Recurrence Time Distribution, Renyi Entropy, and Pattern Discovery

Jianbo Gao[1], Yinhe Cao[2], and Jing Hu[1]
[1]Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611
[2]BioSieve, 1026 Springfield Drive, Campbell, CA 95008
e-mail: gao@ece.ufl.edu, contact@biosieve.com

Abstract —

Entropy and recurrence times are two of the most important complexity measures for both random fields and nonlinear dynamical systems. We report a fundamental relation between recurrence time distribution and Renyi entropy of arbitrary integer order for both ergodic random fields and ergodic nonlinear dynamical systems, thus provide an elegant and comprehensive characterization for these two important systems. The fundamental relation is obtained by parameterizing the dynamics in the state space by a discrete symbol sequence and collectively consider recurrences to all non-empty sub-regions of the state space. Event detection using recurrence time statistics is also considered, including speech endpoint detection, epileptic seizure detection/prediction from continuous EEG measurements, and gene identification from genomic DNA sequences.

I. Introduction

Entropy and recurrence times are two of the most important complexity measures for both random fields and nonlinear dynamical systems. For some systems, there is a fundamental relation between the concepts of entropy and recurrence times: in the study of random fields, it has been found that the ratio between the logarithm of the recurrence time and the pattern length approaches to the Shannon entropy when the pattern size becomes infinite [1], while in the study of nonlinear dynamical systems, it has been found that for systems with a uniform mixing property (such as Anosov-type systems), the inverse of the mean recurrence time is the Kolmogorov-Sinai entropy [2]. The majority of the researchers in the field, however, only consider recurrences to a fixed pattern for random fields or a fixed neighborhood for nonlinear dynamical systems. Since different regions of the state space may be visited by a phase trajectory very nonuniformly, considering recurrences to a fixed region or pattern may not be able to reveal the whole spectrum of the complexities of the motion in the state space. One elegant way of characterizing the full spectrum of the complexities of the motion is by the Renyi entropy of different orders. This consideration propels us to ask a fundamental question: if we partition the state space into many sub-regions and collectively consider recurrences to all these sub-regions, can we find (part of) the Renyi entropy spectrum through the distribution for the recurrence times? Most satisfyingly, we have found that under weak mixing conditions, the distribution for the recurrence times can be explicitly found, and that the entire integer-order Renyi entropy spectrum can be readily computed from the distribution for the recurrence times. Thus, recurrence times provide a comprehensive description for both ergodic random fields and ergodic nonlinear dynamical systems.

Below we briefly explain the theory. Then we discuss event detection using recurrence time statistics.

II. Recurrence time distribution and Renyi Entropy

For ease of exposition, in this section, we shall work with symbol sequences. Symbol sequences frequently occur in practice. For example, they may occur naturally, as in a DNA or protein sequence, or in coin–tossing, where one gets a sequence of heads and tails. For nonlinear dynamical systems, they may be obtained by mapping a continuous dynamics to a coarse-grained discrete symbol sequence. This can be achieved by a number of means including partitioning a phase space into many mini-cells, then representing each mini-cell by a symbol, and mapping a continuous variable to a symbolic sequence by using symbolic dynamics technique.

Let us denote our symbol sequence S by $b_1, b_2, \cdots, b_N$, where each element b_i belongs to a state space SP. For a DNA sequence, SP consists of four nucleotide bases A, C, T, and G. For nonlinear dynamical systems, b_i may be interpreted as representing the i-th mini-cell after the state space is partitioned. Next we group a consecutive symbols of window size w together and call that a word of size w. Using maximal overlapping sliding window, we then obtain $n = N - w + 1$ such words. In chaos theory, these words are considered reconstructed vectors with embedding dimension w and delay time 1 [3]. We associate these words with the positions of the original symbol sequence from 1 to n, i.e., $W_i = b_i b_{i+1} \cdots b_{i+w-1}$ is a word of size w associated with the position i along the original symbol sequence. Two words are considered equal if all of their corresponding symbols match. That is, $W_i = W_j$, if and only if $b_{i+k} = b_{j+k},\ k = 0, \cdots, w-1$. $S[u \to v] = b_u b_{u+1} \ldots b_v$ denotes a subsequence of S from position u to v. We are now ready to define recurrence times.

> **Definition**: The recurrence time $T(i)$ for a position i along the original symbol sequence is the smallest $j - (i + w - 1)$ such that $j > i + w - 1$ and $W_j = W_i$. If no such j exists, then there is no repeat for the word W_i after position i in the sequence S, and we indicate such a situation by $T(i) = -1$.

Before we proceed on, we make a few comments: (i) The condition $j > i + w - 1$ ensures that there is no overlapping between W_j and W_i. Defining T to be $j - (i + w - 1)$ amounts to counting the time from the end of the word W_i. (ii) One may more straightforwardly partition the symbol sequence into non-overlapping blocks of length w and call each block a word, then define the recurrence time. When the symbol sequence is infinite and ergodic, these two definitions are equivalent. When the length of the symbol sequence is finite and a concern, then the definition adopted here uses the symbol sequence more efficiently . (iii) For non-ergodic sequences

such as DNA sequences, simple sequence repeats such as a sequence of $AAA\cdots$ are of biological interest. To identify those sequence segments, it is better to modify the above definition by requiring T to be the smallest $j-i$ such that $j>i$ and $W_j = W_i$. In other words, we do allow overlapping between the words when computing the recurrence times in those situations. Note that in practice these minor subtleties may be ignored, since anyway the probabilities for small T are negligible.

Now let us go on to examine the $T(i)$ sequence. We first filter out all those $T(i)=-1$, then estimate the probability distribution function for those positive $T(i)$. We have the following simple but powerful theorem.

> **Theorem 1**: Given a random sequence of words $W_i,\ i=1,2,\cdots$, of infinite length, where there are a total of m distinct words, each occurs with probability p_i, the probability that the recurrence time $T(i)$ being $T\geq 1$ is given by

$$\psi(T) = P\{T(i)=T\} = \sum_{i=1}^{m} p_i^2 \cdot [1-p_i]^{(T-1)} \quad (T\geq 1). \tag{1}$$

We will omit the proof here. To better appreciate the meaning of $\psi(T)$, we make three comments: (i) $\sum_{T=1}^{\infty}\psi(T)=1$. Hence $\psi(T)$ is a normalized probability distribution function. (ii) Eq. 1 is a special superposition of geometrical distributions. Passing to continuous case, it becomes a special mixture of exponential distributions, hence, the random process involved is related to the compound Poisson process. Note that Poisson heuristic is quite essential in establishing the relation between the recurrence time and entropy for a random field [4]. Exponentially distributed recurrence time to a fixed neighborhood has also been observed in dissipative chaotic systems including the chaotic Lorenz attractor, Rossler attractor, and the Henon map [5], and in measure preserving 1-D and 2-D maps [6]. (iii) Since exponential distribution is memoryless, one might think that we have introduced some assumption to obtain this theorem. This is not so. All we have assumed is that the system is ergodic and p_i for the word W_i is well defined. The nature of the symbol sequence, be it independent, or modeled by a Markov chain, or generated by a chaotic dynamics, etc., only affects how p_i for the word W_i is computed. Hence, except excluding periodic sequences, this theorem is very general.

Next, we briefly overview various entropy definitions for the word sequence, to set the stage for discussing how to compute the entire Renyi entropy spectrum from $\psi(T)$. The usual entropy is defined as

$$H(W) = -\sum_{i=1}^{m} p_i \log p_i, \tag{2}$$

where the unit for H is bit or baud corresponding to base 2 or e in the logarithm. When the source is random, such as modeled by a Markov chain, the entropy is called Shannon entropy, and one usually talks about entropy rate per symbol: $h = -(\sum_{i=1}^{m} p_i \log p_i)/w,\quad w\to\infty$. When the symbol sequence considered is from a nonlinear dynamical system, we may think of partitioning the phase space into a number of mini-cells, and interpreting the probability p_i as the joint probability $p(i_1,i_2,\cdots,i_w)$ that the trajectory $\vec{x}(t=\tau)$ is in box i_1, $\vec{x}(t=2\tau)$ is in box $i_2,\cdots$, and $\vec{x}(t=w\tau)$ is in box i_w, where τ is the sampling time. Let the linear size of the mini-cell be ϵ. Then

$$K = -\lim_{\tau\to 0}\ \lim_{\epsilon\to 0}\ \lim_{w\to\infty}\ \frac{1}{w\tau} H(W),$$

where the summation $\sum_{i=1}^{m}$ is understood to be $\sum_{(i_1,i_2,\cdots,i_w)}$, is the Kolmogorov-Sinai entropy.

Next we discuss Renyi entropy. It is defined by

$$H_q(W) = \frac{1}{1-q}\log\Big(\sum_{i=1}^{m} {p_i}^q\Big) \tag{3}$$

$H_q(W)$ has a number of interesting properties:

- When $q=1$, H_1 is the Shannon entropy: $H_1(W) = H(W)$.
- If W_1 and W_2 are two independent random variables, then $H_q(W_1,W_2) = H_q(W_1) + H_q(W_2)$.
- If $p_1 = p_2 = \cdots = p_m = \frac{1}{m}$, then for all real valued q, $H_q(W) = \log(m)$.
- $H_0(W) = \log(m)$ is the topological entropy.
- $H_2(W)$, after divided by $w\tau$ and taking suitable limits, as in the case of Kolmogorov-Sinai entropy, gives K_2 entropy for chaotic systems [7].
- In the case of unequal probability, let $p_{max} = \max_i\ (p_i)$, then $\lim_{q\to\infty} H_q(W) = -\log(p_{max})$.

Let us now compute the entire H_q spectrum from the recurrence time distribution $\psi(T)$. First we note that the mean recurrence time $\overline{T_i}$ for word W_i is $1/p_i$. Hence Shannon entropy is simply given by the mean of $\log\overline{T_i}$. In the limit of infinite word size, the difference between $\overline{T_i}$ and T_i is negligible, comparing to T_i. So is the difference between mean of $\log T_i$ and $\log\overline{T}_i$. Hence, $\log\overline{T_i}$ can be simply replaced by $\log T$, and we recover the well-known Ornstein-Weiss theorem for random fields [1], which states that the entropy rate per symbol is given by $\log T/w$, when $w\to\infty$. For dynamical systems with uniform mixing, we have $p_i\approx 1/m$, hence, the inverse of the mean recurrence time is the entropy [2]. To find other orders of the H_q spectrum, we note that Eq. 3 can be re-written as $\sum_i^m p_i^q = e^{(1-q)H_q}$. Expanding the binomial term in Eq. 1, we have

$$\psi(T) = \sum_{i=0}^{T-1} (-1)^i\ \frac{(T-1)!}{i!\ (T-1-i)!} e^{-(1+i)H_{2+i}} \tag{4}$$

To understand Eq. 4, let us take $T=1,2,3$, etc. We then have:

- $\psi(1) = e^{-H_2}$,
- $\psi(2) = e^{-H_2} - e^{-2H_3}$,
- $\psi(3) = e^{-H_2} - 2e^{-2H_3} + e^{-3H_4}$, etc.

Hence, all integer-order $H_q,\ q\geq 2$, can be simply found from $\psi(T)$.

The other half of the Renyi entropy spectrum can be found by computing moments of the recurrence time T. More importantly, for computational purpose, it is more advantageous to transform Eq. 1, and then use techniques such as poly-log function and find suitable scaling laws. Due to space limitations, however, we will not give the details here. As a prelude to event detection, we wish to emphasize here that the simple definition of recurrence time not only captures all orders of

the Renyi entropy, but also contains information about periodicity – the recurrence time is just the period of a periodic sequence.

III. Discrete case: Gene finding from genomic DNA sequences

To show how useful the definition of recurrence time is in detecting interesting structures, let us first study the problem of gene identification from genomic DNA sequences. Since a DNA sequence is naturally a symbol sequence, the theory developed in Sec. 2 can be applied without any modification.

Finding genes, especially small ones, by computational approaches is an important yet difficult problem. Current computational approaches for finding genes are either based on comparative search or Markov or hidden Markov models. Both approaches require considerable knowledge of a genome sequence under investigation. The recurrence time based method provides a new means of charactering the period-3 feature of coding DNA sequences. Period-3 is due to the fact that three nucleotide bases encode an amino acid, while the usage of the nucleotide bases in the three reading frames are highly biased.

To appreciate how recurrence time provides a new means of charactering the period-3 feature, we have shown in Fig. 1 the probability distributions for the recurrence times not greater than 40, for the genome sequences of four species, E.Coli, Yeast, C. elegans, and the Human. The black and red curves are for the coding and non-coding regions, respectively. Due to the low percentage of non-coding regions in the E.Coli genome, such a curve is not computed. We observe that the black curves all have very well defined peaks at recurrence times of 3, 6, 9, etc. Also note that the black curves are very similar among the four different species. Such period-3 feature can be conveniently used to define a codon index, which we shall denote by RTI:

$$RTI = \sum_{i=1}^{m} [2p(3i) - p(3i+1) - p(3i+2)] \qquad (5)$$

where $p(i)$ is the probability for the recurrence time $T = i$ calculated for a coding or non-coding sequence of length n, and m is a cutoff parameter typically chosen not to be larger than 20 so that very short sequences can be studied.

To evaluate how effective RTI can be used to identify protein coding regions, we have studied the long genome sequences of yeast, C.elegans, and humans. For yeast, our sample pool is comprised of two sets of DNA segments: the coding set (fully coding regions or exons), which contains 4125 verified ORFs, and the non-coding set, which contains 5993 segments (fully non-coding regions or introns). For C.elegans chromosome 3, there are 2904 genes. Our sample pool consists of 18469 exons and 16421 non-coding regions. For the human chromosome 19, there are 2453 genes. Our sample pool for evaluating the accuracy of RTI consists of 19337 exons and 10785 non-coding regions. For the human chromosome 22, there are only 831 genes, hence, our sample pool only consists of 6618 exons and 4068 non-coding regions. Fig.2 shows the specificity and sensitivity curves for the RTI index evaluated on (a) all of the 16 yeast chromosomes, (b) the chromosome 3 of the C.elegans genome, (c) the chromosome 19 of the human genome, and (d) the chromosome 22 of the human genome. The dashed curve is the cumulative distribution function for RTI for the non-coding regions, and the solid curves are the

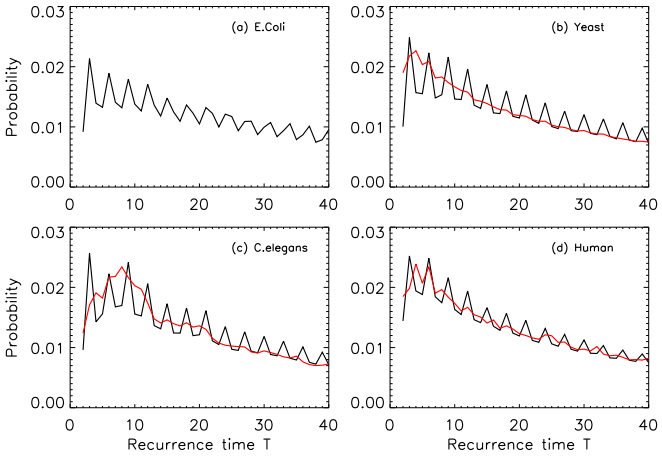

Figure 1: The probability distribution curves computed from the genomes of four organisms studied. The red and black curves are for non-coding and coding sequences, respectively. The window size w is 3 in all the computations. Very similar results have been obtained when $w = 4$ or 5.

complementary cumulative distribution function for the coding regions. To understand the meaning of such curves, let us focus on the intersection of the solid and the dashed curves. When we choose that RTI_0 as a threshold value, then for yeast with 78% probability a coding sequence is characterized as coding sequence, while with 78% probability a non-coding sequence is also taken as a non-coding sequence. It is interesting to note that the percentage of accuracies calculated on the C.elegans and the human genomes are only slightly lower than 78%. This strongly suggests that the method is largely species-independent. It is also worth noting that for the yeast genome, when the period-3 feature is characterized by other methods such as the Fourier transform method, the accuracy is about 71% (see Table 1 below), about 7% worse than the recurrence time based method. It is well known that those methods perform a lot worse when the human genomes are considered. However, the recurrence time based method still performs well.

IV. Continuous case: Speech endpoint detection and epileptic seizure detection/prediction

The close relation between the recurrence time statistics and the Renyi entropy suggests that recurrence time statistics must be very useful for discovering patterns from real world data. This is indeed so, as will be shown below. To set the suitable stage for practical use, we have to first introduce a special type of recurrence time for continuous time systems.

IV.A Recurrence time of the second type

This type of recurrence time was first introduced by Gao [5, 8, 9]. Suppose we are given a scalar time series $\{x(i), i = 1, 2, ...\}$. We first construct vectors of the form: $X_i = [x(i), x(i + L), ..., x(i+(m-1)L)]$, with m being the embedding dimension and L the delay time [3]. $\{X_i,\ i = 1, 2, \cdots, N\}$ then represents certain trajectory in a m-dimensional space. Next, we arbitrarily choose a reference point X_0 on the reconstructed

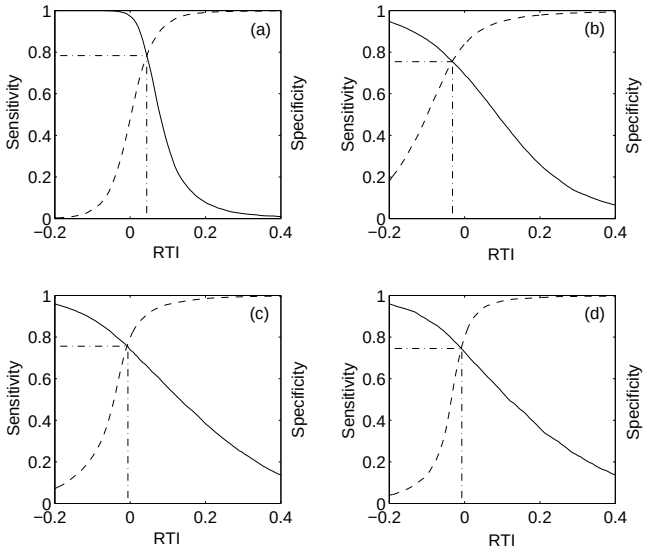

Figure 2: The specificity and sensitivity curves for the RTI index evaluated on (a) all of the 16 yeast chromosomes, (b) chromosome 3 of the C.elegans genome, (c) chromosome 19 of the human genome, and (d) chromosome 22 of the human genome.

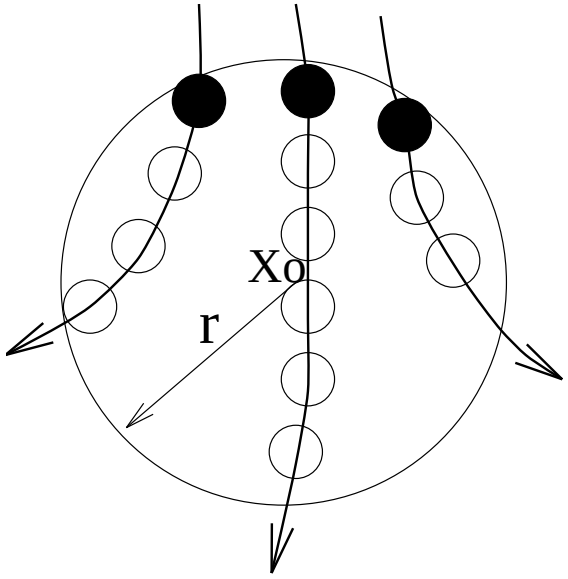

Figure 3: A schematic showing the recurrence points of the second type (solid circles) and the sojourn points (open circles) in $B_r(X_0)$.

trajectory, and consider recurrences to its neighborhood of radius r: $B_r(X_0) = \{X : \|X - X_0\| \leq r\}$. The recurrence points of the 2nd type are defined as the set of points comprised of the first trajectory point getting inside the neighborhood from outside. These are denoted as the dark solid circles in Fig. 3. The trajectory may stay inside the neighborhood for a while, thus generating a sequence of points, as designated as open circles in Fig. 3. These are called sojourn points [5]. It is clear that there will be more such points when the size of the neighborhood gets larger as well as when the trajectory is sampled more densely. The summation of the recurrence points of the second kind and the sojourn points is called the recurrence points of the first kind. These are often called nearest neighbors of the reference point X_0, and have been used by all other chaos theory–based nonlinear methods. Let us be more precise mathematically. We denote the recurrence points of the 1st type to be $S_1 = \{X_{t_1}, X_{t_2}, ..., X_{t_i}...\}$, and the corresponding Poincare recurrence time of the 1st type to be $\{T_1(i) = t_{i+1} - t_i, i = 1, 2, ...\}$. Note the time is computed based on successive returns, not based on the returning points and the reference point. Also note $T_1(i)$ may be 1 (for continuous time systems, this means 1 unit of the sampling time), for some i. This occurs when there are at least 1 sojourn point. Existence of such points makes further quantitative analysis difficult. Thus, we remove the sojourn points from the set S_1 (which can be easily achieved by monitoring whether the recurrence times of the first type are 1 or not). Let us denote the remaining set by $S_2 = \{X_{t'_1}, X_{t'_2}, ..., X_{t'_i}...\}$. S_2 then defines a time sequence $\{T_2(i) = t'_{i+1} - t'_i, i = 1, 2, ...\}$. These are called the recurrence times of the 2nd type. It is clear that for periodic motions, so long as the size of the neighborhood is not too large, $T_2(i)$ accurately estimates the period of the motion. This type of recurrence time is not only related to entropy, but also related to the information dimension of the attractor underlying the data. The conventional energy method basically assumes the dimension of the data to be 1-dimensional. Information dimension is a more accurate way of characterizing an irregular time series. As we shall see, recurrence time method is indeed much more accurate than the energy method in speech end point detection.

IV.B Speech end point detection

In this study, a total of 600 isolated English digits (6 men and 6 women) from the TI46 database were studied. Each speaker pronounced digits zero to nine five times. All of the utterances were manually labeled before the experiment. The sampling rate is 12500 Hz. For generating the noisy speech signals, three different types of noise, white, pink, and babble, from the Signal Processing Information Base (SPIB) collected by Rice University, were used. Pink noise has a power-law decaying power spectral density, and is often called 1/f noise. Babble noise often refers to that the noise has a similar spectrum to that of the speech. Sometimes it may refer to background conversation. In most robust automatic speech recognitions, babble noise is considered the hardest to deal with.

To illustrate the effectiveness of the recurrence time based method for detecting weak transitions, let us study the digit zero, since it contains weak fricatives. The signals shown in Figs. 4(a,b), for the clean and noisy zero, contain 12544 samples. In both figures, two vertical dashed lines were drawn to indicate the beginning and ending positions (often called endpoints) of the speech signal. To apply our recurrence time based method, as our first step, we normalize the original time series into the unit interval [0,1]. Such normalization makes description of the neighborhood size r simple. Next, we partition the speech signals into short data subsets, with each data subset being 1000 points long, and successive data subsets overlapping by 900 points. Since the data subset is quite small, we more or less arbitrarily choose the embedding dimension m to be 4. L, however, is taken to be fairly large, 50, to incorporate the fact that the sampling frequency is fairly

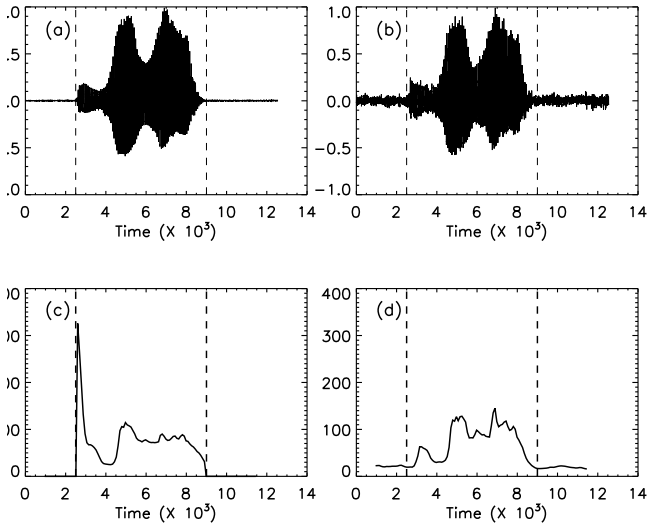

Figure 4: The clean and noisy speech signals for the number "0" (a,b), and their corresponding T_2 curves (c,d). The noise is of babble type.

large (12500 Hz). Figs. 4(c,d) show the $\overline{T}_2(r)$ curve for the signals of Figs. 4(a,b), respectively. Clearly, both the beginning and ending part of the speech signal is accurately identified.

There are two possible ways to evaluate the accuracy of an endpoint detection algorithm. One is to compare the detected results with hand labeled ones. The other is to pass the detected words through a speech recognizer and compare the recognition rates. Here we choose th first option for straightforward comparison. We consider that an endpoint is lost if the error is higher than 75 ms for the beginning and 100 ms for the end. Fig. 5 shows the accuracy of the recurrence time based method (dash-dot curves). For comparison, the results based on the classic energy method is also shown as solid curves. Obviously, the recurrence time based method is much more accurate.

IV.C Epileptic seizure detection/prediction from EEG signals

Epilepsy is one of the most common disorders of the brain. Although epilepsy can be treated effectively in many instances, severe side effects have frequently resulted from constant medication. Even worse, patients may become drug-resistant not long after being treated. To make medication more effective, timely detection of seizure is very important. In the past several decades, considerable efforts have been made to detect/predict seizure through analysis of continuous EEG measurements, and a number of methods have been developed. Here we show that the recurrence time method is excellent in detecting seizures from EEG signals.

We analyzed EEG signals recorded intracranially with approved clinical equipment by the Shands hospital at the University of Florida. Such EEG signals are also called depth EEG, in contrast to scalp EEG. Depth EEG signals are less contaminated by noise or motion artifacts. Typically, a measurement is made with multiple electrodes. Fig. 6 shows a 10-minute duration EEG signals from one electrode. Signals with small amplitudes are considered normal background activities. The clinical equipment used to measure the data has

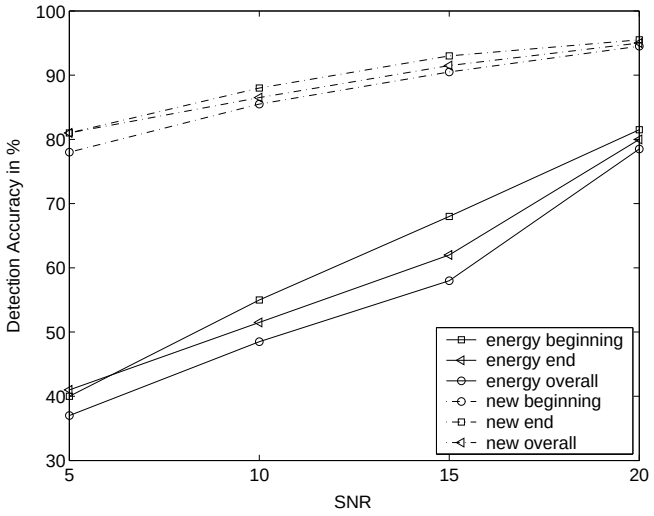

Figure 5: Accuracy of endpoint detection in the babble noise case.

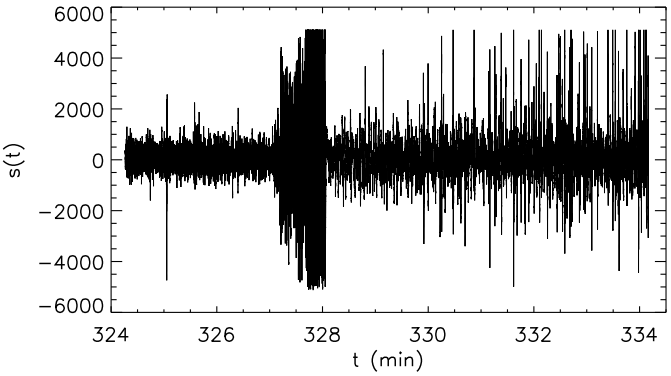

Figure 6: An example of 10-minute long depth EEG signal.

a pre-set, unadjustable maximal amplitude, which is around 5300 μV. This causes clipping of the signals when the signal amplitude is higher than this threshold. This is often the case during seizure episodes, especially for certain electrodes. This is evident in Fig. 6 around minute 328. To a certain extent, this clipping complicates seizure detection, since certain seizure signatures are not captured by the measuring equipment.

We have studied multiple channel EEG signals of a few patients. Each signal is more than five hours long, with a sampling frequency of 200 Hz. Each long signal is partitioned into non-overlapping 10 sec (i.e., 2000 points) window. Within each window, the $\overline{T}_2$ is computed. Figs. 7(a,c,e) show the $\overline{T}_2$ curves for EEG signals of three electrodes. For comparison, the curves based on a popular nonlinear method, the windowed Lyapunov exponent method [10], denoted by STML, are also shown in Figs. 7(b,d,f), respectively. It is clear that the features given by the recurrence time based method is much sharper. Recently, we have found that the nature of EEG signals is such that a number of other nonlinear dynamics or fractal theory based methods are somewhat equivalent to the Lyapunov exponent based method [11, 12, 13, 14, 15, 16]. Therefore, the recurrence time based method is better than

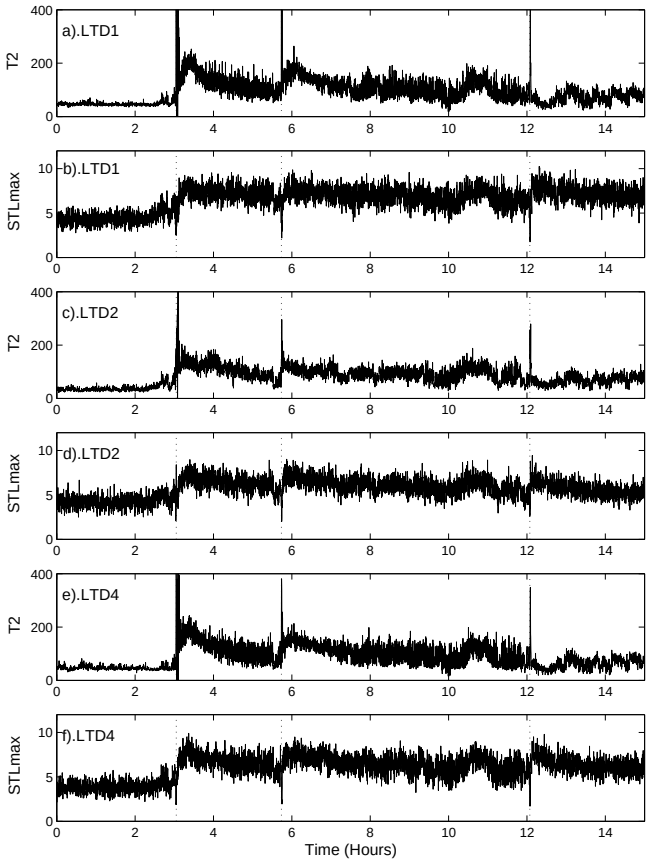

Figure 7: $\overline{T}_2(r)$ and STLmax vs. time curves for the EEG signals measured from three electrodes for a human patient. The parameters are $(m, L, r) = (4, 4, 2^{-4})$ for all the $\overline{T}_2(r)$ curves. The STLmax are extracted from a database maintained by the Epilepsy research Lab at the University of Florida led by Professor Sackellares. Three vertical dotted lines are drawn to indicate seizure occurrence at hour 3.0492, 5.7316, and 12.0690 (i.e., 3:2:57, 5:43:53, and 12:4:8).

any of those methods.

V. Concluding remarks

Entropy and recurrence times are two of the most important complexity measures for both random fields and nonlinear dynamical systems. We have systematically studied the relation between recurrence time distribution and the entire Renyi entropy spectrum. Top illustrate the importance of recurrence time statistics to data analysis, we have considered three important applications, gene identification from genomic DNA sequences, speech endpoint detection, and epileptic seizure detection from EEG data. In all three applications, we have obtained excellent results.

References

[1] D. Ornstein and B. Weiss, *IEEE Trans. Inform. Theory* **48**, 1694, 2002.

[2] G.A. Margulis, *Funkts. Anal. i Prilozh* **3**, 80, 1969.

[3] F. Takens, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, Vol. **898**, edited by D.A. Rand and L.S. Young (Springer-Verlag, Berlin) , p.366, 1981.

[4] A.J. Wyner, *Ann. Appl. Prob.* **9**, 780, 1999.

[5] J.B. Gao, Recurrence Time Statistics for Chaotic Systems and Their Applicaitons. *Phys. Rev. Lett.* **83**, 3178, 1999.

[6] G.H. Choe, *Nonlinearity* **16**, 883, 2003.

[7] P. Grassberger and I. Procaccia, *Phys. Rev. E* **28**, 2591, 1984.

[8] J.B. Gao and H.Q. Cai, On the structures and quantification of recurrence plots. *Phys. Lett. A* **270**, 75-87, 2000.

[9] J.B. Gao, Detecting nonstationarity and state transitions in a time series. *Phys. Rev. E* **63**, 066202, 2001.

[10] L.D. Iasemidis, J.C. Sackellares, H.P. Zaveri, and W.J. Williams, Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures. *Brain Topogr.* **2**, 187-201, 1990.

[11] W. van Drongelen, S. Nayak, D.M. Frim, M.H. Kohrman, V.L. Towle, H.C. Lee, A.B. McGee, M.S. Chico, and K.E. Hecox, Seizure anticipation in pediatric epilepsy: Use of Kolmogorov entropy *Pediatr Neurol* **29**: 207-213, 2003.

[12] K. Lehnertz and C.E. Elger, Spatiotemporal dynamics of the primary epileptogenic area in temporal-lobe epilepsy characterized by neuronal complexity loss. *Electroencephalogr. Clin. Neurophysiol.* **95**, 108-117, 1995.

[13] K. Lehnertz and C.E. Elger, Neuronal complexity loss in temporal lobe epilepsy: effects of carbamazepine on the dynamics of the epileptogenic focus. *Electroencephalogr. Clin. Neurophysiol.* **103**, 376-380, 1997.

[14] J. Martinerie, C. Adam, M. Le Van Quyen, M. Baulac, S. Clemenceau, B. Renault, and F.J. Varela, Epileptic seizures can be anticipated by non-linear analysis. *Nat Med* **4**, 1173, 1998.

[15] R. Aschenbrenner-Scheibe, T. Maiwald, M. Winterhalder, H.U. Voss, J. Timmer, and A. Schulze-Bonhage, How well can epileptic seizures be predicted? An evaluation of a nonlinear method *Brain* **126**, 2616, 2003.

[16] Yinhe Cao, Wen-wen Tung, J.B. Gao, V.A. Protopopescu, and L.M. Hively, Detecting dynamical changes in time series using the permutation entropy. *Phys. Rev. E* **70**, 046217, 2004.