Available online at www.sciencedirect.com
SCIENCE DIRECT®
Physics Letters A 317 (2003) 64–72
PHYSICS LETTERS A
www.elsevier.com/locate/pla

Detection of weak transitions in signal dynamics using recurrence time statistics

J.B. Gao *, Yinhe Cao, Lingyun Gu, J.G. Harris, J.C. Principe

Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA

Received 18 March 2003; received in revised form 29 July 2003; accepted 8 August 2003

Communicated by C.R. Doering

Abstract

Signal detection in noisy and nonstationary environments is very challenging. In this Letter, we study why the two types of recurrence times [Phys. Rev. Lett. 83 (1999) 3178] may be very useful for detecting weak transitions in signal dynamics. We particularly emphasize that the recurrence times of the second type may be more powerful in detecting transitions with very low energy. These features are illustrated by studying a number of speech signals with fricatives and plosives. We have also shown that the recurrence times of the first type, nevertheless, has the distinguished feature of being more robust to the noise level and less sensitive to the parameter change of the algorithm. Since throughout our study, we have not explored any features unique to the speech signals, the results shown here may indicate that these tools may be useful in many different applications.

1. Introduction

Detection of transitional signals in noisy and nonstationary environments is both very important and challenging. Example applications include meteorology, where quantitative description of when and where the weather pattern is going to change is crucial for accurate weather forecasting; and physiology, where timely detection of transitions from a normal to abnormal state may help make medications more effective. Motivated by potential high pay-off in many different areas of science and engineering, this topic has attracted much attention recently, and a number of methods have been proposed using nonlinear dynamical systems theory [1–8]. Most of these methods are based on quantifying certain aspects of the nearest neighbors in phase space. Recently it has been shown [9] that the nearest neighbors in phase space can be broken down into true recurrence points and sojourn points, and two types of recurrence times can be defined. The concepts of sojourn points and two types of recurrence times greatly facilitate the quantification of recurrence plots [1,10], and it is speculated [10] that measures related to these concepts may all be useful to the detection of nonstationarity and state transitions in a time series. Along this line, Marwan et al. [11] have developed an algorithm by quantifying the sojourn points in a recurrence plot to characterize heart rate variability. To deepen our understanding of these new concepts and especially to ease and motivate further ap-

* Corresponding author.
E-mail address: gao@ece.ufl.edu (J.B. Gao).

0375-9601/$ – see front matter
doi:10.1016/j.physleta.2003.08.018

plications of these new concepts to many different fields, in this Letter, we study subtle effects that may be caused by sojourn points and clarify several important differences between the two types of recurrence times in detecting weak transitional signals. We shall illustrate our ideas by studying speech signals with very low energy fricatives and plosives. Such signals can be viewed as naturally occurring weak transitional signals. Accurate detection of such low energy fricatives and plosives in noisy environments is the hardest part of speech endpoint detection. We study both clean and noisy such signals to illustrate that the two different types of recurrence times [9], the recurrence times of the first and second type, have different detection power. Specifically, the recurrence times of the second type is able to detect very weak transitions with high accuracy, both in clean and noisy environments. The recurrence times of the first type, on the other hand, has the distinguished merit of being more robust to the noise level and not sensitive to the parameter change of the algorithm.

The rest of the Letter is organized as follows. In Section 2, we give an overview of the two types of recurrence times, and discuss why they may be useful for weak signal detection and how they may be different. In Section 3, we illustrate our ideas by studying speech signals with weak fricatives and plosives. Section 4 contains the conclusion.

2. Recurrence time based methods for detecting weak transitions

The method first partitions a time series into (overlapping or nonoverlapping) blocks of data sets of short length k, and computes the so-called mean recurrence time of the 1st and the 2nd type, $\bar{T}_1(r)$ and $\bar{T}_2(r)$, for each data subset. For piecewise stationary or transient time series, it has been found [9,10,12] that $\bar{T}_2(r)$ will be different for different blocks of data subsets.

Let us first define the recurrence time of the 2nd type. Suppose we are given a scalar time series $\{x(i),\ i = 1, 2, \ldots\}$. We first construct vectors of the form [13–15]: $X_i = [x(i), x(i + L), \ldots, x(i + (m - 1)L)]$, with m being the embedding dimension and L the delay time. $\{X_i,\ i = 1, 2, \ldots, N\}$ then represents certain trajectory in a m-dimensional space. Next, we arbitrarily choose a reference point

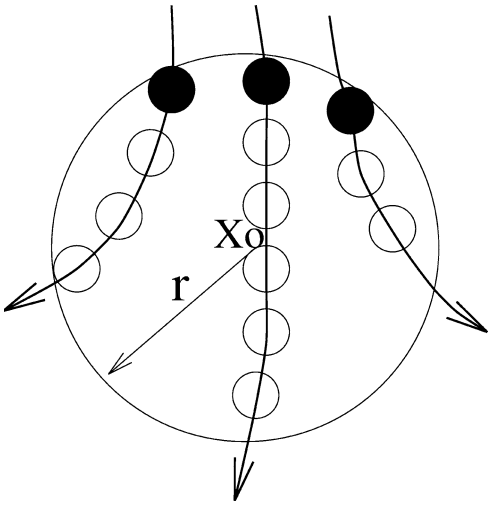

Fig. 1. A schematic showing the recurrence points of the second type (solid circles) and the sojourn points (open circles) in $B_r(X_0)$.

X_0 on the reconstructed trajectory, and consider recurrences to its neighborhood of radius r: $B_r(X_0) = \{X\colon \|X - X_0\| \leqslant r\}$. We then define the recurrence points of the 2nd type 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. 1. The trajectory may stay inside the neighborhood for a while, thus generating a sequence of points, as designated as open circles in Fig. 1. These are called sojourn points [9]. 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. Sojourn points form vertical and horizontal lines, thus square textures, in recurrence plots [10]. The summation of the recurrence points of the second kind and the sojourn points is called the recurrence points of the first kind.

Let us be more precise mathematically. We denote the recurrence points of the 1st type to be $S_1 = \{X_{t_1}, X_{t_2}, \ldots, X_{t_i}, \ldots\}$, and the corresponding Poincaré recurrence time of the 1st type to be $\{T_1(i) = t_{i+1} - t_i,\ i = 1, 2, \ldots\}$. Such times are computed for a protein dynamics model by Manetti et al. [16]. Note the time is computed based on successive returns, not based on the returning points and the reference point. This way, the nearest neighbors of the reference point X_0 are assigned a natural time ordering, and the reference point does not play a central role. We 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, since the number of sojourn points sensitively depends on the embedding parameters, the sampling rate, and the neighborhood size. 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 $S_2 = \{X_{t_1'}, X_{t_2'}, \ldots, X_{t_i'}, \ldots\}$. S_2 then defines a time sequence $\{T_2(i) = t_{i+1}' - t_i',\ i = 1, 2, \ldots\}$. These are called the recurrence times of the 2nd type.

To understand why these two types of recurrence times may be useful for detecting weak transitions, we start from the following scaling laws [9] for the mean recurrence times of the 1st and 2nd type, $\bar{T}_1(r)$ and $\bar{T}_2(r)$, for chaotic attractors,

$$\bar{T}_1(r) \sim r^{-d_1} \tag{1}$$

and

$$\bar{T}_2(r) \sim r^{-(d_1-\alpha)}, \tag{2}$$

where d_1 is the information dimension of the attractor, and α takes on value 0 or 1, depending on whether the sojourn points form very few isolated points inside $B_r(X_0)$, thus contribute dimension 0, or form a smooth curve inside $B_r(X_0)$, thus contribute dimension 1. Since a measured time series may not form an attractor, especially a short segment of the time series may not be able to resolve the whole attractor, hence, d_1 in practice depends on the embedding dimension and the actual dimension of that short segment of the time series. Now suppose during the whole time span of observation, the amplitude or energy has been changing. If the neighborhood size r is the same for all the sub-segments, then the effective r will be different for different sub-segments when their amplitude differs. Then both $\bar{T}_1(r)$ and $\bar{T}_2(r)$ will be different for those sub-segments. This reminds one of the energy method. However, methods based on either $\bar{T}_1(r)$ or $\bar{T}_2(r)$ will be more powerful than the energy method in detecting transitions, since recurrence time based methods work in phase space, with the dimension of the phase space usually larger than 2. Even more important, even if the amplitude remains the same for the whole time series, so long as other features of the signals such as period, dimension, or complexity (as characterized, say, by Lyapunov exponents or entropy) change with time, $\bar{T}_1(r)$ and $\bar{T}_2(r)$ will also be different for different blocks of the sub dataset. These are the fundamental reasons that recurrence time based methods may be very powerful in detecting transitions in signal dynamics.

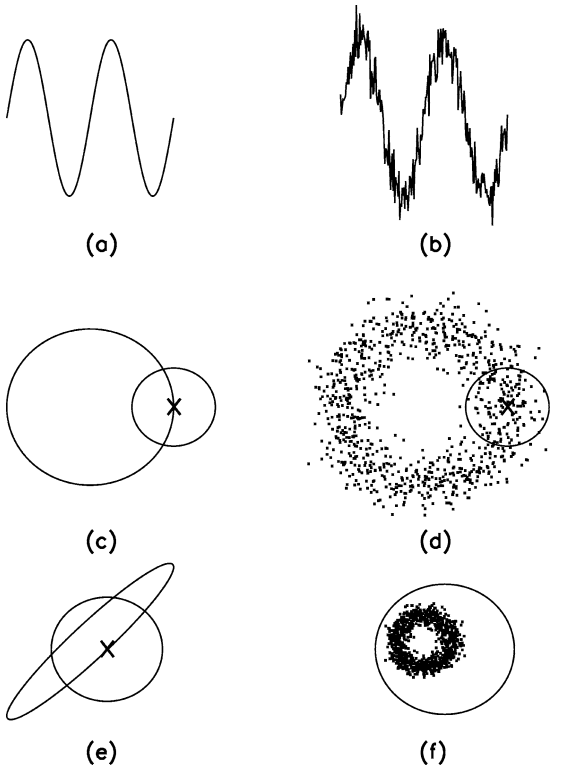

Fig. 2. (a), (c) Sine wave (unit amplitude) and its embedding (large circle). A small circle is also drawn to represent a neighborhood of the reference point denoted by the cross; (b), (d) same as (a), (c) except now the sine wave signal is corrupted by Gaussian white noise with mean 0 and variance 0.3; (e) a schematic showing that a neighborhood of some reference point intersects disjoint regions of the phase space; and (f) a neighborhood of some reference point contains the entire phase space.

Next we discuss the differences between $\bar{T}_1(r)$ and $\bar{T}_2(r)$. We first discuss periodic signals both without and with noise, as schematized in Fig. 2(a) and (b). Fig. 2(c), (d) show the corresponding embeddings. We notice that so long as the size of the neighborhood is not too large, $T_2(i)$ accurately estimates the period of the motion. In fact, even if a signal is noise-corrupted, the period can still be accurately estimated, as shown

by Fig. 2(b) and (d). Due to sojourn points, however, $\bar{T}_1(r)$ is different from the period of the motion.

At this point, it is worth pointing out that although the embedding dimension m and delay time L have to be chosen carefully for most chaos-based methods, this is often not the case for computing $\bar{T}_2(r)$. This can be readily appreciated by observing Fig. 2(c), (d). If we choose other embedding parameters, the limit cycle gets deformed or even twisted. However, the period of the motion remains the same and can be readily estimated by $\bar{T}_2(r)$ for most parameter combinations of m and L. On the other hand, if a reconstructed attractor has an ill-defined structure due to poor embedding, then computations of fractal dimension, Lyapunov exponents, Kolmogrov–Sinai entropy, etc., may become unfeasible.

The above discussion is valid only when the signal amplitude is large compared to the size of the neighborhood. When this is not the case, then the neighborhood of some reference point may contain disjoint subsets of the phase space, as schematized in Fig. 2(e). When this happens, then $\bar{T}_2(r)$ is no longer the period, but a fraction of the period (half of the period for Fig. 2(e) for some reference points).

At this point, it is appropriate for us to make a few comments on the interesting notion of coping nonstationarity by over-embedding suggested by Hegger et al. [17] (see also the in-depth discussion on this issue by Zbilut et al. [18]). Over-embedding is often advantageous, especially when the dataset considered is large. This can be appreciated by considering a noisy limit cycle (Fig. 2(d)). It occupies a larger domain in a space with a higher dimension. Since the noise level at any specific direction remains the same, the ratio between the size of the limit cycle and the standard deviation of the noise thus becomes larger in a higher-dimensional space, resulting in a better defined limit cycle. However, when the dataset available is small, we would only recommend over-embedding with a not too high embedding dimension m, for two primary reasons: (i) when the embedding dimension is too large, then one may not be able to find sufficient number of recurrences; (ii) when the dataset is small, its "effective" dimension is often also small. For example, a few cycles of data from the chaotic Lorenz attractor would appear to be more like periodic than chaotic. Hence, a very high dimensional embedding space may not be necessary.

Let us now discuss the major difference between $\bar{T}_1(r)$ and $\bar{T}_2(r)$ in detecting weak transitions. In order to fully explore the amplitude difference along the entire time series, in practice it is often advantageous to keep the neighborhood size r fixed for all the sub-datasets. Hence, when the energy of some sub-segment is low, the ratio between r and the amplitude of that sub-segment is large. In other words, r is big compared to some sub-segments of the data where the energy is very low. In the extreme case, the neighborhood of any reference point contains the entire phase space spanned by the sub-dataset. When this happens, then there are no recurrence points of the second type. For simplicity, we take $\bar{T}_2(r) = 0$ (which really means $\bar{T}_2(r) = \infty$, i.e., no returns). On the other hand, $\bar{T}_1(r)$ equals to 1 unit of the sampling time, since all the points in the phase space are sojourn points. Now suppose there are some sub-datasets, where the majority of the points are sojourn points (due to low energy). This amounts to saying that for each reference point belonging to that sub-dataset, the number of sojourn points is on the order of k, where k is the length of the sub-dataset. Hence, the summation of $T_1(r)$ from all the reference points has a component on the order of k^2. We further assume there are only a few true recurrence points. Hence, $T_2(r)$ is on the order of k, which is a finite but large number, thus vastly different from 0, indicating transitions. This $T_2(r)$ contributes another component to the summation of $T_1(r)$ from all the reference points, which is on the order of k. Hence, $\bar{T}_1(r) \approx (k^2 + k)/k^2 \sim 1$, which is not very different from the data subset with small enough energy to be completely contained in the neighborhood of some reference point. Hence, $\bar{T}_1(r)$ is not able to detect such weak transitions. These arguments signify that $\bar{T}_2(r)$ is more powerful in detecting very weak transitions in signal dynamics than $\bar{T}_1(r)$. Below, we shall illustrate this feature by studying speech signals.

The above argument implies that the neighborhood size r can act as a filter. Observing again Fig. 2(b) and (d), we see that if we reduce r till passing certain threshold which is proportional to the standard deviation of the noise, then we will no longer be able to estimate the period. Put in another way, if we make r reasonably large, then noise can be easily and completely filtered out.

3. Speech endpoint detection using recurrence time statistics

In this study, a number of signals for isolated words, such as English digits, obtained from the TI46 database, have been studied. Below, we shall present typical results. Since our purpose is to study very weak transitions, we shall use numbers "0" and "7" for illustration, as they contain weak fricatives. The "0" used here was pronounced by a female speaker, while "7" by a male speaker. Both were sampled with a frequency of 12500 Hz. A total of 12544 and 13056 samples were collected for "0" and "7", respectively. Three different types of noise, white, pink, and babble, were studied. 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. In nonlinear time series analysis, it is often called surrogate data. Sometimes it may refer to background conversation. In most robust automatic speech recognitions, babble noise is considered the hardest to deal with. These different types of noise were obtained from the Signal Processing Information Base (SPIB) collected by the Rice University.

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 large (12500 Hz). Fig. 3(a) and (b) show the clean and noisy signals for the number "0", respectively. In both figures, two vertical dashed lines were drawn to indicate the starting and ending positions (often called endpoints) of the speech signal. In Fig. 3(b), the type of noise is babble. Its strength is 15 dB, where dB is defined as 10 times the logarithm of the ratio between signal energy and noise variance. In automatic speech recognitions, two types of SNR have been used, one is global, where the entire signal is used to define the ratio between signal energy and noise variance. The other is local, where the entire signal is partitioned into many small segments, then the ratio between signal energy and noise variance is found for each segment. We adopt global SNR here. This way, the effective noise level for endpoints is much higher than that defined by the local SNR method as 15 dB.

Fig. 3(c), (e) and (d), (f) show the $\bar{T}_2(r)$ and $\bar{T}_1(r)$ curves for the clean and noisy signals, respectively. The neighborhood size r is chosen to be $r = 2^{-9/2}$, in all these computations. We observe several distinct features from these $\bar{T}_2(r)$ and $\bar{T}_1(r)$ curves: (i) $\bar{T}_2(r)$ picks up very well the starting and ending positions of the speech signal, for both clean or noisy signals. We especially emphasize that due to change of the number of true recurrence points from zero to a few, $\bar{T}_2(r)$ curve has a huge jump corresponding to the starting position of the speech for the clean signal; (ii) $\bar{T}_1(r)$ also indicates the starting and ending positions of the speech signal to some degree, both for the clean and noisy signals. However, the indication is less accurate than that using $\bar{T}_2(r)$; (iii) The shape of $\bar{T}_2(r)$ curve changes a lot when noise is added. The shape of $\bar{T}_1(r)$ curve, however, is largely independent of noise. This is a salient feature of the $\bar{T}_1(r)$ statistic.

Next, we consider the number "7". The clean and noisy signals are shown in Fig. 4(a), (b). Again, the noise is of babble type, and the SNR is 15 dB. We notice that the signal waveform is very different from that of the number "0". We choose $r = 2^{-9/2}$ to compute the recurrence times. They are shown in Fig. 4(c), (e) and (d), (f), for the clean and noisy signals, respectively. We also observe a few interesting features here: (i) the indication of the starting and ending positions of the speech signal from $\bar{T}_2(r)$ is quite accurate, both for the clean and the noisy signals; (ii) $\bar{T}_1(r)$ only indicates the major part of the speech signal, not the endpoints where the energy of the signal is very low. Hence, this example illustrates the major difference between $\bar{T}_1(r)$ and $\bar{T}_2(r)$ more vividly. Nevertheless, $\bar{T}_1(r)$ again shows the remarkable feature of being very robust to the noise level.

We should emphasize that when the signal is corrupted by babble noise, $\bar{T}_2(r)$ also shows considerable activity in the time intervals where there is only noise. This has both advantages and disadvantages: on the one hand, this complicates automatic speech segmentation problem, but on the other hand, this indicates $\bar{T}_2(r)$ has the great power of detecting all types of weak changes in a signal.

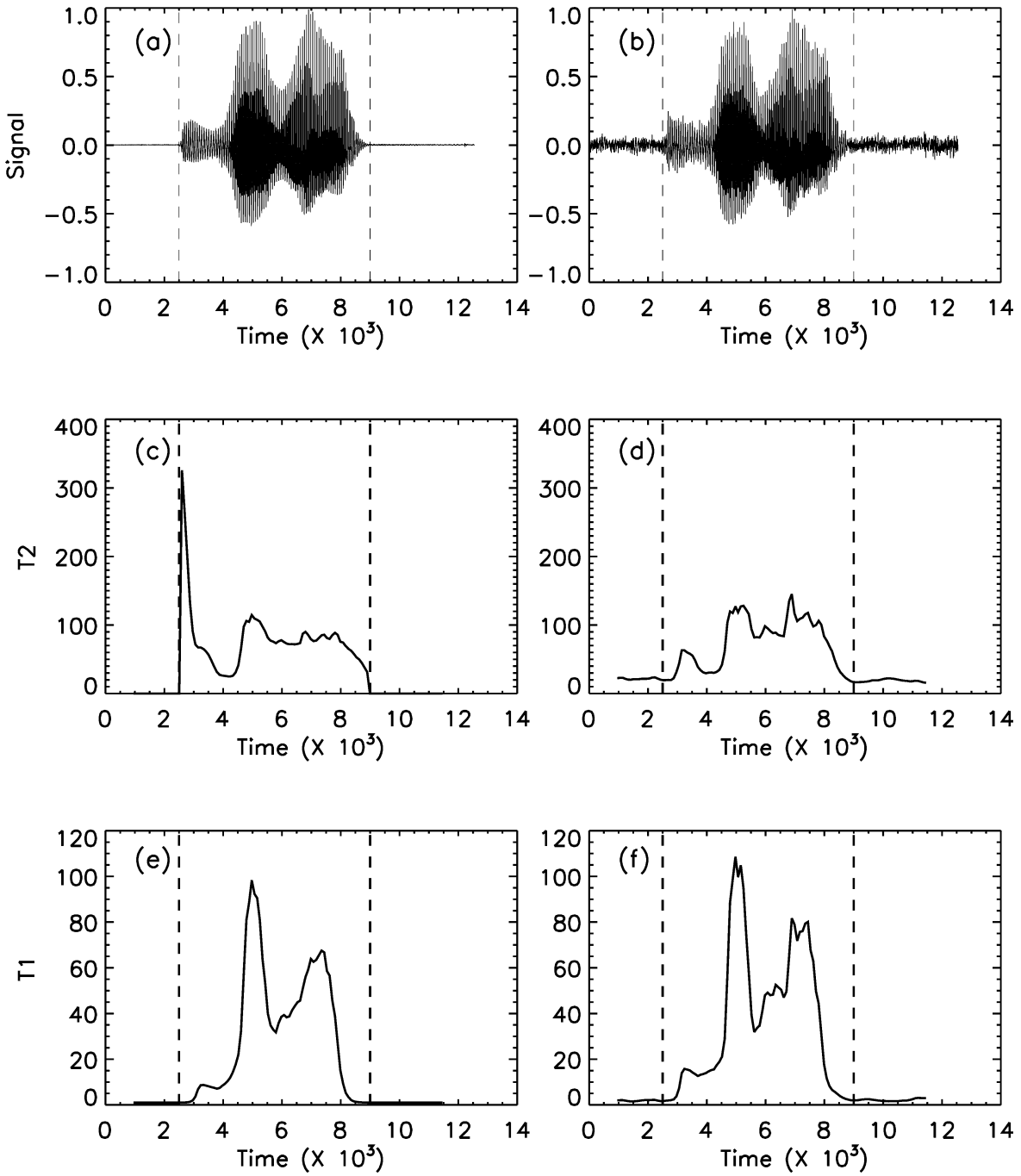

Fig. 3. (a), (b) The clean and noisy signals for the number "0". The noise is of babble type; (c), (e) and (d), (f) $\bar{T}_2(r)$ and $\bar{T}_1(r)$ curves for the clean and noisy signals, respectively. The parameters are $m = 4$, $L = 50$, and $r = 2^{-9/2}$. Vertical dashed lines are drawn to indicate the positions of the starting and ending positions of the speech signal.

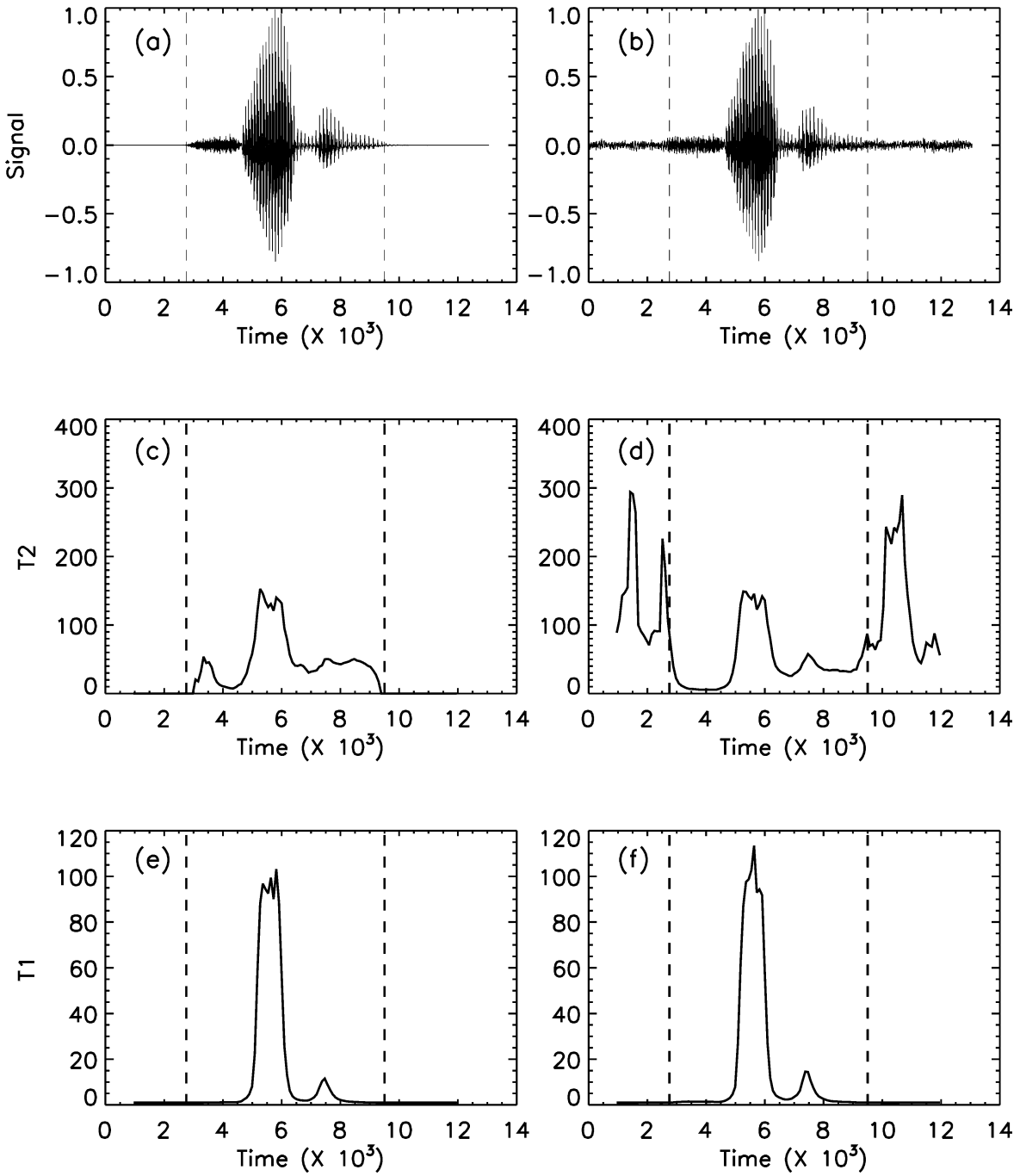

Fig. 4. Same as Fig. 3 except now the signal is for the number “7”.

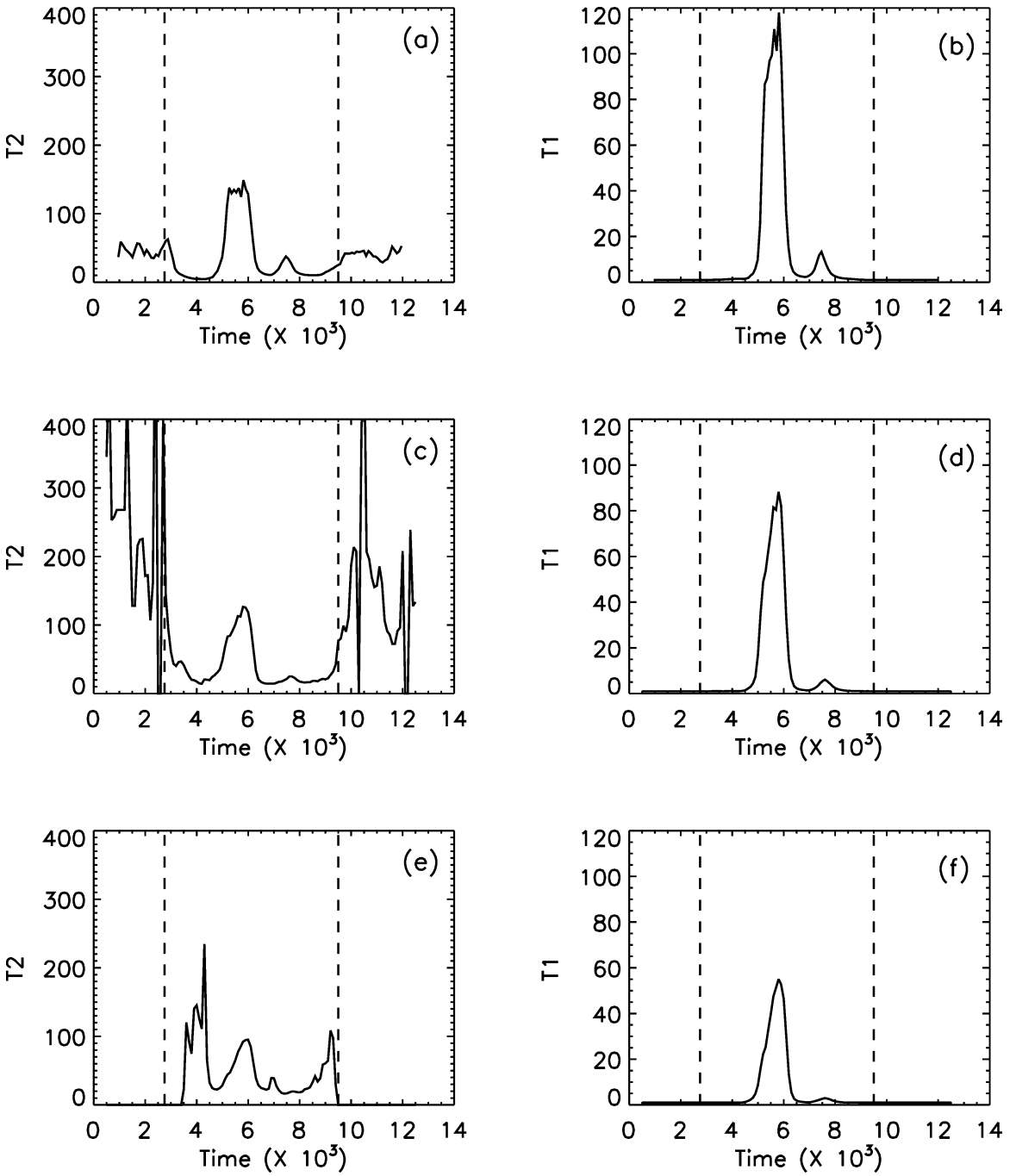

Fig. 5. $\bar{T}_2(r)$ and $\bar{T}_1(r)$ curves for the noisy signal of the number "7". The noise is now white instead of babble. The parameters are $m = 4$, $L = 50$, and $r = 2^{-9/2}$ for (a), (b), $r = 2^{-8/2}$ for (c), (d), and $r = 2^{-7/2}$ for (e), (f).

For the purpose of automatic speech recognition, there is a way to disable $\bar{T}_2(r)$ from picking up all the activities. This can be easily achieved by using larger r when the noise level is below certain threshold. Fig. 5 shows a few $\bar{T}_2(r)$ and $\bar{T}_1(r)$ curves with $r = 2^{-9/2}$ (Fig. 5(a), (b)), $r = 2^{-4}$ (Fig. 5(c), (d)), and $r = 2^{-7/2}$ (Fig. 5(e), (f)). Here, we have used white noise (also 15 dB) instead of babble noise, to illustrate an interesting difference between white noise and babble noise. Comparing between Fig. 4(d) and Fig. 5(a), we find that in the time intervals where there is only white noise, $\bar{T}_2(r)$ curves are fairly flat. This contrasts with the wide variations shown in Fig. 4(d). The reason that Fig. 5(c) also shows considerable variations in those time intervals is quite different. That is simply because of poor statistics based on very few recurrences to a neighborhood. In fact, when the neighborhood size is increased from $r = 2^{-4}$ to $r = 2^{-7/2}$, then all the noise is filtered out, and no recurrences are recorded. Accompanying with this filtering there is a small penalty as shown by Fig. 5(e): the starting position of the signal is slightly delayed than that shown by Fig. 4(c).

A note on automatic speech endpoint detection using $\bar{T}_2(r)$ is in order. The method computes $\bar{T}_2(r)$ twice for two different r, one large and one small. The large r case would give us a $\bar{T}_2(r)$ curve with clean boundaries, such as shown in Fig. 5(e). Those boundaries may, however, be not very accurate. The $\bar{T}_2(r)$ curve for the small r, on the other hand, would be both messy and indicate the boundaries accurately, as shown in Figs. 4(d) and 5(a). To accurately locate the true boundaries, we can extend from the clean $\bar{T}_2(r)$ curve for large r till we reach the location of $\bar{T}_2(r)$ for small r where first abrupt change occurs. This way, implementation will be simple and detection will be accurate.

Before leaving this section, we note that noise level up to SNR of 5 dB has been studied and similar performance results obtained. Noise level of 5 dB is considered very noisy in speech recognitions.

4. Conclusions

In this Letter, we have both qualitatively and quantitatively argued why the two types of recurrence times may be very useful for detecting weak transitions in signal dynamics. We have particularly emphasized that the recurrence times of the second type may be more powerful in detecting transitions with very low energy. These features are illustrated by studying a number of speech signals with fricatives and plosives. We have also shown that the recurrence times of the first type, nevertheless, has the distinguished feature of being more robust to the noise level and parameter change of the algorithm. Hence, when the signatures of the signals being detected are very strong, the recurrence times of the first type may be easier to use than the recurrence times of the second type. Since throughout our study, we have not tried to explore any features unique to the speech signals, the results shown here may indicate that these novel tools may be useful in many different applications, such as epileptic seizure detection from EEG signals, sleep stage detection, network anomaly detection, and earthquake prediction.

References

[1] J.P. Eckmann, S.O. Kamphorst, D. Ruelle, Europhys. Lett. 4 (1987) 973.
[2] L.L. Trulla, A. Giuliant, J.P. Zbilut, C.L. Webber, Phys. Lett. A 223 (1996) 255.
[3] A. Provenzale, L.A. Smith, R. Vio, G. Murante, Physica D 58 (1992) 31.
[4] D.J. Yu, W.P. Lu, R.G. Harrison, Phys. Lett. A 250 (1998) 323.
[5] R. Manuca, R. Savit, Physica D 99 (1996) 134.
[6] M.B. Kennel, Phys. Rev. E 56 (1997) 316.
[7] H. Kantz, Phys. Rev. E 49 (1994) 5091.
[8] T. Schreiber, Phys. Rev. Lett. 78 (1997) 843.
[9] J.B. Gao, Phys. Rev. Lett. 83 (1999) 3178.
[10] J.B. Gao, H.Q. Cai, Phys. Lett. A 270 (2000) 75.
[11] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths, Phys. Rev. E 66 (2002) 026702.
[12] J.B. Gao, Phys. Rev. E 63 (2001) 066202.
[13] N.H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw, Phys. Rev. Lett. 45 (1980) 712.
[14] F. Takens, Detecting strange attractors in turbulence, in: D.A. Rand, L.S. Young (Eds.), Dynamical Systems and Turbulence, in: Lecture Notes in Mathematics, Vol. 898, Springer-Verlag, Berlin, 1981, p. 366.
[15] T. Sauer, J.A. Yorke, M. Casdagli, J. Stat. Phys. 65 (1991) 579.
[16] C. Manetti, A. Giuliani, M.-A. Ceruso, C.L. Webber, J.P. Zbilut, Phys. Lett. A 281 (2001) 317.
[17] R. Hegger, H. Kantz, L. Matassini, T. Schreiber, Phys. Rev. Lett. 84 (2000) 4092.
[18] J.P. Zbilut, N. Thomasson, C.L. Webber, Medical Engrg. Phys. 24 (2002) 53.