Study on Complexity of EEG Time Series Under Different Mental States*

Hu Xiao Yao Dezhong

(Dept. of Automation, UEST of China Chengdu 610054)

Abstract In this paper, two kinds of surrogate data are designed according to electroencephalography (EEG) data. They are a binomial random series with the same source entropy as that of EEG and an AR series that is the output of AR model constructed from EEG data and driven by white noise. Significant differences among the complexity of these three series show that EEG is far from being random and has certain pattern in it, which can be partially modeled by AR. The complexity of EEG series recorded during three different mental states is computed and the result of two-way analysis of variance shows that the complexity can discriminate the three mental states significantly.

Key words electroencephalography; mental state; nonlinear; complexity; analysis of variance

EEG is both the temporal and spatial summary of the electric activities of millions of cortex neurons interconnected by axons and dendrites. The way neurons function and the complicated neural network formed by them are nonlinear in nature. Recent studies on EEG time series have shown that its nonlinear dynamic analysis may be a new promising technique for studying the intrinsic non-linearity of brain[1~3].

One popular nonlinear analysis of EEG is the phase space reconstruction method. Correlation dimension D_2 and Lyapunov exponents (LE) are estimated as relative measures for comparing different brain states. However, much more efforts should be given to improve the algorithm to calculate them, especially for short data segment.

In this paper, the algorithm complexity was used to characterize three different brain states[4,5]. Though similar tests have been reported[6~10] , it has never been verified that the decrease of EEG complexity, when a mental task is under progress, is caused by the formation of certain pattern in EEG not by the mere reduction of randomness, namely the decrease of source entropy. One of the main contributions of this paper is a method based on proposed surrogate data test to tackle this problem. The detail is given in the following section.

1 Methods

1.1 Experiment Setup and Data Preprocessing

Five healthy male subjects participated in the experiment. Most subjects were graduate students from the automation department of UESTC. Their mean age was 25.2.

EEG recordings were based on referential montage. A_1 and A_2 were connected to serve as the reference electrodes, which were placed on the lobes of subjects. Data was recorded using SHIP9316 EEG data acquisition system with analog filter setting from 0.5 Hz to 30 Hz and a sampling rate of 160 Hz. The total channel number was 16 and the 16 electrodes were positioned according to the popular international 10~20 system. Explanation of denotations of the 16 electrodes (Fp_1,Fp_2,et al) can be found in Ref. [11].

EEG from each subject was recorded at three metal states: wake with eyes closed, multiplying by heart, visualizing a 3-D object in brain. The first state served as a control.

Received October 27,1998 and revised November 25,1998

* The projcet supported by the Huo Yingdong Foundation for young investingatior and the National Nature Science Foundation of China,No.39770215

After recording, off-line IIR filter was implemented to remove the 50Hz noise and all EEGs were inspected visually to check for eye movements and other artifacts. On the basis of this processing, one 20-second artifact-free epoch per state was kept for each subject and saved as binary file for further analysis.

1.2 Algorithm to Compute Complexity of EEG

According to the algorithm to compute the complexity of time series introduced in Ref.[4~6], one has to scale the raw EEG data down to more coarse level. In this case, we supposed data larger than the average to be 1,otherwise to be 0 and thus we got a binary EEG series.

Here we briefly review the procedure to calculate the complexity $c(n)$ of a given binary string[4]. Let S (s_1, $s_2 \ldots s_n$) denote the binary string whose complexity is to be calculated where s_i is a letter of S, which is 1 or 0. And n is the length of S. A sub-string of S, which starts at the position i and ends at the position j, is denoted as $S(i, j)$, where i and j satisfy $1 \le i \le j \le n$. Vocabulary of S is denoted as $V(S)$, which is the set formed by all the sub-strings of S. Let Q (q_1, $q_2, \ldots, q_m$)denote another binary string. Then SQ is the concatenation of S and Q, and $SQ\pi$ is the string(s_1, $s_2, \ldots, s_n$, q_1, $q_2, \ldots, q_{m-1}$).

To illustrate the above definitions, an example is given as follows

Given S=(00101),Q=(101) ,then $V(S)$={0,1,00,10,01,001,010,101,0010,0101,00101}

$V(Q)$={0,1,10,01,101},SQ=(00101101) and $SQ\pi$ =(0010110)

Using the above denotation we proceed to compute the complexity of a string (s_1, $s_2, \ldots, s_n$). At start, $c(n)$=1,S=(s_1), let Q=(s_2),so $SQ\pi$=(s_1).Now if we can find Q in the $V(SQ\pi)$, a new letter,s_3 ,is added to the Q. So S=(s_1),Q=(s_2, s_3)and $SQ\pi$=(s_1, s_3) . Else Q is cascaded to S, S=(s_1 s_2), and Q is re-formed by adding the next letter s_3 in the string, Q=(s_3). This procedure is repeated until reaching the end of the string. Every time we cascaded Q to S, we add 1 to $c(n)$. The final $c(n)$ plus 1 is the complexity of string(s_1, $s_2, \ldots, s_n$).

For example, (0010)'s complexity can be obtained as follows:

1) $c(n)$=1,S=(0),Q=(0),SQ=(00), $SQ\pi$=(0),Q belongs to $V(SQ\pi)$.

2) $c(n)$=1,S=(0),Q=(01),SQ=(001),$SQ\pi$=(00), Q does not belongs to $V(SQ\pi)$,$c(n)$=$c(n)$+1.

3) $c(n)$=2,S=(001),Q=(0),SQ=(0010), $SQ\pi$=(001),Q belongs to $V(SQ\pi)$.

4) $c(n)$=$c(n)$+1=3,S=(0010).

So, (0010)'s complexity=3.

It has been proved[4,5]

$$\lim_{n \to \infty} c(n) = b(n) \equiv n / \log_2 n$$

where $b(n)$ gives the asymptotic value of $c(n)$ when n approximates to ∞ and $c(n)/b(n)$ gives a relative measure of complexity of a given string. In the following presentation, we will use $c(n)/b(n)=K_C$ as the complexity measure of time serie[4, 5].

1.3 Using Surrogate Data to Reveal the Underlying Pattern in EEG

For a random binary series, the asymptotic value $b(n)$ can be reached by $c(n)$ when the binary string's letter has equal probability to be 0 or 1.On the other hand, when this probability deviates from 0.5 then the asymptotic value will deviate from $b(n)$ and it tends to be $h*b(n)$ where[5]

$$h = -p \log_2 p - (1-p) \log_2 (1-p)$$

is the source entropy that has its maximum value 1.0 at p=0.5 and p is the probability of finding 1 or 0 in the string.

So the deviation of $c(n)/b(n)$ from 1 can be attributed to two factors: one is the existence of certain pattern in the symbolic representation of system, which characterizes the formation of certain order in the underlying system, the other is the reduction of source entropy of system, which means that the underlying system remains random and the only thing has changed is the probability of system's state.

Now, we need to know whether the deviation of K_C of EEG from 1 is due to the change of source entropy or due to certain pattern formed in EEG data.

We used two kinds of surrogate data.The first one was the random series from binomial distribution that had the same source entropy as the EEG series. If the complexity measure of EEG was largely determined by its randomness, the K_C of EEG would be comparable to that of its random series counterpart. Source entropy of binary EEG series was estimated from its histogram. If total number of data points is N, and $n_1(n_2)$ is the number of 0(1) in a series , source entropy h of this series is estimated as.

$$p_1 = n_1 / N; p_2 = n_2 / N; \quad h = -p_1 \log_2 p_1 - p_2 \log_2 p_2$$

The second surrogate data was based on AR model. First, we obtained the AR model coefficients from the real EEG data by Burg algorithm[12] and the order of model was set to be 9. Then white noise whose variance was determined by the model was inputted to the system to generate the surrogate data, which was converted to binary series using the same method as to convert EEG series. When $S(n)$ denotes the EEG signal and $S'(n)$ the output of the model constructed from $S(n)$, $S'(n)$ could be viewed as the pseudo-EEG, then we proceeded to compute the K_C of its binary series.

We know that AR model incorporates both stochastic and determinate information of a series[12]. Its stochastic information is reflected in the white driven noise and determinate information in the model coefficients, which we suppose correspond to the underlying pattern of EEG. If EEG is nonlinear in nature, the linear AR model only preserves part of its information. So we deduce that K_C of the surrogate data generated by AR model will be less different from that of raw EEG data than binomial random series.

2 Results

2.1 Results of Surrogate Data Test

EEG segments were extracted from the saved data files of the 16 channels of each subject in control state. Segment duration was set to 7s. Thus a total number of 80 (e.g. 16 channels×5 subjects) segments were obtained. Surrogate data was generated according to its corresponding EEG segment.

Means and standard deviations of K_C of binomial random series, AR series and EEG series are given in Tab. 1.

Table 1 Mean and Standard deviation of K_C of EEG

Series	EEG	Binomial	AR
Mean	0.331 9	0.684 8	0.346 4
Std	0.037 4	0.050 2	0.040 7

At a first glance, Fig.1 shows that there exists apparent difference between binomial random series and EEG. In order to check the difference of K_C among the three series, t-test was used to compare K_C of AR series and binomial random series with that of EEG respectively. The hypothesis was:

H0: mean of K_C of 80 EEG segments was the same as that of both AR series and binomial random series. H0's alternative hypothesis: mean of K_C of 80 EEG segments was smaller than that of both AR series and binomial random series.

T-test shows that K_C of binomial random series, though their source entropy equals to their EEG counterpart, is significantly (p=0) larger than that of EEG, and that K_C of AR series also is significantly larger than that of EEG at a less significant level (p=0.05).

In a word, AR series constructed from the model determined by its corresponding EEG segments preserves some underlying determinant information of EEG. Binomial random series is constructed based on the entropy information of EEG, which solely describes randomness in EEG. The above results of surrogate test show that EEG series is far from being random and it is its underlying pattern that significantly reduces its complexity and that because underlying system generating EEG series is complicated and nonlinear in nature, AR model can only reveal some of its pattern, which makes AR series' complexity is significantly different from that of its EEG counterpart.

2.2 Results of the Experiment of Three Different Mental States of Brain

To compute K_C of EEG, we decomposed the 20 s artifact-free epoch of EEG into 7s segments with overlapped 6 s duration. Thus we got 13 segments for each state of a subject and the final K_C was the average of these 13 segments.

Mean values of K_C at sixteen channels during the three mental tasks are plotted in Fig.2.Horizontal axis denotes channel number N_c and the mapping of N_c to the denotation of electrodes is as this: Channel:1-FP1,2-FP2, 3-F3, 4-F4, 5-C3, 6-C4, 7-P3, 8-P4, 9-O1, 10-O2, 11-F7, 12-F8, 13-T3, 14-T4, 15-T5, 16-T6. Based on the two-way analysis of variance (ANOVA2) [13] (two within-subject factors were brain state and locus of electrodes), it is revealed that there is a significant overall decrease of complexity during. visualization and multiplication and that the locus effect is also significant. There is no significant interaction between two factors.

The results of these experiments are consistent with the previous report[8]. The decrease of complexity

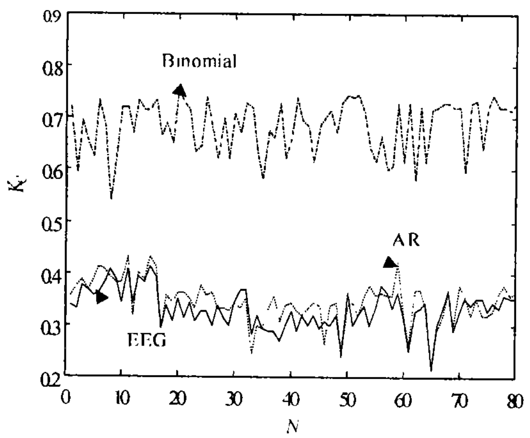

Fig.1　K_C versus sample number

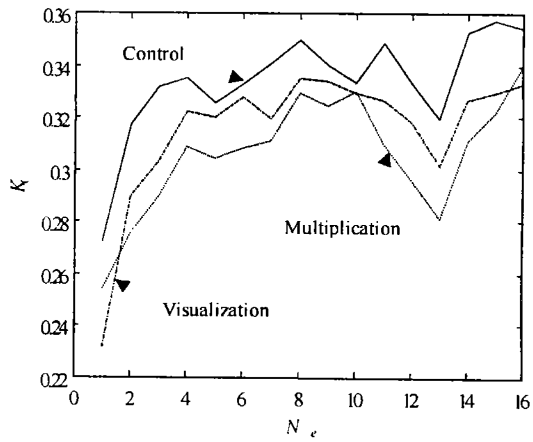

Fig.2　K_C of EEG series under 3 mental states

of EEG during visualization and multiplication compared to the control state may be explained as this: when one is thinking, whether it is an imaging or logical one, his neurons have to work in a more "cooperative" and synchronous mode and this mode reduces the complexity of EEG. As for the difference between visualization and doing multiplication, it may be due to the fact that all subjects reported to us that doing multiplication made them pay more attention than visualization did.

3 Conclusion

Nonlinear dynamic analysis is a promising tool to disclose much implicit information in EEG. K_C can be calculated with reasonable computation burden and used to characterize underlying pattern in EEG, which is sensitive to the change of brain states.

Reference

1 Pritchard W S. A tutorial teriew of the use and benefits of non-linear dynamical analysis of the EEG in analysis of the electrical activity of the brain. New Jork:John Wiley&Sons Ltd, 1997:4~18

2 Popivanov D, Mineva A, Dushanova J. Tracking EEG signal dynamics during mental tasks. IEEE EMB, 1998:89~95

3 Stam C J, van Woerkom T C A M, Pritchard W S. Use of non-linear EEG measures to characterize EEG changes during mental activity. Electroencephalography and clinical Neurophysiology, 1996,99:214~224

4 Lempel A, Ziv J. On the complexity of finite sequences. IEEE Trans on I T,1976, 22(1):75~81

5 Kaspar F, Schuster H G. Easily calculable measure for the complexity of spatiotemporal patterns.Physical Review A. 1987, 36:842~848

6 徐京华，吴祥宝.以复杂性测度刻划人脑皮层上的信息传输.中国科学（B辑）,1994,24(1): 57~62

7 刘建平，郑崇勋，马建青.不同睡眠期脑电图复杂性研究.中国生物医学工程学会95学术年会论文集，1995 ,203~204

8 顾凡及，范思陆，阮 炯等.不同状态下脑电图复杂性探索. 生物物理学报，1994,10(3): 439~445

9 欧阳楷，孟 欣.基于符号序列的生物信号处理方法. 中国生物医学工程学会 95 学术年会论文集，1995 :268~270

10 童勤业，刘刚强，刘加友. EEG 的信息熵分析. 生物物理学报，1993,13(9): 645~649

11 John R. Hughes 著，马仁飞译. 临床实用脑电图学. 第二版. 北京：人民卫生出版社，1997：1~11

12 杨福生，高上凯. 生物医学信号处理. 北京：高等教育出版社，1989：318~326

13 Olive Jean Dunn, Virgina Clark A. Applied statistics: analysis of variance and regression. New York:John Wiley&Sons World Publishing Corp, 1987:123~151

不同精神状态下EEG序列复杂性研究*

胡 晓** 尧德中

（电子科技大学自动化系 成都 610054）

【摘要】 构造了与脑电的源熵相同的（0，1）分布随机序列，以及用白噪激励依据脑电构造出的AR模型而得到的AR序列来作为伪脑电信号。通过比较这三种序列的复杂度，证明了脑电远非随机信号，而是存在某种模式，这种模式可以由AR模型部分表示出来。在此基础上，对三种精神状态下的脑电序列的复杂度进行了双因素方差分析，结果表明复杂度可以显著地区分这三种状态。

关 键 词 脑电图；精神状态；非线性；复杂度；方差分析

中图分类号 R318.4

* 国家教委霍英东青年基金和国家自然科学基金资助项目，基金号：39770215

** 男 24岁 硕士生