journal homepage: www.intl.elsevierhealth.com/journals/cmpb

A novel encoding Lempel–Ziv complexity algorithm for quantifying the irregularity of physiological time series

Yatao Zhang [a,b], Shoushui Wei [a,*], Hai Liu [b], Lina Zhao [a], Chengyu Liu [a,*]

[a] School of Control Science and Engineering, Shandong University, Jinan 250061, China
[b] School of Mechanical, Electrical & Information Engineering, Shandong University, Weihai 264209, China

ARTICLE INFO

Article history:
Received 3 September 2015
Received in revised form
21 April 2016
Accepted 19 May 2016

Keywords:
Lempel–Ziv complexity
Encoding LZ complexity
Coarse-graining process
Physiological time series

ABSTRACT

Background and objective: The Lempel–Ziv (LZ) complexity and its variants have been extensively used to analyze the irregularity of physiological time series. To date, these measures cannot explicitly discern between the irregularity and the chaotic characteristics of physiological time series. Our study compared the performance of an encoding LZ (ELZ) complexity algorithm, a novel variant of the LZ complexity algorithm, with those of the classic LZ (CLZ) and multistate LZ (MLZ) complexity algorithms.

Methods and results: Simulation experiments on Gaussian noise, logistic chaotic, and periodic time series showed that only the ELZ algorithm monotonically declined with the reduction in irregularity in time series, whereas the CLZ and MLZ approaches yielded overlapped values for chaotic time series and time series mixed with Gaussian noise, demonstrating the accuracy of the proposed ELZ algorithm in capturing the irregularity, rather than the complexity, of physiological time series. In addition, the effect of sequence length on the ELZ algorithm was more stable compared with those on CLZ and MLZ, especially when the sequence length was longer than 300. A sensitivity analysis for all three LZ algorithms revealed that both the MLZ and the ELZ algorithms could respond to the change in time sequences, whereas the CLZ approach could not. Cardiac interbeat (RR) interval time series from the MIT-BIH database were also evaluated, and the results showed that the ELZ algorithm could accurately measure the inherent irregularity of the RR interval time series, as indicated by lower LZ values yielded from a congestive heart failure group versus those yielded from a normal sinus rhythm group ($p < 0.01$).

1. Introduction

The complexity analysis method proposed by Lempel and Ziv (LZ) in 1976 [1] is a useful approach for evaluating the irregularity of physiological time series [1–3]. Variants generated by defining different coarse-graining processes have been extensively applied to a variety of physiological time series including intracranial pressure [4], electroencephalogram [5–8], electromyographic [9], mechanomyographic [2,10], nocturnal oximetry [11], and electrocardiographic (ECG) signals [12–14].

In most cases, the classic LZ complexity (CLZ) algorithm is executed by transforming an original signal into a binary sequence by comparing it with a preset median [4,6,7] or mean

* *Corresponding authors.* School of Control Science and Engineering, Shandong University, Jinan 250061, China. Mobile: 86 13964062926; fax: 86 (531)88395827.
E-mail addresses: sswei@sdu.edu.cn (S. Wei), bestlcy@sdu.edu.cn (C. Liu).

http://dx.doi.org/10.1016/j.cmpb.2016.05.010
0169-2607/

value [12,13] as the threshold. However, the binary coarse-graining process is associated with a risk of losing essential data, in addition to a radical distortion of dynamic characteristics in certain instances. When higher quantification levels (or symbols) are employed in the coarse-graining process, more detailed information in the original signal can be preserved. As an improvement of the binary coarse-graining process, a method entailing the use of three quantification levels has been employed in certain studies [5,14]. Furthermore, Sarlabous et al. [2,3] recently proposed a multistate LZ (MLZ) complexity algorithm to quantify both the irregularity and the amplitude variations in the original signal, obtaining more favorable results when the number of quantification levels was set to more than 40 [2,3]. Nevertheless, determining the optimal quantification level in a practical environment is difficult because superfluous information within the signal (e.g. spikes of high amplitude and unexpected noise) may be unnecessarily quantified when higher quantification levels are used.

Several LZ algorithms exist, but they do not clarify what type of signal property they directly quantify: irregularity or complexity. This situation exists in nonlinear analysis methods, such as entropy measures, in addition to LZ algorithms. Recent studies have closely linked existing LZ algorithms (CLZ and MLZ) with signal irregularity rather than complexity [1–3]. Furthermore, the LZ values from random signals overlap with those from chaotic signals, corroborating to the inaccuracies found in LZ algorithms.

Our study applied a novel encoding LZ (ELZ) algorithm to directly and accurately quantify the irregularity, rather than the complexity, of a physiological time series. The performance of the ELZ algorithm was compared with those of the existing CLZ and MLZ methods on both the artificial and the MIT-BIH interbeat (RR) interval time series.

2. Method

2.1. CLZ and MLZ algorithms

For CLZ complexity, the coarse-graining process is performed by comparing signal X with a threshold to transform X into a binary sequence R. That is, whenever the signal is larger than the threshold, one maps the signal to 1, otherwise, to 0. The mean or median of the signal is usually selected as the threshold [4,12,15]. In this study, the mean value was used for calculating CLZ.

The MLZ converts signal $X = x_1 x_2, \ldots, x_n$ to a 0, 1, 2, …, $\gamma - 1$ sequence S, where γ is an integer number higher than 3. The coarse-graining process is described in Refs. [2, 3]. In the current study, γ was set to 90 to execute the MLZ to derive more accurate results.

After the coarse-graining process, the LZ complexity counter $c(n)$ of the new symbol sequence can be calculated according to the following rules [4,12,16]:

Let S and Q respectively denote two strings, and SQ is the concatenation of S and Q, whereas string $SQ\pi$ is derived from SQ after its last character is deleted (π means the operation to delete the last character in the string). Let $v(SQ\pi)$ denote the vocabulary of all different substrings of $SQ\pi$. Initially, $c(n) = 1$, $S = s_1$, and $Q = s_2$, and thus $SQ\pi = s_1$. In summary, $S = s_1 s_2, \ldots, s_r$ and $Q = s_{r+1}$, and thus $SQ\pi = s_1 s_2, \ldots, s_r$. If Q belongs to $v(SQ\pi)$, then s_{r+1}, that is, Q is a substring of $SQ\pi$, and so S does not change, and renew Q to be $s_{r+1}s_{r+2}$, and then judge if Q belongs to $v(SQ\pi)$ or not. This process is repeated until Q does not belong to $v(SQ\pi)$. Next, $Q = s_{r+1}s_{r+2}, \ldots, s_{r+i}$, which is not a substring of $SQ\pi = s_1 s_2, \ldots, s_r s_{r+1}, \ldots, s_{r+i-1}$; therefore, $c(n)$ is increased by 1. Subsequently, S is renewed to be $S = s_1 s_2, \ldots, s_{r+i}$, and $Q = s_{r+i+1}$. The procedures are repeated until Q is the last character. Concurrently, $c(n)$ is the number of different substrings (new pattern) contained in the new sequence. Finally $c(n)$ can be normalized as

$$C(n) = c(n)\frac{\log_\alpha(n)}{n}, \tag{1}$$

where n is the length of signal X, α is the number of possible symbols contained in the new sequence, and $C(n)$ is the normalized LZ complexity and denotes the arising rate of new patterns within the sequence. In practice, the normalized complexity $C(n)$, instead of $c(n)$, is considered [12]. The detailed calculation process of LZ complexity is described in Refs. [1, 4, 12, 14, 16].

2.2. Encoding LZ complexity

ELZ transforms each x_i contained within the original signal $X = x_1, x_2, \ldots, x_n$ into a 3-bit binary symbol $b_1(i)b_2(i)b_3(i)$, and the process is described as follows:

(1) The first binary digit $b_1(i)$ is determined by comparing x_i with a threshold T_{mean} which is the mean of signal X, and it is defined as follows:

$$b_1(i) = \begin{cases} 0 & if \quad x_i < T_{mean} \\ 1 & if \quad x_i \geq T_{mean} \end{cases}, \quad i = 1, 2, \ldots, n, \tag{2}$$

(2) The second binary digit $b_2(i)$ is determined by the difference between x_i and x_{i-1}, and it is defined as follows:

$$b_2(i) = \begin{cases} 0 & if \quad x_i - x_{i-1} < 0 \\ 1 & if \quad x_i - x_{i-1} \geq 0 \end{cases}, \quad i = 2, 3, \ldots, n, \tag{3}$$

where $b_2(1)$ is set to 0.

(3) For the third binary digit $b_3(i)$, a variable $Flag$ is first denoted as follows:

$$Flag(i) = \begin{cases} 0 & if \quad |x_i - x_{i-1}| < dm \\ 1 & if \quad |x_i - x_{i-1}| \geq dm \end{cases}, \quad i = 2, 3, \ldots, n, \tag{4}$$

where dm is the mean distance between adjacent points within signal X. If $Flag(i)$ is 0, point x_i is relatively close to point x_{i-1}; otherwise, the two points are relatively far away. Subsequently, $b_3(i)$ is calculated as follows:

$$b_3(i) = NOT\,(b_2(i)\,XOR\,Flag(i)), \quad i = 2, 3, \ldots, n, \tag{5}$$

where $b_3(1)$ is 0. Moreover, $b_2(i) = 1$ and $Flag(i) = 1$ mean that x_i is not only higher than x_{i-1} but also is relatively farther from x_{i-1} compared with $b_2(i) = 1$ and $Flag(i) = 0$.

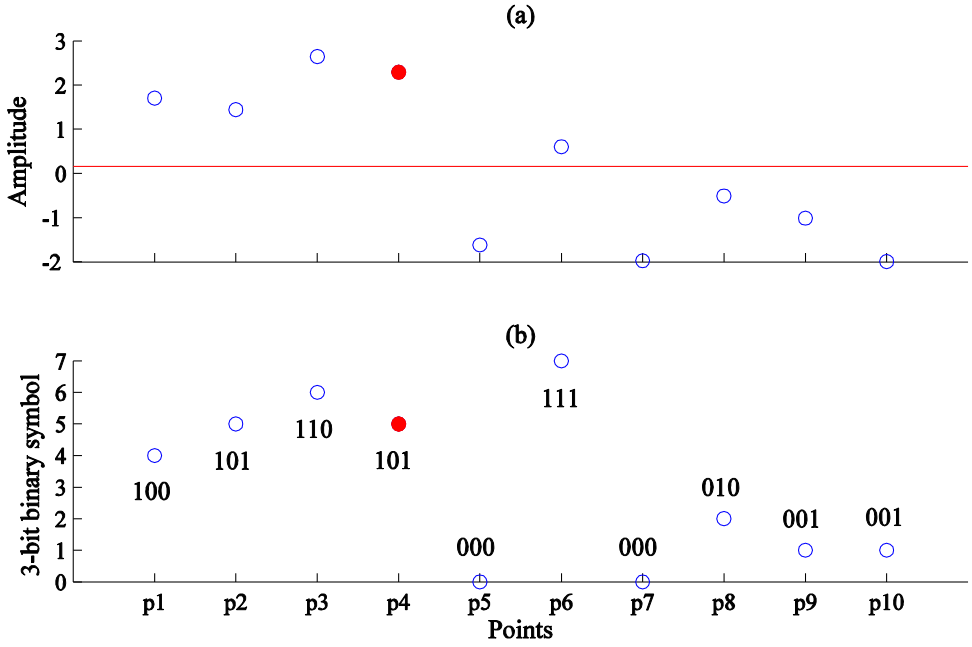

Fig. 1 – An example for the proposed coarse-graining process. (a) Before coarse-graining and (b) after coarse-graining. The red line represents the mean of the original sequence. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 1 illustrates an example for understanding the aforementioned process. Fig. 1(a) shows a 10-point sequence, and the red line represents the mean value. For point $p4$, $b_1(p4)$ is 1 because $p4$ is higher than the mean value. Furthermore, $b_2(p4)$ is 0 because $p4$ is lower than $p3$. $Flag(p4)$ is 0 because the distance between $p4$ and $p3$ is less than the mean distance dm; therefore, $b_3(p4)$ is set to 1 according to Eq. (5). Finally, point $p4$ is transformed into a 3-bit binary symbol 101, and its corresponding decimal value is 5. Fig. 1(b) presents the results of the coarse-graining process. Fig. 1 indicates that the new sequence waveform after the coarse-graining process is approximately consistent with the original sequence.

An 8-level sequence is finally derived from the original signal X to calculate the complexity $C(n)$ by using Eqs. (2–5).

2.3. *Simulation test*

2.3.1. *Artificial sequences*

Various artificial sequences, including Gaussian noise, MIX(p) processes, chaotic sequences, and periodic signals, were employed to observe the performance of ELZ. Logistic (Logi) mapping $x_{n+1} = \mu \times x_n \times (1 - x_n)$, $1 < \mu \leq 4$, is considered with $\mu = 4.0$ (Logi(4.0)) and $\mu = 3.8$ (Logi(3.8)) for chaotic sequences and $\mu = 3.5$ for periodic sequence, respectively; the closer the value is to 4, the more chaotic the sequence is. An MIX(p) process is inherently a sinusoid signal of length N, where $N \times p$ randomly chosen points are replaced with independent identically distributed random noise [17]. In the current study, $p = 0.2$ and $p = 0.4$ were used to generate two MIX(p) processes with different irregularity levels. Gaussian white noise was generated using the function (wgn) in MATLAB software.

For each type of sequence, 20 samples were employed, including 20 realizations for Gaussian noise and MIX(p) processes, respectively, and 20 randomly selected initials adopted for logistic attractors for generating 20 chaotic ($\mu = 4.0$ and $\mu = 3.8$) and 20 periodic ($\mu = 3.5$) sequences.

For further verification, surrogate data analysis was employed to clearly understand the performance of the three LZ algorithms. The surrogate data analysis entails initially specifying a linear process as a null hypothesis, then generating surrogate data sets that are consistent with this null hypothesis, and finally calculating a discriminating statistic for the original data and for each of the surrogate data sets. If the values computed from the surrogate data are significantly different from those computed from the original data, then the null hypothesis is rejected and nonlinearity is detected [18]. In the current study, the null hypothesis was that the surrogate data were consistent with the mean and variance of the original time series, and this hypothesis was generated through the linear correlation of Gaussian process. According to the null hypothesis, 20 surrogates for each realization of the logistic mapping (Logi(4.0)) process were initially generated using the Fourier transform algorithm [18]. The surrogate data could contaminate the complex structures in the logistic mapping process and increase the irregularity of the time series.

2.3.2. *Analysis of typical artificial sequence*

This analysis was conducted to compare the performance levels of the CLZ, MLZ, and ELZ algorithms by using the artificial sequences mentioned in Section 2.3.1. The lengths of the sequences, including short-time (100 points), medium-time (500 points), and long-time (5000 points), were used because these were the common length types in clinical practice.

We also analyzed the effects of sequence length on the CLZ, MLZ, and ELZ algorithms by using aforementioned typical sequences with different lengths N (i.e., 50, 100, 200, 300, 400, 500, 750, 1000, 2000, and 5000).

2.3.3. *Sensitivity analysis*

This analysis was aimed at evaluating the sensitivity of the three LZ algorithms toward the dynamic characteristics of time series. A logistic sequence was employed for the analysis; such a sequence maintains period doublings when parameter μ is less than μ^* (3.569945672) and then becomes irregular and chaotic after μ^*. We set the value of μ to the range 2.8–4.0 to test the sensitivity of the three LZ algorithms.

2.4. *MIT-BIH RR interval time series*

Normal sinus rhythm (NSR) and congestive heart failure (CHF) RR interval time series from the PhysioNet website were used as clinical data [19]. The database provided RR interval recordings from 54 individuals with NSR and 29 individuals with CHF. The original ECG signals for both the NSR and the CHF RR interval databases were digitized at 128 Hz, and the beat annotations were obtained through an automated analysis with manual review and correction.

First, RR intervals greater than 2 s were removed from the raw RR interval recordings to avoid the influence of artifacts [20]. Each beat in the raw ECG signals was annotated as a normal (denoted as "N") or an abnormal heartbeat. Abnormal heartbeats were determined to be usually caused by supraventricular or ventricular ectopic beats and were removed from the RR interval recordings. Because a baseline drift could lead to inaccurate results, the baseline curve was also extracted using an improved fast median filtering algorithm [21,22] and subsequently removed.

Finally, the RR time series were segmented with different window lengths of 50, 100, 200, 300, 400, 500, 750, 1000, 2000, and 5000 points in sequence for the subsequent complexity computation.

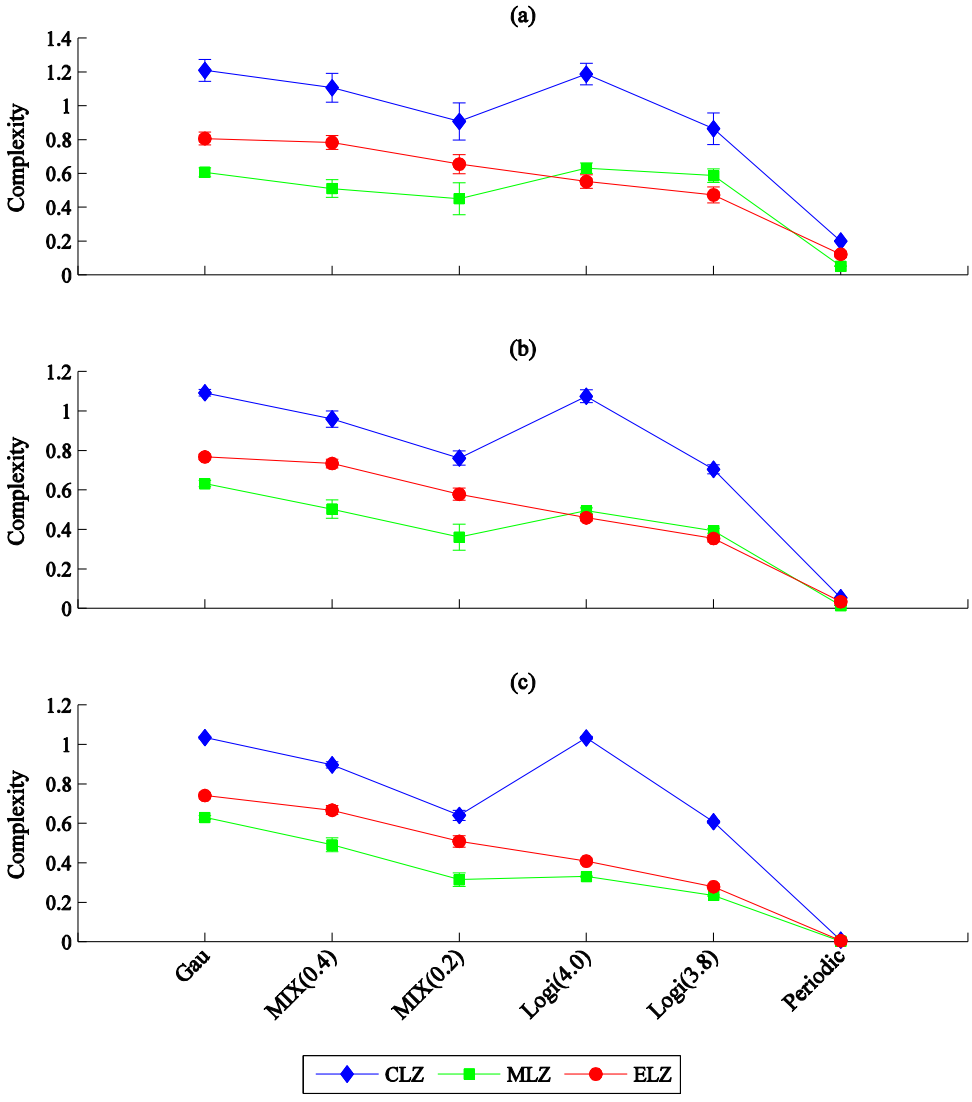

Fig. 2 – Complexity results of the six kinds of sequence for CLZ, MLZ and ELZ using three time lengths. (a) Short-time (100 points), (b) medium-time (500 points), and (c) long-time (5000 points).

3. Results

3.1. LZ results for typical artificial sequences

The complexity results of the CLZ, MLZ, and ELZ algorithms for the six artificial sequences with short-time, medium-time, and long-time length types are shown in Fig. 2(a), (b), and (c), respectively. Because MIX(*p*) processes were added by Gaussian noise, the irregularity of the MIX(*p*) processes was higher than that of the logistic sequences. The irregularity should decrease in the order of pure Gaussian noise, mixed noise sequences (MIX(*p*)), chaotic sequences (i.e., Logi(4.0) and Logi(3.8)), and periodic sequence (Logi(3.5)). From Fig. 2, for all three time length types, the complexity of the ELZ algorithm decreased stepwise in the aforementioned order, whereas those of CLZ and MLZ exhibited fluctuations at the Logi(4.0) and Logi(3.8) sequences. Therefore, the ELZ algorithm presented the ability of characterizing the irregularity rather than the complexity of the physiological time series. However, both the CLZ and the MLZ failed to distinguish between the irregularity and the complexity of the physiological time series, as indicated by the overlap in the results derived when the MIX(*p*) process and chaos logistic sequences were used.

Fig. 3 shows the results of the surrogate data analysis. The ELZ results increased significantly for all three signal length types. The CLZ results did not demonstrate discernible increases when surrogate data were used. The MLZ algorithm distinguished between the original and the surrogate data when the medium-time and long-time sequences were used, but it failed when the short-time sequence was used.

3.2. Signal length effect on LZ results

Fig. 4 illustrates the CLZ, MLZ, and ELZ results at different signal length settings. For each signal length, Gaussian noise had the highest ELZ value for quantifying irregularity, followed by the MIX(0.4) and MIX(0.2) processes, Logi(4.0) and Logi(3.8) sequences, and periodic sequence (Logi(3.5)). In addition, the ELZ algorithm yielded consistent values at signal lengths greater than 300, indicating the stability of this algorithm. However, the MLZ algorithm depended on the signal length for the logistic series (Logi(4.0) and Logi(3.8)), supporting our observation of a change in the MLZ results when the medium-time and long-time surrogate data were used and a lack of change when the short-time sequence was used, as presented in Fig. 3.

3.3. Results of the analysis of sensitivity to data change

The results derived from analyzing the sensitivity of the three LZ algorithms to changes in data sequences are illustrated in Fig. 5. Both the MLZ and the ELZ algorithms responded quickly when parameter μ approached the threshold μ^*, whereas the CLZ algorithm demonstrated a delayed response to the dynamic change in the sequence.

3.4. Results for MIT-BIH RR interval time series

Table 1 presents the mean and standard deviation of the values of the CLZ, MLZ, and ELZ algorithms for both the CHF and the NSR groups. The independent *t* test was used to compare the statistical differences among the values of the three LZ

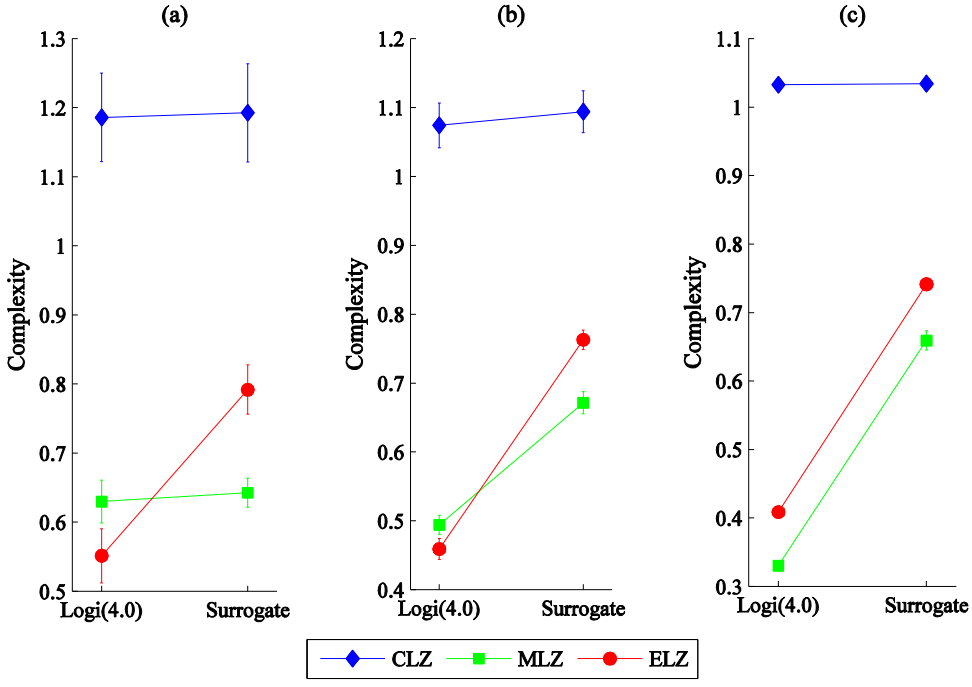

Fig. 3 – Complexity results of surrogate data test for CLZ, MLZ and ELZ. (a) Short-time (100 points), (b) medium-time (500 points), and (c) long-time (5000 points).

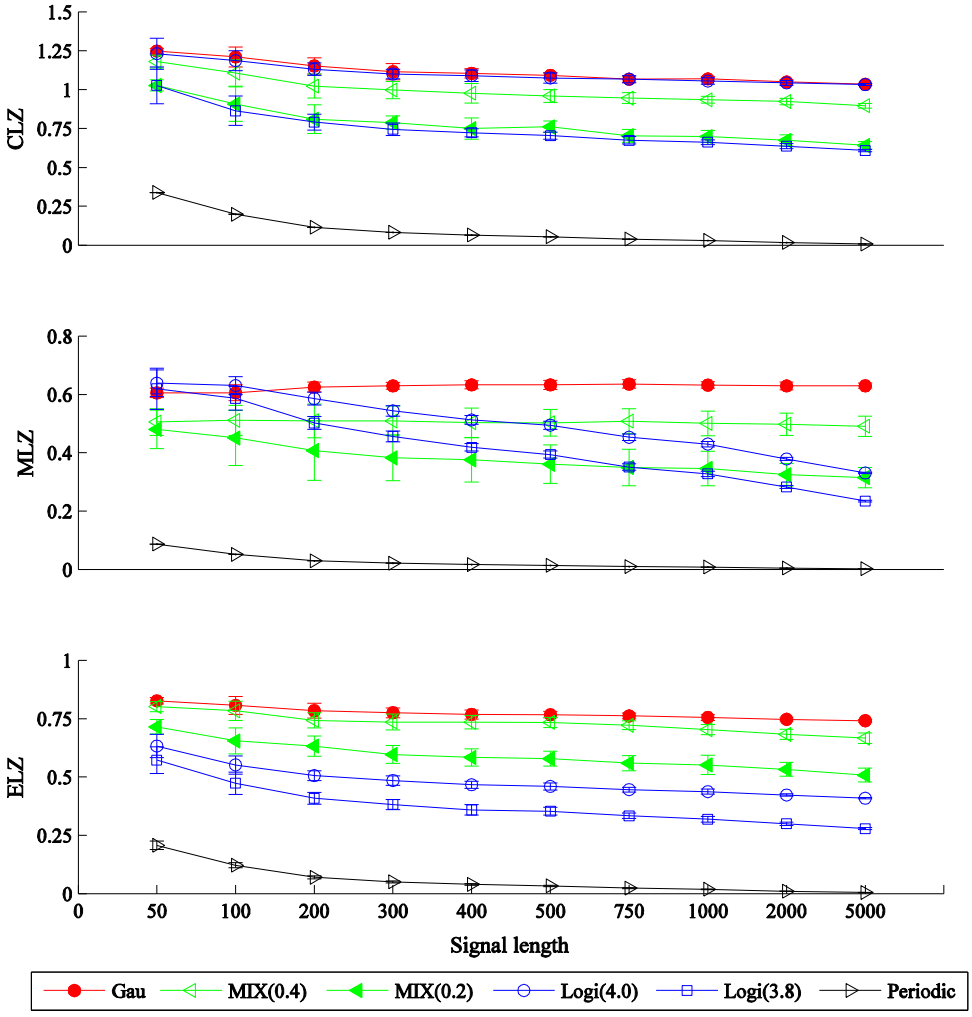

Fig. 4 – The effect of data length on the CLZ, MLZ and ELZ results. Error bar indicates the standard deviation from the 20 realizations for each of signal types.

algorithms between the groups. At various window lengths, higher CLZ values were recorded in the CHF group than that in the NSR group. Both the MLZ and the ELZ algorithms had lower LZ values for the CHF group than for the NSR group, and significant differences were observed at various window lengths. The differences in the values of the ELZ and MLZ algorithms between the groups were significant because the associated p value was considerably lower than 0.01.

4. Discussion and conclusion

We propose a novel ELZ algorithm to quantify the inherent irregularity of physiological time series. The ELZ algorithm can reflect more detailed information from an original signal than does the CLZ algorithm. Moreover, ELZ algorithm can avoid the indefinite increase in quantization levels, a phenomenon observed in the MLZ method.

An effective algorithm should sufficiently differentiate noise, chaotic, and periodic signals accurately. LZ algorithms typically evaluate the irregularity (i.e., "randomness"), rather than complexity, of time series [1,23]. As illustrated in Fig. 2, the ELZ algorithm could distinguish between chaotic and random sequences more effectively than the CLZ and MLZ algorithms did because the ELZ values monotonically decreased with the random signal components. The ELZ values from the chaotic sequences were lower than those from Gaussian noise and the Gauss-mixed MIX(p) signals, and this is because a chaotic sequence is not inherently a pure random process. By contrast, both the CLZ and the MLZ algorithms failed to distinguish between noisy sequences and chaotic sequences, particularly the CLZ algorithm. The results of the surrogate data analysis also verify that the ELZ algorithm can identify the randomness and chaotic characteristics of a signal under short-time (100 points), medium-time (500 points), and long-time (5000 points) sequences.

At various signal lengths, the ELZ and CLZ values stabilized when the sequence length exceeded 300, whereas the MLZ

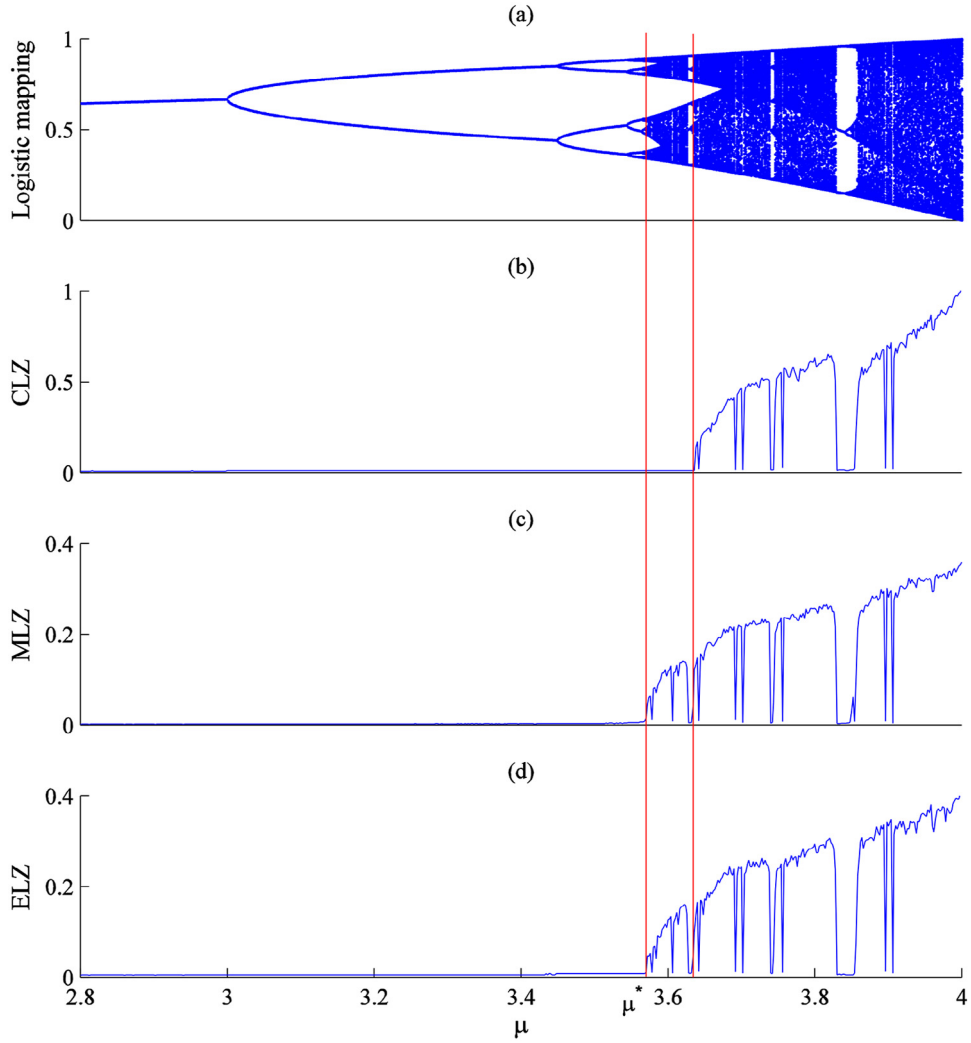

Fig. 5 – Results of the logistic mapping for CLZ, MLZ and ELZ algorithms. (a) Bifurcation diagram, result of (b) CLZ, (c) MLZ, and (d) ELZ.

Table 1 – The means and standard deviations of CLZ, MLZ and ELZ values for both CHF and NSR groups from the MIT/BIH RR interval database.

Window length	CLZ			MLZ			ELZ		
	CHF	NSR	*p*	CHF	NSR	*p*	CHF	NSR	*p*
50	0.957 ± 0.087	0.915 ± 0.060	>0.01	0.387 ± 0.070	0.470 ± 0.036	<0.01	0.745 ± 0.053	0.800 ± 0.014	<0.01
100	0.844 ± 0.096	0.810 ± 0.060	>0.05	0.401 ± 0.075	0.488 ± 0.038	<0.01	0.729 ± 0.054	0.785 ± 0.015	<0.01
200	0.782 ± 0.093	0.752 ± 0.060	>0.05	0.410 ± 0.077	0.498 ± 0.039	<0.01	0.712 ± 0.055	0.771 ± 0.016	<0.01
300	0.758 ± 0.090	0.734 ± 0.060	>0.05	0.413 ± 0.077	0.502 ± 0.039	<0.01	0.704 ± 0.057	0.763 ± 0.016	<0.01
400	0.745 ± 0.088	0.723 ± 0.059	>0.05	0.415 ± 0.077	0.503 ± 0.038	<0.01	0.698 ± 0.057	0.759 ± 0.016	<0.01
500	0.735 ± 0.088	0.717 ± 0.059	>0.05	0.416 ± 0.077	0.504 ± 0.038	<0.01	0.695 ± 0.057	0.755 ± 0.017	<0.01
750	0.723 ± 0.085	0.708 ± 0.058	>0.05	0.417 ± 0.077	0.505 ± 0.038	<0.01	0.688 ± 0.058	0.750 ± 0.017	<0.01
1000	0.714 ± 0.083	0.703 ± 0.058	>0.05	0.418 ± 0.077	0.506 ± 0.038	<0.01	0.684 ± 0.059	0.746 ± 0.018	<0.01
2000	0.702 ± 0.084	0.694 ± 0.057	>0.05	0.418 ± 0.077	0.506 ± 0.038	<0.01	0.677 ± 0.060	0.739 ± 0.018	<0.01
5000	0.695 ± 0.083	0.689 ± 0.057	>0.05	0.418 ± 0.076	0.506 ± 0.038	<0.01	0.671 ± 0.061	0.733 ± 0.018	<0.01

Data are expressed as mean ± standard deviation.

algorithm was dependent on the sequence length. Because physiological signals are typically restricted by long-term recordings, a stable LZ method is essential for clinical applications.

Another common viewpoint is that an effective algorithm should be able to respond to the dynamic and temporal characteristics of a signal. The ELZ and MLZ algorithms were sensitive to data variations, whereas the CLZ algorithm was not (Fig. 5). Therefore, the ELZ and MLZ algorithms are more effective in responding to the ever-changing challenges of time series.

For the MIT-BIH RR interval time series, the signal irregularity of CHF patients is usually considered to be less than that of NSR patients [24]. In Table 1, the ELZ and MLZ algorithms accurately measured the inherent irregularity of the RR interval time series by outputting lower LZ values for the CHF group, in which the CLZ algorithm failed to execute accurate measurements.

The reason for the superior performance of the proposed ELZ method to those of the other methods should be discussed. For calculating the normalized LZ complexity $C(n)$, the upper bound of $c(n)$ (the number of new patterns contained in the sequence) is limited as

$$c(n) < \frac{n}{(1-\varepsilon_n)\log_\alpha(n)}, \tag{6}$$

where n is the sequence length and α is the number of different symbols in the symbol set; in addition, ε_n is a small quantity, and $\varepsilon_n \to 0$ $(n \to \infty)$ [1,4,12,14,15]. Specifically,

$$\lim_{n\to\infty} c(n) = b(n) = \frac{n}{\log_\alpha(n)}, \tag{7}$$

For a finite time sequence, $c(n)$ is limited to a range determined by α. For a specified sequence, different α values can produce different $c(n)$ and $C(n)$ values. Whether the calculated LZ value can reflect the real situation of the sequence is contingent on α. For the CLZ approach, α is a relatively low value ($\alpha = 2$) because the coarse-grained sequence is a 0–1 sequence. Therefore, the upper bound of $c(n)$ calculated by Eq. (7) is lower than that in a practical situation. For example, $b(n)$ is approximately 15 when n and α are equal to 100 and 2, respectively. One possible theory is the loss of information contained in the original sequence after the coarse-graining process. For the MLZ method, α is a relatively high value ($\alpha = 90$). Hence, the upper bound of $c(n)$ calculated by Eq. (7) is higher than that in a practical situation. For example, $b(n)$ can be 98 when n and α are equal to 100 and 90, respectively. In this situation, unnecessary information contained in the original sequence is kept after the coarse-graining process. Therefore, the CLZ and MLZ values cannot accurately reflect the real situation of the sequence. The experiment proves that the ELZ algorithm is superior to its outdated counterparts, the CLZ and MLZ approaches.

Conflict of interest

The authors declare that there are no conflicts of interest to this work.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under grants 61473174 and 61201049, Shandong Provincial Natural Science Foundation in China under grant ZR2014EEM003, and the Excellent Young Scientist Awarded Foundation of Shandong Province in China under grant BS2013DX029. We thank the MIT-BIH for providing the invaluable data used in our research.

REFERENCES

[1] A. Lempel, J. Ziv, On the complexity of finite sequences, IEEE Trans. Inf. Theory 22 (1976) 75–81.
[2] L. Sarlabous, A. Torres, J.A. Fiz, J. Morera, R. Jané, Index for estimation of muscle force from mechanomyography based on the Lempel-Ziv algorithm, J. Electromyogr. Kinesiol. 23 (2013) 548–557.
[3] L. Sarlabous, A. Torres, J.A. Fiz, J. Gea, J.B. Galdiz, R. Jané, Multistate Lempel-Ziv (MLZ) index interpretation as a measure of amplitude and complexity changes, in: Proc. 31st Annual International Conference IEEE-EMBS 2009 (EMBS2009), IEEE, 2009, pp. 4375–4378.
[4] M. Aboy, R. Hornero, D. Abásolo, D. Álvarez, Interpretation of the Lempel-Ziv complexity measure in the context of biomedical signal analysis, IEEE Trans. Biomed. Eng. 53 (2006) 2282–2288.
[5] D. Abásolo, R. Hornero, C. Gómez, M. García, M. López, Analysis of EEG background activity in Alzheimer's disease patients with Lempel-Ziv complexity and central tendency measure, Med. Eng. Phys. 28 (4) (2006) 315–322.
[6] J. Dauwels, K. Srinivasan, M. Ramasubba Reddy, T. Musha, F.B. Vialatte, C. Latchoumane, et al., Slowing and loss of complexity in Alzheimer's EEG: two sides of the same coin? Int. J. Alzheimer's Dis. 2011 (2011) 539621.
[7] L. Ling, W. Ruiping, Complexity analysis of sleep EEG signal, in: The 4th International Conference on Bioinformatics and Biomedical Engineering 2010, 2010, pp. 1–3.
[8] J.C. McBride, X. Zhao, N.B. Munro, C.D. Smith, G.A. Jicha, L. Hively, et al., Spectral and complexity analysis of scalp EEG characteristics for mild cognitive impairment and early Alzheimer's disease, Comput. Methods Programs Biomed. 114 (2) (2014) 153–163.
[9] R. Nagarajan, Quantifying physiological data with Lempel-Ziv complexity-certain issues, IEEE Trans. Biomed. Eng. 49 (2002) 1371–1373.
[10] A. Torres, J.A. Fiz, R. Jané, E. Laciar, J.B. Galdiz, J. Gea, et al., Rényi entropy and Lempel-Ziv complexity of mechanomyographic recordings of diaphragm muscle as indexes of respiratory effort, in: Proc. 30th Annual International Conference IEEE-EMBS, 2008 (EMBS2008), IEEE, 2008, pp. 2112–2115.
[11] J.V. Marcos, R. Hornero, D. Álvarez, F. del Campo, C. Zamarrón, M. López, Utility of multilayer perceptron neural network classifiers in the diagnosis of the obstructive sleep apnoea syndrome from nocturnal oximetry, Comput. Methods Programs Biomed. 92 (1) (2008) 79–89.
[12] X.S. Zhang, Y.S. Zhu, N.V. Thakor, Z.Z. Wang, Detecting ventricular tachycardia and fibrillation by complexity measure, IEEE Trans. Biomed. Eng. 46 (1999) 548–555.
[13] X.S. Zhang, Y.S. Zhu, X.J. Zhang, New approach to studies on ECG dynamics: extraction and analyses of QRS complex irregularity time sequence, Med. Biol. Eng. Comput. 35 (1997) 467–473.

[14] D. Abásolo, R. Alcaraz, J.J. Rieta, R. Hornero, Lempel-Ziv complexity analysis for the evaluation of atrial fibrillation organization, in: Proc. 8th IASTED International Conference on Biomedical Engineering, 2011, pp. 30–35.
[15] J. Hu, J.B. Gao, J.C. Príncipe, Analysis of biomedical signals by the Lempel-Ziv complexity: the effect of finite data size, IEEE Trans. Biomed. Eng. 53 (2006) 2606–2609.
[16] X.S. Zhang, R.J. Roy, E.W. Jensen, EEG complexity as a measure of depth of anesthesia for patients, IEEE Trans. Biomed. Eng. 48 (2001) 1424–1433.
[17] S.M. Pincus, Approximate entropy as a measure of system complexity, Proc. Natl. Acad. Sci. U.S.A. 88 (1991) 2297–2301.
[18] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, J.D. Farmer, Testing for nonlinearity in time sequence: the method of surrogate data, Physica D 58 (1) (1992) 77–94.
[19] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, et al., PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (2000) e215–e220.
[20] L.N. Zhao, S.S. Wei, C.Q. Zhang, Y.T. Zhang, C.Y. Liu, Determination of sample entropy and fuzzy measure entropy parameters for distinguishing congestive heart failure from normal sinus rhythm subjects, Entropy 17 (2015) 6270–6288.
[21] Y.T. Zhang, C.Y. Liu, S.S. Wei, C.Z. Wei, F.F. Liu, ECG quality assessment using a kernel support vector machine and genetic algorithm with a feature matrix, J. Zhejiang Univ.-SCI. C (Comput. & Electron.) 15 (2014) 564–573.
[22] E. Ataman, V. Aatre, K. Wong, A fast method for real-time median filtering, IEEE Trans. Acoust. Speech Sig. Process. 28 (1980) 415–421.
[23] J. Ladyman, J. Lambert, K. Wiesner, What is a complex system? Philos. Sci. 3 (1) (2013) 33–67.
[24] C.Y. Liu, C.C. Liu, P. Shao, L.P. Li, X. Sun, X.P. Wang, et al., Comparison of different threshold values r for approximate entropy: application to investigate the heart rate variability between heart failure and healthy control groups, Physiol. Meas. 32 (2011) 167–180.