Time-Frequency Representations as Phase Space Reconstruction in Recurrence Symbolic Analysis

Mariia Fedotenkova, Peter Beim Graben, Jamie Sleigh, Axel Hutt

Time-Frequency Representations as Phase Space Reconstruction in Recurrence Symbolic Analysis

Mariia Fedotenkova[1,2,3], Peter beim Graben[4], Jamie Sleigh[5] and Axel Hutt[1,2,3]

[1] NEUROSYS team, INRIA, Villers-lès-Nancy, F-54600, France
[2] UMR n° 7503, CNRS, Loria, Vanduvre-lès-Nancy, F-54500, France
[3] Université de Lorraine, Villers-lès-Nancy, F-54600, France
[4] Bernstein Center for Computational Neuroscience, Berlin, Germany
[5] Waikato Clinical School of the University of Auckland, New Zealand
maria.fedotenkova@gmail.com

Abstract. Recurrence structures in univariate time series are challenging to detect. We propose a combination of symbolic and recurrence analysis in order to identify recurrence domains in the signal. This method allows to obtain a symbolic representation of the data. Recurrence analysis produces valid results for multidimensional data, however, in the case of univariate time series one should perform phase space reconstruction first. In this paper, we propose a new method of phase space reconstruction based on signal's time-frequency representation and compare it to delay embedding method. We argue that the proposed method outperforms delay embedding reconstruction in the case of oscillatory signals. We also propose to use recurrence complexity as a quantitative feature of a signal. We evaluate our method on synthetic data and show its application to experimental EEG signals.

Keywords: recurrence analysis, symbolic dynamics, time-frequency representation, Lempel-Ziv complexity, EEG

1 Introduction

Recurrent temporal dynamics is a phenomenon frequently observed in time series measured in biological systems. For instance, bird songs exhibit certain temporal structures, that recur in time [29]. Other examples are returning epileptic seizures [2], recurrent cognitive states in neural language processing [12] and in early auditory neural processing [14]. All these latter phenomena are observed in electroencephalographic data (EEG). To detect such temporal recurrent structures, typically one applies recurrence analysis [5, 23] based on Poincaré's theorem [25]. This approach allows to detect recurrence structures in multivariate time series. To retrieve recurrence structures from univariate time, several methods have been suggested, such as delay embedding techniques.

However, most existing methods do not take into account specifically the oscillatory nature of the signals as observed in biological systems. To this end,

we propose a technique to embed the univariate time series in a multidimensional space to better consider oscillatory activity. The approach is based on the signals time-frequency representation. In a previous work we have sketched this approach [28] already but without discussing its performance subject to different time-frequency representations. The present work shows this detailed discussion and suggests a new method to classify signals according to their recurrence complexity. Applications to artificial data permits to evaluate the method and compare it to results gained from the conventional delay embedding technique. Final applications to experimental EEG data indicates the method's future application.

2 Analysis Methods and Data

2.1 Recurrence Symbolic Analysis

Recurrence is a fundamental property of nonlinear dynamical systems, which was first formulated by Poincaré in [25]. It was further illustrated in recurrence plot (RP) technique proposed by Eckmann et al. [5]. This relatively simple method allows to visualize multidimensional trajectories on a two-dimensional plane. The RP can be obtained by plotting the recurrence matrix:

$$R_{ij} = \Theta\left(\varepsilon - ||\boldsymbol{x}_i - \boldsymbol{x}_j||\right), \qquad i, j = 1, 2, \ldots, N \ , \tag{1}$$

where $\boldsymbol{x}_i \in \mathbb{R}^d$ is the state of the complex system in the phase space of dimension d at a time instance i; $||\cdot||$ denotes a metric, Θ is the Heaviside step function, and ε is a threshold distance.

It can be seen from (1), that if two points in the phase space are relatively close, the corresponding element of the recurrence matrix $R_{ij} = 1$, which would be represented by a black dot on the RP.

Instead of analyzing RPs point-wise we concentrate our attention on recurrence domains, labeling each domain with a symbol, thus obtaining recurrence plots of symbolic dynamics. The RP from symbols were successfully used in several studies (see, for instance, [6], [18], and [4]). Here, we use recurrence symbolic analysis (RSA) proposed in [11], this technique allows to obtain symbolic representations of the signal from the RP, the latter being interpreted as a rewriting grammar. The RP contains rewriting rules, which substitute large time indices with smaller ones when two states, occurring at these times, are recurrent. These rules can be summarized as follows:

$$\begin{array}{ll} i \to j & \text{if } i > j \text{ and } R_{ij} = 1 \\ \left.\begin{array}{l} i \to k \\ j \to k \end{array}\right\} & \text{if } i > j > k \text{ and } R_{ij} = 1,\ R_{ik} = 1 \ . \end{array} \tag{2}$$

More detailed description of the method and examples can be found in [11, 12].

By examining (1) one can see that the resulting recurrence matrix and, thus, symbolic sequence strongly depends on distance threshold parameter ε. Several

techniques for optimal ε estimation exist [22], most of which are empirical. RSA aims to obtain an optimal value of ε from the data.

Our approach to optimal ε estimate is based on the principle of maximal entropy, which implies that the system spends equal amount of time in each recurrence domain [11]. Thus, optimal ε is chosen such that the entropy of symbolic sequence is maximal. The entropy of the symbolic sequence is given by:

$$H(\varepsilon) = -\sum_{k+1}^{M(\varepsilon)} p_k \log p_k \ , \tag{3}$$

where p_k is the relative frequency of the symbol k, $M(\varepsilon)$ is the cardinality of the alphabet. Here, we use $h(\varepsilon) = H(\varepsilon)/M(\varepsilon)$ in order to compensate for the influence of the alphabet size. This entropy ratio allows us to determine the optimal threshold distance as follows:

$$\varepsilon^* = \arg\max_{\varepsilon} h(\varepsilon) \ . \tag{4}$$

2.2 Phase Space Reconstruction

A dynamical system is defined by an evolution law in a phase space. This space is d-dimensional, where each dimension is a certain property of a system (for instance, position, and velocity). Each point of the phase space correspond to a possible state of the system. An evolution law, which is normally given by a set of differential equations, defines system's dynamics, shown as a trajectory in a phase space.

In certain cases only discrete measurements of single observable are available, in this situation a phase space should be reconstructed according to Takens's theorem [27], which states that phase space presented with a d-dimensional manifold can be mapped into $2d+1$-dimensional Euclidean space preserving dynamics of the system. Several method of phase space reconstruction exist: method of delays (MOD) [27], numerical derivatives [24] and others (see for instance [17]).

In this paper we propose a new method of phase space reconstruction based on the time-frequency representation of the signal. A time-frequency representation (TFR) is a distribution of the power of the signal over time and frequency. Here, the power in each frequency band contributes to a dimension of the reconstructed phase space. This approach is well-adapted for non-stationary and, especially, for oscillatory data, allowing better detection of oscillatory components rather than creating RPs point-wise from the signal. In this article we compare performance of RSA with different reconstruction methods, MOD and three different TFRs: spectrogram, reassigned spectrogram and continuous wavelet transform.

Method of Delays. Assume, we have a time series which represents scalar measurements of a system's observable in discrete time:

$$x_n = x(n\Delta t), \quad n = 1, \ldots, N \ , \tag{5}$$

where Δt is measurement sampling time. Then reconstructed phase space is given by:

$$\boldsymbol{s}_n = \left[x_n,\, x_{n+\tau},\, x_{n+2\tau},\, \ldots,\, x_{n+(m-1)\tau}\right], \quad n = 1,\, \ldots,\, N-(m-1)\tau \ , \qquad (6)$$

where m is the embedding dimension and τ is the time delay. These parameters play an important role in correct reconstruction and should be estimated correctly. Optimal time delay τ should be chosen such that delay vectors from (6) are sufficiently independent.

Several approaches to estimate this parameter exist. Amongst them, one well-established technique is based on average mutual information [7, 20], which we have chosen to estimate the proper value of the time delay τ. Moreover, the main attribute of appropriately chosen dimension m is that the original d-dimensional manifold will be embedded into an m-dimensional space without ambiguity, i.e., self-crossing and intersections. We apply the method of false nearest neighbors [16, 15], which permits to estimate the minimal embedding dimension.

Time-Frequency Representation. Time-frequency representation of a signal shows the signal's energy distribution in time and frequency. In this paper we analyze three different types of TFR: spectrogram, reassigned spectrogram and continuous wavelet transform.

The spectrogram $S^h(t,\omega)$ of a signal $x(t)$ is the square magnitude of its short-time Fourier transform (STFT)

$$X^h(t,\omega) = \int_{-\infty}^{+\infty} x(\tau) h^*(t-\tau) \mathrm{e}^{-i\omega\tau}\, \mathrm{d}\tau \ , \qquad (7)$$

where $h(t)$ is a smoothing window and $*$ denotes the complex conjugate, i.e., $S^h(t,\omega) = \left|X^h(t,\omega)\right|^2$.

Spectrogram reassignment [3] is a method of decreasing the spreading of the conventional spectrogram by moving its value $S(t,\omega)$ from original location where it has been computed to the centroids of the energy distribution $\hat{t}, \hat{\omega}$ (one can think of such centroids as "centers of gravity" of the distribution). More detailed information about the implementation of the method can be found in [10].

The continuous wavelet transform (CWT) [1] is obtained by convolving the signal with a set of functions $\psi_{ab}(t)$ obtained by translation and dilation of a mother wavelet function $\psi_0(t)$:

$$T_\psi(b,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t) \psi_0^* \left(\frac{t-b}{a}\right) \mathrm{d}t \ , \qquad (8)$$

Then the squared magnitude of the CWT is called scalogram and is given by:

$$W_\psi(b,a) = \left|T_\psi(b,a)\right|^2 \ . \qquad (9)$$

In practice, the scale a can be mapped to a pseudo-frequency f and the dilation b represents a time instance and hence the time-frequency distribution is given by $W_\psi(t, f)$.

Reassigned and conventional spectrograms and scalogram were obtained using toolboxes for MATLAB®[6]. The CWT was computed using analytical Morlet wavelet, and a Hamming window with 80% overlap was chosen for the spectrogram. In all the methods the window length and scales locations were chosen such as to achieve a frequency resolution of 0.3Hz for synthetic data and 1Hz for experimental data.

2.3 Complexity Measure

To quantitatively measure the obtained symbolic sequences we propose to measure its complexity. We present here three different complexity measures. These are the cardinality of the sequence and the number of distinct words obtained from the sequence [13], where a word is a unique group of the same symbols. In addition, we compute the well-known Lempel-Ziv (LZ) complexity [19], which is related to the number of distinct substrings and the rate of their occurrence along the symbolic sequence. All of the complexity measures have in common the notion of complexity, that is the number of distinct elements required to encode the symbolic string. The more complex the sequence is the more of such elements are needed to present it without redundancy.

To demonstrate these measures we generated 100 artificial signals of two kinds (see below) with random initial conditions and random noise.

2.4 Synthetic data

Transient Oscillations. The signal is a linear superposition of three signals, which exhibit sequences of noisy transient oscillations at a specific frequency [28]. These frequencies are 1.0Hz, 2.25Hz and 6.3Hz, cf. Fig.1. The sampling frequency is 50Hz and the signal has a duration of 70s. Figure 1 shows the three different transient oscillations whose sum represents the signal under study.

Lorenz System. The solution of the chaotic Lorenz system [21, 11] exhibits two wings which are approached in a unpredictable sequence. These wings represent metastable signal states. Figure 2 shows the time series of the z-component of the model.

2.5 Experimental data

We investigate electroencephalographic data (EEG) obtained during surgery under general anaesthesia [26]. The EEG data under investigation has been captured at frontal electrodes 2 minutes before (pre-incision phase) and 2 minutes

[6] They are openly available at https://github.com/mfedoten/reasspectro and https://github.com/mfedoten/wavelets

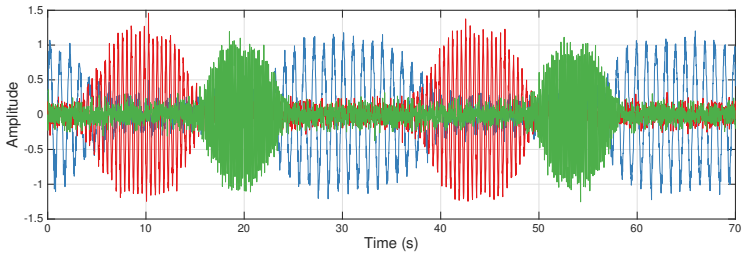

Fig. 1. Three signals whose superposition yield transient oscillations. Different frequencies of each oscillating component is shown in different color: blue color corresponds to 1Hz, red – 2.25Hz and green – 6.3Hz. Here no experimental noise is added (high SNR).

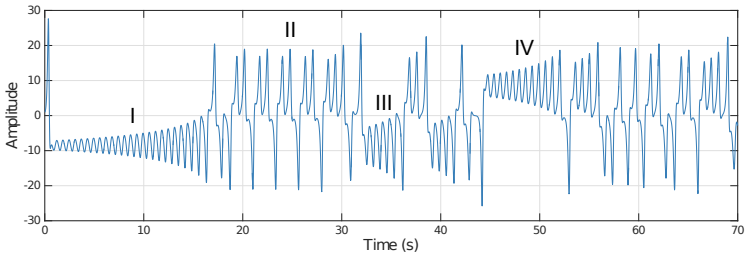

Fig. 2. Solution of the Lorenz system without noise.

after (post-incision phase) skin incision and lasts 30 seconds. The raw signal was digitized at a rate of 128 Hz and digitally band-pass filtered between 1 Hz and 41 Hz using a 9th order Butterworth filter. The question in the corresponding previous study [26] was whether it is possible to distinguish the pre-incision from post-incision phase just on the basis of the captured EEG time series.

3 Results

3.1 Synthetic data

Time-Frequency Embedding. To illustrate the method, Fig. 3 shows three different time-frequency representations of the transient oscillation signal. Spectrogram and reassignment spectrogram yield time-frequency intervals of high power at very good accordance to the underlying dynamics, cf. Sect. 2.4. In contrast, wavelet analysis smears out upper frequencies as a consequence of their intrinsic normalization of power. The symbolic sequence and the corresponding recurrence plot (right-hand side of the panel) derived from the spectrogram fits perfectly to the underlying dynamics. They exhibit three different symbols in the symbolic sequence color-coded in blue, red and orange in Fig. 3(A) and alternate in very good accordance to the three different transient oscillations. They are also visible as three rectangles of different size in the symbolic recurrence plot.

Conversely, the reassignment spectrogram and wavelet representation yield two recurrent signal features only not reflecting the underlying dynamics.

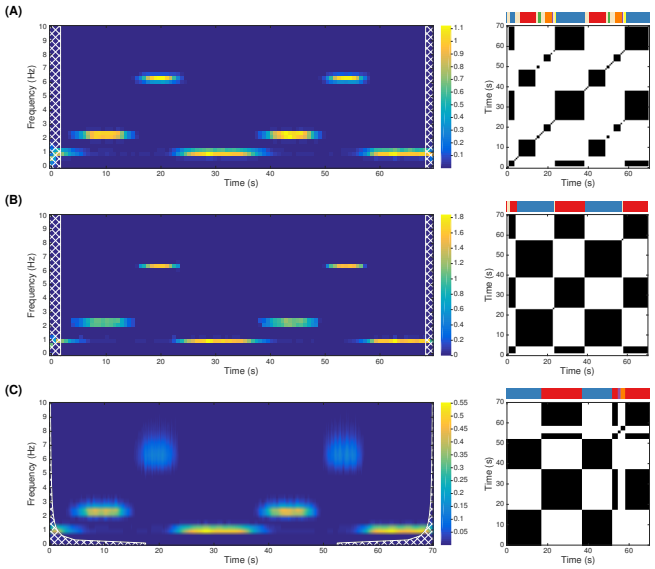

Fig. 3. Results for the transient oscillation signal. (A) spectrogram; (B) reassigned spectrogram; (C) continuous wavelet transform. On each subfigure, left: time-frequency representation, right: RPs with corresponding symbolic sequences above them.

Typically experimental neurophysiological signals exhibit a less regular temporal structure than given in the transient oscillations example. Solutions of the Lorenz system exhibit chaotic behavior, that is rather unregular and exhibits metastable oscillatory states. Since experimental EEG may exhibit chaotic behavior [9, 8], the Lorenz signal is closer to neurophysiological data. Figure 4 shows TFR of the Lorenz signal without experimental noise. For the three time-frequency representations, one can well identify visually the four signal states I to IV marked in Fig.2. The color-coded symbolic sequences extracted from the spectrogram (seen in Fig.4(A) on the right-hand side) identify correctly the time windows of the signal states I to IV. The states I, II and IV are well captured, whereas the short state III is not well identified. The result for reassigned spectrogram is similar, whereas wavelet results are much worse.

Delay embedding. To illustrate the power of the method proposed, we compare our results to recurrence analysis results utilizing delay embedding, cf. Sect. 2.2. We consider the transient oscillations and the Lorenz signal, compute the optimal delay embedding parameters and apply the recurrence analysis technique to gain the symbolic sequences and the recurrence plots. Figure 5 reveals

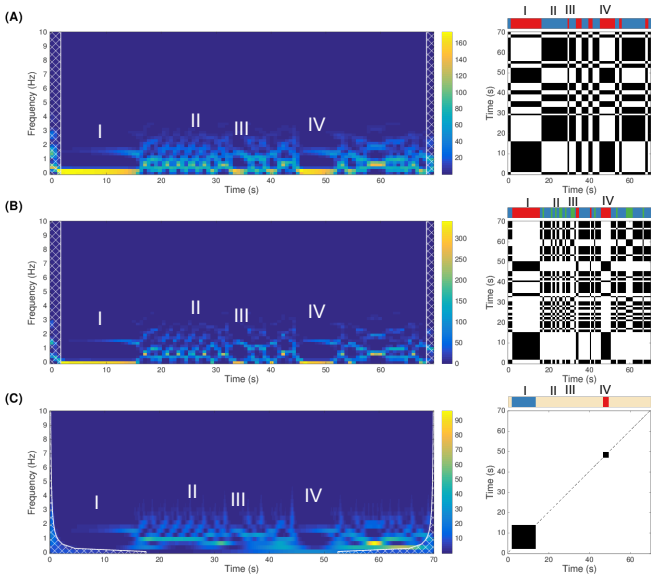

Fig. 4. Results for the Lorenz system. (A) spectrogram; (B) reassigned spectrogram; (C) continuous wavelet transform. On each subfigure, left: time-frequency representation, right: RPs with corresponding symbolic sequences above them.

that the delay embedding leads to a bad detection of the recurrence domains in the transient oscillations compared to the time-frequency embedding. In the Lorenz signal all states I-IV are captured in the symbolic sequence and visible in the recurrence plot, however the detection is much worse than with time-frequency embedding, cf. Fig. 4.

Complexity Measure. To quantify the intrinsic temporal structure, in addition we compute three complexity measures for each of the signals based on the spectrograms and the delay embedding technique. Table 1 indicates that transient oscillations have a higher complexity than the Lorenz signal. Moreover, the three complexity measures are similar in size for the time-frequency embedding, whereas complexity measures of delay-embedded signals exhibit dramatic differences. It is important to note that the degree of complexity based on the spectrogram is in the range of the number of recurrent states of the signals, whereas the complexity values of delay-embedded signals well exceed the number of recurrent states.

Since these results on single datasets do not allow to evaluate whether complexity measures are well adapted to distinguish temporal structures, Fig. 6 gives the distribution of complexity measures for both artificial datasets. We observe that all complexity measures show significantly different distributions. Qualitatively, the largest difference between both signals is reflected in the LZ complexity measure.

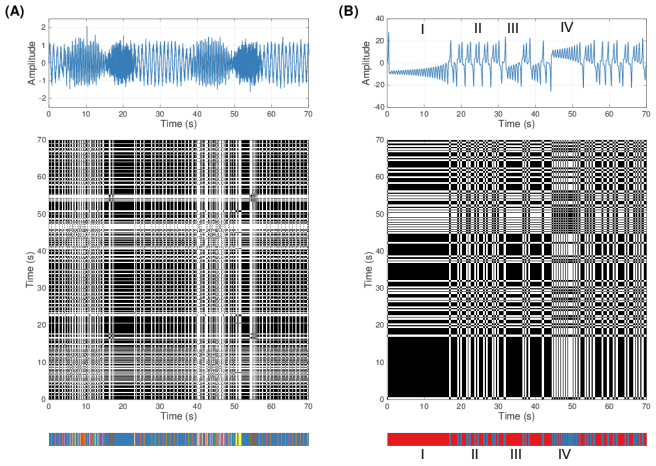

Fig. 5. Results obtained with method of delays. (A) The transient oscillations, reconstruction parameters: $m = 5$, $\tau = 0.1$ s; (B) the Lorenz system, reconstruction parameters: $m = 3$ and $\tau = 0.18$ s.

Table 1. Complexity measures of transient oscillations and Lorenz signal with spectrogram and delay embedding.

Complexity measure	Transient oscillations	Lorenz system
Spectrogram		
Alphabet size	8	2
Nr. of words	12	11
Lempel-Ziv	13	10
Delay embedding		
Alphabet size	20	2
Nr. of words	80	29
Lempel-Ziv	285	35

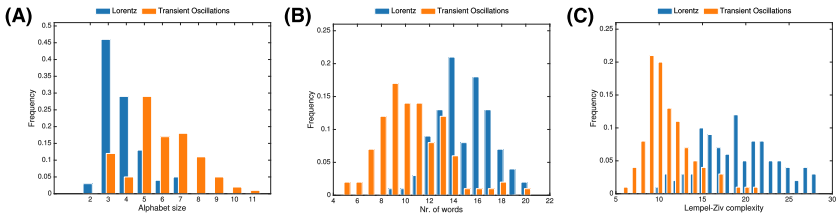

Fig. 6. Three complexity measure distributions for Lorenz system and transient oscillations. (A) Alphabet size; (B) number of words; (C) Lempel-Ziv complexity. For each complexity measure, both distributions are significantly different (Kolmogorov-Smirnov test with $p < 0.001$).

3.2 EEG data

Finally, we study experimental EEG data. Figure 7 shows time-frequency plots (spectrogram) with corresponding symbolic sequences for two patients before and after incision during surgery. We observe activity in two frequency bands, namely strong power in the δ-band ($1-5$ Hz) and lower power in the α-range ($8-12$ Hz). This finding is in good accordance to previous findings in this EEG dataset [26]. The corresponding spectral power is transient in time in both frequency bands, whose temporal structure is well captured by the recurrence analysis as seen in the symbolic sequence.

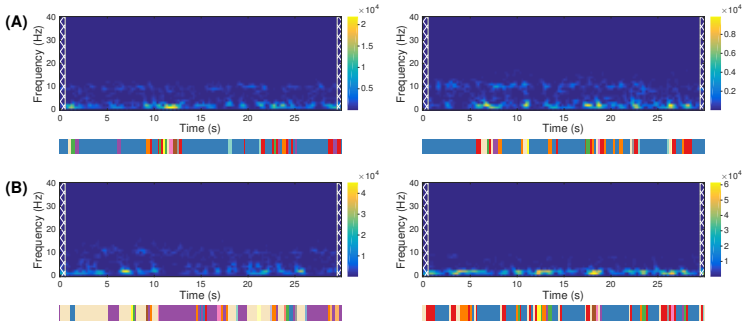

Fig. 7. Results for EEG signals obtained with spectrogram. Pre-incision is on the left and post-incision is on the right. (A) Patient #1065; (B) Patient #1099.

In order to characterize the temporal structure, we compute the symbolic sequences' recurrence complexity, cf. Table 2. We observe that the complexity values of the various complexity measures are very similar in pre- and post-incision data and close between patients. Since the time periods of pre- and post-incision data are captured several minutes apart and hence the corresponding data are uncorrelated, their similarity of complexity measures is remarkable pointing out to a constant degree of complexity in each patient. This is in line with the different complexity measures in both patients indicating different complexity measures.

4 Discussion

The present work shows that recurrence analysis can be employed on univariate time series if, at first, the data is transformed to its time-frequency representation. This transform provides a multivariate time series whose number of dimensions is equal to the number of frequencies considered. We show that the best time-frequency representation for the synthetic time series studied is the spectrogram. The recurrence structures extracted can be represented by a symbolic sequence whose symbolic complexity may serve as an indicator of the time series

Table 2. Complexity measures of EEG signals (spectrogram).

Complexity measure	Pre-incision	Post-incision
Patient #1065		
Alphabet size	11	12
Nr. of words	25	26
Lempel-Ziv	29	27
Patient #1099		
Alphabet size	16	11
Nr. of words	31	26
Lempel-Ziv	30	31

complexity. The application to experimental EEG data reveals small differences in the symbolic complexities between two experimental conditions and larger differences between subjects. This indicates that the symbolic complexity may serve as a classifier to distinguish temporal structures in univariate time series.

References

1. Addison, P.S.: The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine, and finance. Institute of Physics Pub. (2002)
2. Allefeld, C., Atmanspacher, H., Wackermann, J.: Mental states as macrostates emerging from eeg dynamics. Chaos 19, 015102 (2009)
3. Auger, F., Flandrin, P., Lin, Y.T., McLaughlin, S., Meignen, S., Oberlin, T., Wu, H.T.: Time-Frequency Reassignment and Synchrosqueezing: An Overview. IEEE Signal Process. Mag. 30(6), 32–41 (2013)
4. Donner, R., Hinrichs, U., Scholz-Reiter, B.: Symbolic recurrence plots: A new quantitative framework for performance analysis of manufacturing networks. Eur. Phys. J. Spec. Top. 164(1), 85–104 (2008)
5. Eckmann, J.P., Kamphorst, S.O., Ruelle, D.: Recurrence Plots of Dynamical Systems. Europhys. Lett. EPL 4(9), 973–977 (1987)
6. Faure, P., Lesne, A.: Recurrence plots for symbolic sequences. Int. J. Bifurc. Chaos 20(06), 1731–1749 (2010)
7. Fraser, A.M., Swinney, H.L.: Independent coordinates for strange attractors from mutual information. Phys. Rev. A 33(2), 1134–1140 (1986)
8. Freeman, W.J.: Evidence from human scalp EEG of global chaotic itinerancy. Chaos 13(3), 1069 (2003)
9. Friedrich, R., Uhl, C.: Spatio-temporal analysis of human electroencephalograms: Petit-mal epilepsy. Phys. D 98, 171–182 (1996)
10. Fulop, S.A., Fitz, K.: Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications. J. Acoust. Soc. Am. 119(1), 360 (2006)
11. beim Graben, P., Hutt, A.: Detecting Recurrence Domains of Dynamical Systems by Symbolic Dynamics. Phys. Rev. Lett. 110(15), 154101 (2013)

12. beim Graben, P., Hutt, A.: Detecting event-related recurrences by symbolic analysis: applications to human language processing. Philos Trans A Math Phys Eng Sci 373(2034) (2015)
13. Hu, J., Gao, J., Principe, J.C.: Analysis of Biomedical Signals by the Lempel-Ziv Complexity: the Effect of Finite Data Size. IEEE Trans. Biomed. Eng. 53(12), 2606–2609 (2006)
14. Hutt, A., Riedel, H.: Analysis and modeling of quasi-stationary multivariate time series and their application to middle latency auditory evoked potentials. Phys. D 177, 203–232 (2003)
15. Kennel, M.B., Abarbanel, H.D.I.: False neighbors and false strands: A reliable minimum embedding dimension algorithm. Phys. Rev. E 66(2) (2002)
16. Kennel, M.B., Brown, R., Abarbanel, H.D.I.: Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys. Rev. A 45(6), 3403–3411 (1992)
17. Kugiumtzis, D., Christophersen, N.D.: State space reconstruction: method of delays vs singular spectrum approach. Res. Rep. Httpurn Nb NoURN NBN No-35645 (1997)
18. Larralde, H., Leyvraz, F.: Metastability for Markov processes with detailed balance. Phys. Rev. Lett.
19. Lempel, A., Ziv, J.: On the complexity of finite sequences. IEEE Trans. Inf. Theory 222(1), 75–81 (1976)
20. Liebert, W., Schuster, H.G.: Proper choice of the time delay for the analysis of chaotic time series. Phys. Lett. A 142(2), 107–111 (1989)
21. Lorenz, E.N.: Deterministic nonperiodic flow. J. Atmos. Sci. 20(2), 130–141 (1963)
22. Marwan, N., Carmenromano, M., Thiel, M., Kurths, J.: Recurrence plots for the analysis of complex systems. Phys. Rep. 438(5-6), 237–329 (2007)
23. Marwan, N., Kurths, J.: Line structures in recurrence plots. Phys. Lett. A 336, 349–357 (2005)
24. Packard, N.H., Crutchfield, J.P., Farmer, J.D., Shaw, R.S.: Geometry from a time series. Phys. Rev. Lett. 45(9), 712 (1980)
25. Poincaré, H.: Sur le problème des trois corps et les équations de la dynamique. Acta Math. 13(1), 3–270 (1890)
26. Sleigh, J.W., Leslie, K., Voss, L.: The effect of skin incision on the electroencephalogram during general anesthesia maintained with propofol or desflurane. J. Clin. Mon. Comput. 24, 307–318 (2010)
27. Takens, F.: Detecting strange attractors in turbulence. Springer (1981)
28. Tošić, T., Sellers, K.K., Fröhlich, F., Fedotenkova, M., beim Graben, P., Hutt, A.: Statistical frequency-dependent analysis of trial-to-trial variability in single time series by recurrence plots. Fronti. Syst. Neurosci. 9(184) (2016)
29. Yildiz, I.B., Kiebel, S.J.: A hierarchical neuronal model for generation and online recognition of birdsongs. PLoS Comput. Biol. 7, e1002303 (2011)