Kolmogorov complexity of finite sequences and recognition of different preictal EEG patterns

Conference Paper *in* Proceedings of the IEEE Symposium on Computer-Based Medical Systems · July 1995

DOI: 10.1109/CBMS.1995.465426 · Source: IEEE Xplore

1 author:

Arthur Petrosian
American University of Armenia

Kolmogorov Complexity of Finite Sequences and Recognition of Different Preictal EEG Patterns

Arthur Petrosian, Member, IEEE

Texas Tech University Health Sciences Center,
Lubbock, TX 79430

Abstract*. The problem of an adequate quantitative interpretation of epileptic EEG recordings is of great importance in the understanding, recognition, and treatment of epilepsy. In recent years, much effort has been made to develop computerized methods which can characterize different interictal, ictal, and postictal stages. The main issue of whether there exist a preictal phenomenon is unresolved. In the present work we address this issue making use of the most basic representation of data complexity, namely, the algorithmic information content. In general this measure, also known as Kolmogorov complexity, represents the compressibility of the data strings. It can also be used to describe properties (linear and nonlinear) of the underlying dynamical system. We analyze Kolmogorov complexity and related characteristics of intracranial EEG recording, containing preictal, ictal, and postictal segments.*

I. Introduction

In his late work [1] Andrey Kolmogorov introduced a new algorithmic approach to the quantitative definition of information, that used the concept of recursive functions. He analyzed its relation with respect to two other common approaches: combinatorial and probabilistic. Alternative to the probabilistic notion of the information content I(x,y) conveyed by one random object x with respect to another y, he linked a complexity of an object to that of an algorithm which can generate it from another one. Naturally, all these measures are directly related to data compressibility and coding characteristics [2]. It was mentioned in [1] that although the proposed approach yields in principle a correct definition of the quantity of “hereditary information”, it would be difficult to obtain a reliable estimate of it. Later, however, an easily calculable measure for finite sequences was suggested [3,4] which closely relates to their Kolmogorov complexity. An attempt to make use of such measures to distinguish different EEG signal patterns was made recently in [5]. This was motivated by the presumption that various EEG time series contain a mixture of limit cycle (chaotic) and random attractors. Despite the numerous positive results and investigations of EEG time series based on purely nonlinear dynamical characteristics (e.g. fractal dimension of the strange attractor), the limitation of these studies is that a certain brain functional state may often last only for a brief time period, whereas, in order to obtain reliable estimates for the mentioned characteristics, long-term time series are needed. Also, the algorithms involved are extremely sensitive to the noise. Therefore, better alternative EEG analysis methods need to be discovered [6]. The purpose of this work is to explore the relationship between various informational measures, including Kolmogorov complexity, information content, and fractal dimension of epileptic EEG series. The special point of interest is to determine whether the combined analysis of these characteristics can provide a reliable tool for distinguishing preictal stages, and thus may be useful for prediction of epileptic seizures.

II. Methods and Results

The EEG data processed were obtained from an epileptic patient with depth electrodes surgically implanted into the epileptic focus. A 32-channel EEG with a sampling rate of 200 Hz was recorded using Stellate's Monitor System. The visual based segmentation of these recordings into interictal, preictal, ictal, and postictal stages was performed by an expert epileptologist to provide feasible transitional changes. In addition, to designate at what point preictal EEG changes indicated that an epileptic seizure would inevitably follow, the preictal stage was further divided into preictal 1 (seizure non inevitable) and preictal 2 (seizure inevitable) segments. Examples of these extracted patterns are presented in Fig.1. The short-term epochs of each interictal, preictal 1, preictal 2, ictal (seizure), and postictal stages considered are shown. The total recording, stored for computer analysis, contained about 1.5 min. interictal/preictal stage, 45 sec. of ictal stage (seizure), and 30 sec. of postictal stage.

In this study we investigated the behavior of several basic informational measures for the described EEG signal activity during interictal/preictal/ictal/postictal stages. Given an EEG segment - a sequence of variables capable of taking values *k*, *min≤k≤max*, in a finite set of *n=max-min+1* sample values - we first consider along with an entropy of variable:

$$H_1 = \sum_k p(k) \cdot \log p(k) \,,$$

the conditional entropy: $H_2 = \sum_k p(k) \cdot h_k(l) = \sum_k p(k) \cdot \left[\sum_l p_d(k,l) \cdot \log p_d(k,l) \right]$,

where $p_d(k,l)$ is the joint probability of the signal samples *k,l*, occurred at a distance *d* over the entire considered segment. In our previous work [6] we made use of the latter matrix $p_d(k,l)$ of second order histogram to extract and analyze signal texture content for seizure prediction (we will put further on *d*=1, as it was in [6]). It is to be emphasized, that the conditional entropy H_2, $H_2 \leq H_1$, represents, in this case, the interrelationship between neighboring samples-variables, and that the average amount of information, $I = H_2 - H_1$, characterizes the "closeness of the relation" between them

[1]. Yet, these quantities are related to signal fractal dimension and its algorithmic complexity. The fractal dimension of planar curve is defined as [7]:

$D=\log(N)/(\log(N)+\log(d/L))$, where N is the number of samples in a segment, d is a planar diameter of the waveform, and L is the total length of it. This measure actually represents the precise quantity of spatial information embodied in a waveform pattern [7]. The Kolmogorov complexity $c(N)$ of the signal samples sequence, on the other hand, is given by the length of the shortest binary program which can generate that sequence [4]. We will make use of quantitative measure $h(N)$, $h(N)=c(N)/b(N)$, $b(N)=N/\log N$, istead, which was proposed by Lempel and Ziv [3] and was proven to be an appropriate measure of signal Kolmogorov complexity. The computations of $c(N)$ were based upon the algorithm described in [4]. The results are presented in Fig. 2.

We also converted the data into binary sequences prior to obtaining the described informational measures from time series. Several simple conversion methods were applied:

a) average method - the EEG sample was assigned to be 1 if it was above the signal average value, and 0 otherwise;

b) modified zone method - it was assigned to be 1 if it was out of bounds of average plus or minus standard deviation, and 0 otherwise;

c) “differential” method - the sample was given value 1 if the difference between two consecutive samples is positive, and 0 if it is negative;

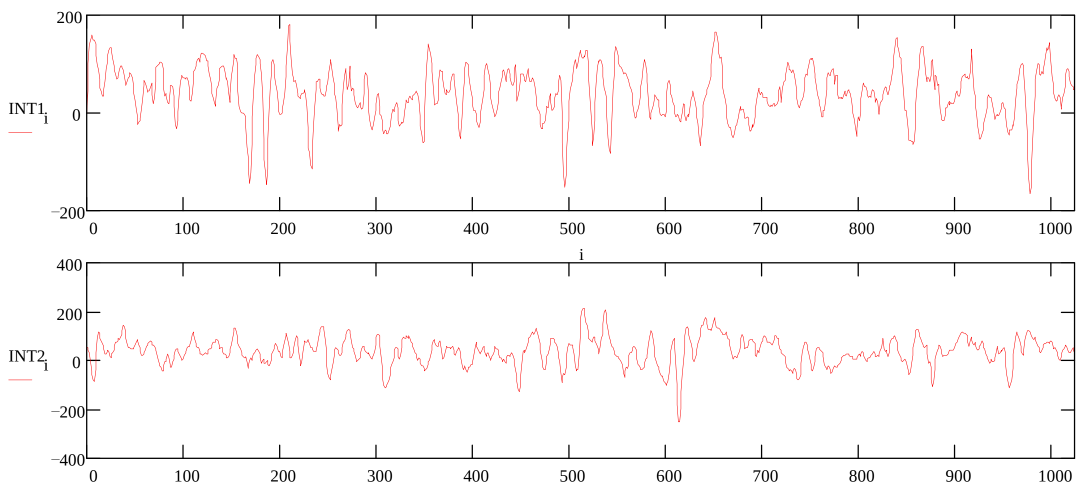
d) “zone differential” method - the consecutive samples with positive or negative “differentials” were given value 1 if the “local extremum” exceeded standard deviation value, and 0 otherwise;

e) “modified zone differential” method - similar to above d) with an *apriori* chosen boundary value Δ instead of standard deviation value.

Fig. 3 represents the computed measures over the same segments converted to the binary sequences according to the methods a-e. The computation of fractal dimension was based on the following approximate relation [7]:

$$D\cong\log(N)/(\log(N)+\log(N/(N+1.4N_\delta-N_\delta)),$$

where N_δ is the number of dissimilar pairs in the underlying binary sequence.



1) INTERICTAL STAGE

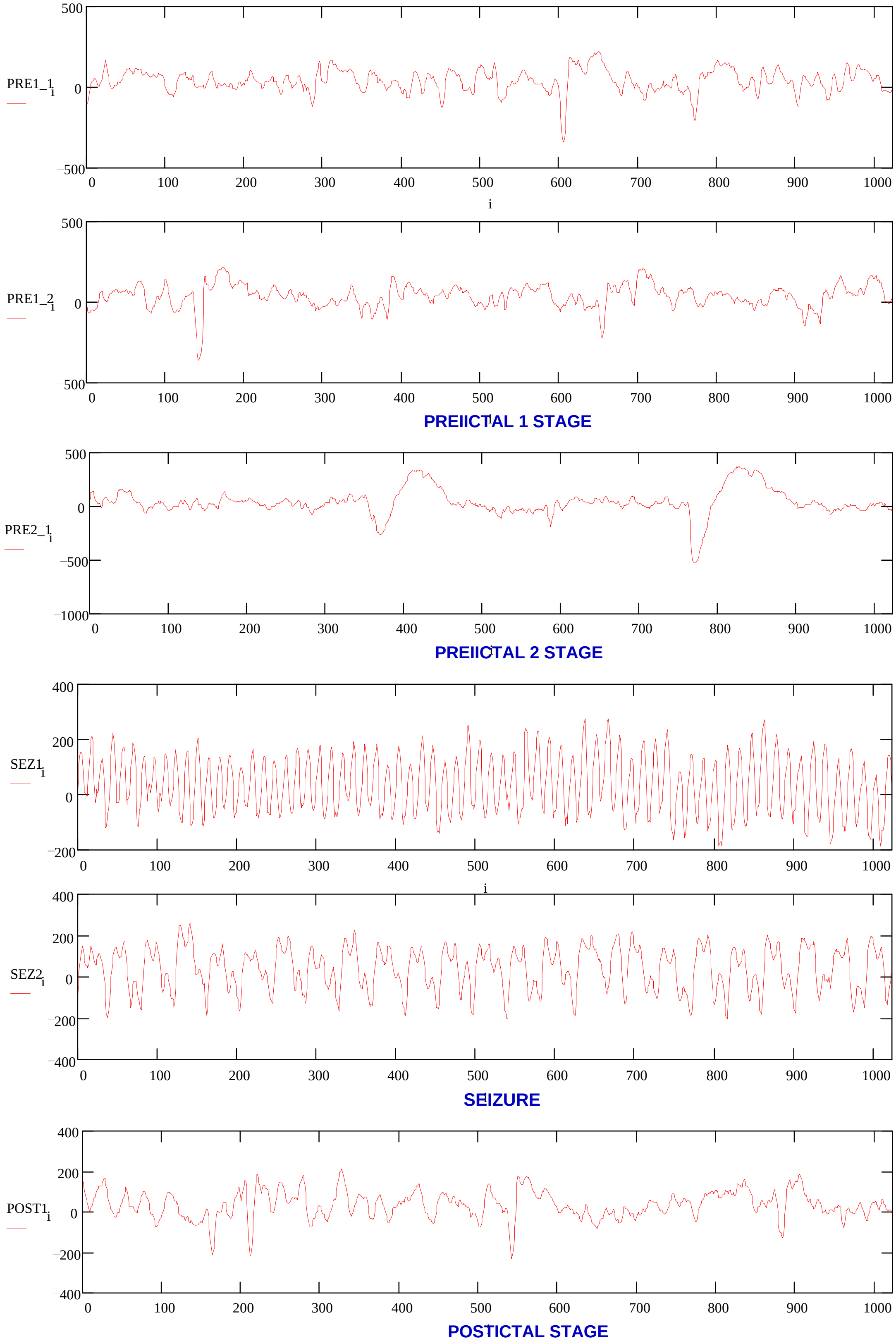


Fig.1. The epochs of about 5 sec. each, extracted from interictal, preictal 1, preictal 2, ictal, and postictal segments.

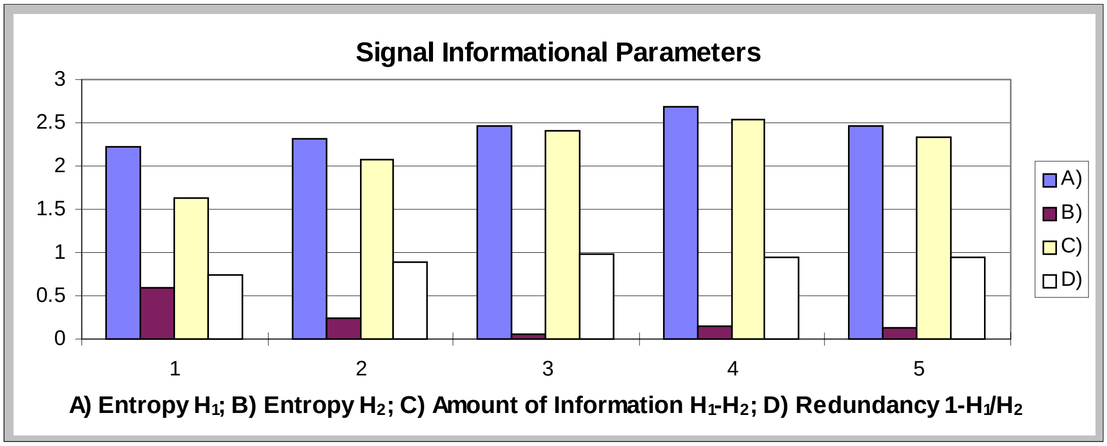


Fig. 2. The columns 1,2,3,4,5 represent values for interictal, preictal 1, preictal 2, ictal, and postictal segments respectively.

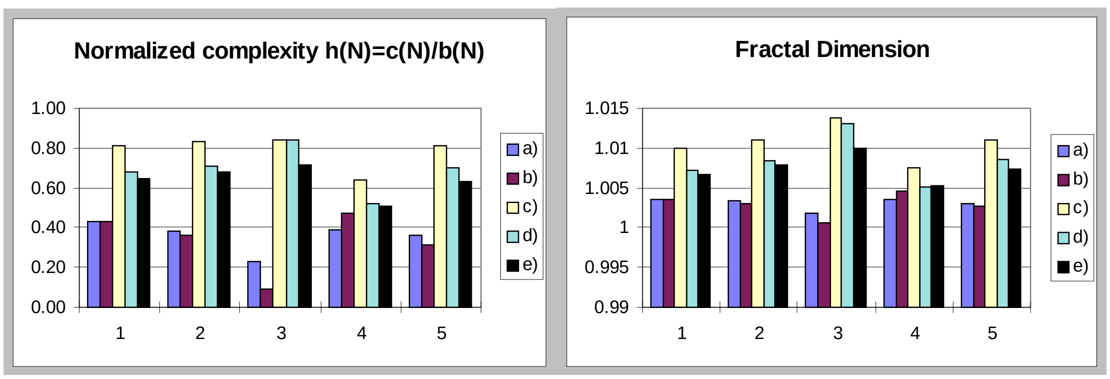


Fig. 3. 1 - Interictal; 2 - Preictal 1; 3 - Preictal 2; 4 - Ictal; 5 - Postictal; a)-e): described in the text signal→binary conversion methods.

Thus, the behavior of displayed measures over the entire EEG segment depends on which signal→binary conversion method was chosen prior to their computation. As it follows from Fig. 3, the normalized complexity as well as fractal dimension of converted sequences tend to decrease during preictal stages and increase during seizure if methods a) and b) were chosen. On the other hand, the opposite tend for these measures was observed if methods c), d), and e) were applied.

III. Discussion

This study represents an attempt to incorporate the basic informational characteristics for analysis of EEG signals. In particular, it shows the feasibility of combined consideration of those measures for recognition of epileptic EEG patterns. However, the data are preliminary as only one subject was recorded. Obviously, the considered original-to-binary conversion methods can be optimized for a given application by the parameters involved. For instance, the boundary values of methods b) and e) can be adjusted to provide the best recognition in a specific task. In addition, the essence of differences of calculated measures when different conversion methods are applied needs to be explored. This, combined with interrelationships of these measures may provide guidelines for distinguishing different preictal EEG patterns. Future studies will determine the optimal length of analyzed segments based upon testing on larger data sets, the value of higher resolution EEG recordings and the findings during sleep REM and non-REM segments.

Acknowledgment.

Author would like to thank Dr. Richard W. Homan (Professor and Chairman, Department of Neurology, Texas Tech University Health Sciences Center) for providing an expert review of patient data.

References.

1. A.N. Kolmogorov. Three approaches to the quantitative definition of information. *Problemy Peredachi Informatsii*, Vol. 1, No.1, pp. 3-11, 1965 (in Russian).

2. A.A. Petrosian. Optimal zonal coding of digital signals by transforms. *Problemy Peredachi Informatsii*, Vol. 27, No.2, pp. 46-58, 1991 (in Russian).

3. A.Lempel, J.Ziv. On the complexity of finite sequences. *IEEE Transactions on Information Theory*, IT-22, pp. 75-81, 1976.

4. F.Kaspar, H.G.Schuster. Easily calculable measure for the complexity of spatiotemporal patterns. *Physical Review*, Vol.36, No.2, pp. 842-848, 1987.

5. Q.Guo, G.Jin, Y.Shu, G.Zhou, J.Zhang. Zone method - a new approach to brain nonlinear dynamics. *IEEE Proc., 16th EMBS Intl. Conference*, Nov.3-6, 1994, Baltimore, MD, pp.1322-23.

6. A.Petrosian., R.W.Homan The analysis of EEG texture content for seizure prediction. *IEEE Proc., 16th EMBS Intl. Conference*, Nov.3-6, 1994, Baltimore, MD, pp.231-32.

7. M.J. Katz. Fractals and the analysis of waveforms. *Comput. Biol. Med.*, Vol.18, No. 3, pp.145-156, 1988.