        | 0.9556        |
| 2              | 0.8111        | 0.5778        | 0.6333        | 0.8778        | 1.0000        | 1.0000        | 0.8778        | 0.9667        | 0.8666        |
| 3              | 0.6333        | 0.8222        | 0.8111        | 1.0000        | 0.9667        | 0.9667        | 1.0000        | 1.0000        | 0.9889        |
| 4              | 0.7000        | 0.7778        | 0.8445        | 0.9889        | 0.8222        | 0.9667        | 0.8000        | 0.9222        | 0.9667        |
| 5              | 1.0000        | 0.9444        | 0.9445        | 0.6556        | 0.9333        | 0.9667        | 1.0000        | 1.0000        | 1.0000        |
| 6              | 0.9889        | 0.9000        | 0.5222        | 0.9333        | 0.8667        | 0.9667        | 0.9667        | 0.9444        | 0.9445        |
| 7              | 0.9889        | 1.0000        | 1.0000        | 0.8778        | 0.9445        | 1.0000        | 1.0000        | 1.0000        | 1.0000        |
| 8              | 1.0000        | 1.0000        | 1.0000        | 1.0000        | 1.0000        | 1.0000        | 1.0000        | 1.0000        | 1.0000        |
| 9              | 0.8778        | 0.9555        | 1.0000        | 1.0000        | 0.9333        | 1.0000        | 0.9889        | 1.0000        | 1.0000        |
| 0              | 1.0000        | 0.8222        | 0.9667        | 1.0000        | 0.9333        | 1.0000        | 0.9778        | 0.9889        | 0.9889        |
| <b>Average</b> | <b>0.8822</b> | <b>0.8356</b> | <b>0.8656</b> | <b>0.9333</b> | <b>0.9333</b> | <b>0.9844</b> | <b>0.9600</b> | <b>0.9822</b> | <b>0.9711</b> |

TABLE III: THE RECOGNITION RESULTS WITH HKM

|                | 24*2          | 24*3          | 24*4          | 24*5          | 24*6          | 24*7          | 24*8          | 24*9          | 24*10         |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 1              | 0.4500        | 0.7333        | 0.8667        | 0.9667        | 0.9333        | 0.9334        | 0.9667        | 0.9333        | 0.9667        |
| 2              | 0.5333        | 0.9000        | 0.9000        | 0.7167        | 0.7667        | 0.7667        | 0.8500        | 0.9000        | 0.8333        |
| 3              | 0.9000        | 0.9000        | 0.7333        | 0.8000        | 0.8333        | 0.8000        | 0.9000        | 0.8500        | 0.9667        |
| 4              | 0.8166        | 0.6667        | 0.8833        | 0.8334        | 0.8000        | 0.8667        | 0.8333        | 0.9667        | 0.9334        |
| 5              | 0.7333        | 0.9166        | 0.8000        | 0.8500        | 0.9667        | 0.9333        | 0.8667        | 1.0000        | 0.9333        |
| 6              | 0.8667        | 0.5167        | 0.6333        | 0.8500        | 0.8000        | 0.6500        | 0.5834        | 0.6667        | 0.6666        |
| 7              | 0.6667        | 0.9833        | 0.7667        | 0.9667        | 0.9833        | 1.0000        | 1.0000        | 0.9000        | 0.8333        |
| 8              | 0.8666        | 1.0000        | 0.9334        | 1.0000        | 0.9333        | 1.0000        | 1.0000        | 1.0000        | 1.0000        |
| 9              | 0.5167        | 0.8834        | 0.9667        | 0.8000        | 0.8834        | 0.9667        | 0.8667        | 0.8333        | 0.8667        |
| 0              | 0.8834        | 1.0000        | 0.9000        | 0.6333        | 0.9000        | 1.0000        | 0.9333        | 0.9500        | 1.0000        |
| <b>Average</b> | <b>0.7233</b> | <b>0.8500</b> | <b>0.8383</b> | <b>0.8417</b> | <b>0.8800</b> | <b>0.8917</b> | <b>0.8800</b> | <b>0.9000</b> | <b>0.9000</b> |

After training, the output of OB layer is stored as a cognition standard. We give the spoken digit in unknown class as input and get the output vector from OB layer. The nearest neighbor principle is employed to recognize the speech according to the Euclidean distances from unknown spoken digit to the different centers (attractor).

### B. Experiment Results

Mandarin digital speeches are collected using microphone in quite condition. The data is the pronunciation of 0~9 in Mandarin. Each digit is read 30 times. Each digital speech is learned 10 times. The experiment results are shown in Table I, II and III. In the Table II and III, the first column denotes the different Mandarin digital speech and the first row denotes the dimension of feature matrix,  $m*n$ . And  $m$  denotes the dimension of feature extracted from one frame of each digital speech, so if the digital speech is divided into  $r$  frame, the dimension of feature matrix is  $m*r$ . Based on FCM or HKM, the feature matrix of  $m*r$  can be clustered into the feature matrix of  $m*n$  ( $r>n$ ), so  $n$  denotes the number of class after clustering each digital speech. In Table I, the performance of KIII model is compared with BP neural network. And the parameters of BP neural network are set as follows: there are 40 neuron and hyperbolic tangent sigmoid transfer function in hidden layer; there are 10 neuron and transform function is linear transfer function in output layer; the learning function is gradient descent and the training set is equal to the KIII model's (100 samples, 10 samples of each classes).

From Table II and III, it is shown that the performance of FCM is better than the performance of HKM. But from Table I, it is shown that both of them (FCM+KIII and HKM+KIII) are better than BP neural networks.

### V. CONCLUSION

The KIII network is derived directly from the biological neural system. Its mechanism for pattern recognition is totally different from other ANNs. In this paper, FCM is used to extract feature of Mandarin digital speech and KIII model is used as classifier to recognize the different digital speech. Experimental results show that KIII model has better performance and has the potential to recognize more complex pattern. Otherwise, it is also shown that FCM clustering is suitable to speech recognition than HKM clustering. Taking after the biological olfactory neural system being good at learning new odors, KIII model requires only a few learning trials to set up the basins of attraction of the memory patterns with robust tolerance for noise and speech variability. But, BP neural networks needs about hundreds of thousands trials. Mimicking biological neural systems is shown to be an efficient way to handle complicated pattern recognition problems. But many problems, such as the optimization of parameter, are open and are the targets of our further work.

## REFERENCES

- [1] L.W. Zhao, *Chinese Speech Recognition Based on HMM and ANN*, Ph.D. Dissertation, Harbin Engineering University, 2006. (in Chinese)
- [2] Y. Deng, T. Huang and B. Xu, "Towards high performance continuous mandarin digit string recognition," *Proc. Sixth International Conference on Spoken Language Processing*, Beijing, China, Oct. 2000, pp.642-645.
- [3] E. Trentin and M. Gori, "A survey of hybrid ANN/HMM models for automatic speech recognition," *Neurocomputing*, vol.37, pp.91-126, April 2001.
- [4] I.A. Maaly and M. El-Obaid, "Speech Recognition using Artificial Neural Networks," *Proc. Second Information and Communication Technologies (ICTTA '06)*, Damascus, Syria, Apr. 2006, pp.1246-1247.
- [5] R.C. Price, P. Willmore, W.J. Robert, K.J. Zyga, "Genetically Optimized Feedforward Neural Networks for Speaker Identification", *Fourth International Conference on Knowledge-Based Engineering System & Allied Technologies*, Sep.2000, Brighton, U.K.
- [6] Y. Yao, W. J. Freeman, "Pattern recognition in olfactory systems: modeling and simulation," *Proc. Of the International Joint Conference on Neural Networks (IJCNN-1989)*, Wash. D.C., June 1989, pp. 699-704.
- [7] G. Li, J. Zhang, Y. Wang, W. J. Freeman, "Face Recognition Using a Neural Network Simulating Olfactory Systems," *Lecture Notes in Computer Science*, vol. 3972, pp. 93-97, Jun. 2006.
- [8] J. Zhang, G. Li, W. J. Freeman, "Application of Novel Chaotic Neural Networks to Text Classification Based on PCA," *Lecture Notes in Computer Science*, vol.4319, pp. 1041-1048, Dec. 2006.
- [9] L. Rabiner and B. Hwang Juang, *Fundamentals of Speech Recognition*, Prentice Hall PTR, 1993.
- [10] J. Zhang, G. Li, W. J. Freeman, "Application of Novel Chaotic Neural Networks to Mandarin Digital Speech Recognition," *Proc. of International Joint Conference on Neural Networks 2006 (IJCNN'06)*, Vancouver, BC, Canada, Jun. 2006, pp. 653- 658.
- [11] J. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Norwell: Plenum Press, USA, 1981.
- [12] J.C. Dunn, "A Fuzzy Relative of the ISODATA Process and its Use in Detecting Compact, Well Separated Clusters", *Journal of Cybernetics*, vol.3, pp. 32-57, 1974.
- [13] W. J. Freeman, Y. Yao, B. Burke, "Central pattern generating and recognizing in olfactory bulb: a correlation learning rule," *Neural Networks*, vol. 1, pp. 277-288, 1988.
- [14] H.J. Chang, W. J. Freeman, "Local homeostasis stabilizes a model of the olfactory system globally in respect to perturbations by input during pattern classification," *Int. J. Bifurcation and Chaos*, vol.8, pp. 2107-2123, Nov. 1998.
- [15] R. Kozma, W. J. Freeman, "Chaotic resonance – methods and applications for robust classification of noisy and variable patterns," *Int. J. Bifurcation and Chaos*, vol.11, pp. 1607-1629, Jun. 2001.
- [16] X. Li, G. Li, L. Wang, W. J. Freeman, "A Study on a Bionic Pattern Classifier Based on Olfactory Neural System," *International Journal of Bifurcation and Chaos*, vol. 16, pp: 2425-2434, Aug. 2006.
- [17] S. Albayrak, F. Amasyali, "Fuzzy c-means clustering on Medical Diagnostic Systems", *Proc. International XII. Turkish Symposium on Artificial Intelligence and Neural Networks*, Antalya, Turkey, Jul. 2003.
- [18] J.S. R. Jang, C.T. Sun, E. Mizutani, *Neuro-Fuzzy and Soft Computing*, NJ:Prentice Hall, 1997.