

Article

# Multivariate Multiscale Symbolic Entropy Analysis of Human Gait Signals

Jian Yu <sup>1</sup>, Junyi Cao <sup>1,\*</sup>, Wei-Hsin Liao <sup>2</sup>, Yangquan Chen <sup>3</sup>, Jing Lin <sup>1</sup> and Rong Liu <sup>4</sup>

<sup>1</sup> Research Institute of Diagnostics and Cybernetics, Xi'an Jiaotong University, Xi'an 710049, China; yujian\_05\_31@163.com (J.Y.); jinglin@mail.xjtu.edu.cn (J.L.)

<sup>2</sup> Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China; whliao@cuhk.edu.hk

<sup>3</sup> School of Engineering, University of California, Merced, CA 95343, USA; ychen53@ucmerced.edu

<sup>4</sup> Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hong Kong, China; rong.liu@polyu.edu.hk

\* Correspondence: caojy@mail.xjtu.edu.cn; Tel.: +86-139-9131-6384

Received: 25 September 2017; Accepted: 17 October 2017; Published: 19 October 2017

**Abstract:** The complexity quantification of human gait time series has received considerable interest for wearable healthcare. Symbolic entropy is one of the most prevalent algorithms used to measure the complexity of a time series, but it fails to account for the multiple time scales and multi-channel statistical dependence inherent in such time series. To overcome this problem, multivariate multiscale symbolic entropy is proposed in this paper to distinguish the complexity of human gait signals in health and disease. The embedding dimension, time delay and quantization levels are appropriately designed to construct similarity of signals for calculating complexity of human gait. The proposed method can accurately detect healthy and pathologic group from realistic multivariate human gait time series on multiple scales. It strongly supports wearable healthcare with simplicity, robustness, and fast computation.

**Keywords:** complexity; entropy; symbolic entropy; multivariate multiscale symbolic entropy; human gait

## 1. Introduction

The human gait is a nonlinear dynamic behavior based on the feedback of space and time, mainly controlled by the nerve and locomotor systems. Its outputs exhibit significant fluctuations on account of multiple interacting components commanded by various complex physiological systems [1]. Several conventional approaches and standard physical examinations cannot provide a complete pathological description about these complex fluctuations and the emerging complexity of an abnormal human gait. In recent decades, it has been clearly demonstrated that the complexity of the human gait can be adequately analyzed by stride interval time series, which is the gait cycle and is defined as the time interval of the same limb from a heel-strike to heel-strike again [2].

Hausdorff et al. [3–5] discussed human gait variability with aging, certain disease states, and even different walking conditions by applying detrended fluctuation analysis (DFA) and observed that there are more random or less correlated in elderly subjects and in subjects with Parkinson's disease (PD) and Huntington's disease (HD). The multiscale entropy (MSE) method proposed by Costa et al. [6] defined a quantitative measure of complexity that is large for both correlated stochastic processes and normal walking conditions. Aziz et al. [7,8] employed symbolic entropy (SyEn) to characterize human gait signals of pathological subjects with neurodegenerative diseases such as PD, HD, amyotrophic lateral sclerosis (ALS) and healthy subjects. Goshvarpour et al. [9] evaluated nonlinear and complexity characteristics of gait signals in healthy subjects who walked at their usual, slow, and fast paces

with Poincare plots, Hurst exponents, and the Lyapunov exponents. Obviously, various nonlinear methods were introduced to study complex human gait and physiological signals. Among these methods, entropy-based algorithms have received considerable attention in quantifying the complexity of physiological systems for the sake of potential implications with respect to evaluating dynamical models of biologic control systems and bedside diagnostics [10]. For instance, loss of complexity due to the reduction of physiologic information content and individual adaptive capacity resulted in aging and disease states has been proposed as a generic feature of pathologic dynamics [11,12].

Much work has been devoted to develop entropy-estimation algorithms applied to complexity measure for distinguishing physiologic signals in health and disease. Traditional entropy-based algorithms, such as Shannon entropy [13], Kolmogorov entropy [14], spectral entropy [15], wavelet entropy [16], approximate entropy [17], and sample entropy [18], are not always associated with dynamical complexity for estimating entropy. For example, some algorithms assign a larger entropy for certain pathologic processes that are generally presumed to represent less complexity than for healthy dynamics [11]. In fact, this is misleading, especially when the signal comes from more complex systems, with underlying significant correlation over multiple spatio-temporal scales. Therefore, the MSE technique [10] and its modified algorithms [19–22] are introduced to applicable analysis of various time series. Permutation entropy (PE) [23] and its improved algorithms [24,25] have been proposed as characteristics extraction in order to make a comprehensive analysis of biological and economical systems [26]. However, Kumar et al. [27] concluded SyEn is a more statistically significant separation than MSE between normal and walking under various stress conditions. Moreover, SyEn has higher calculation efficiency on the whole for different time series data lengths. PE is simple and computationally fast [23]. Nevertheless, PE does not consider the influence of the difference between amplitude values for a given time series and takes a litter more computation time than SyEn [26,28]. Consequently, SyEn [7,8,27,29] can provide an accurate assessment of the dynamic behavior and takes less computation time for various signals.

Recently, to improve the predictable accuracy of pathologic signals, there are multichannel physiologic signals or synchronous coupling multivariate time series in experimental measurements [19,25]. SyEn may be inapplicable due to the fact that these signals are statistically dependent or correlated at a certain degree. More importantly, the wearable health monitoring approach prefers simple and fast computation capability and robustness in the presence of noise. To meet these demands, a multivariate multiscale complexity measure method is proposed to robustly distinguish physiologic signals in health and disease with high computation efficiency.

In this paper, multivariate multiscale symbolic entropy (MMSyEn) is proposed to accurately quantify the complexity measure considering both within- and cross-channel dependencies and coupling in multiple channels complex signals over a range of scales. The MMSyEn calculation results of simulated stochastic data and experimental gait signals obtained under different disease states and walking conditions demonstrate the advantageous performance in the complexity quantification and characteristics extraction of real-world time series.

## 2. Multivariate Multiscale Symbolic Entropy

For a given  $p$ -variate time series  $\{x_{k,i}\}_{i=1}^N$ ,  $k = 1, 2, \dots, p$ , the consecutive moving-averaging multivariate time series  $\{y_{k,j}^\varepsilon\}$  is constructed at scale factor  $\varepsilon$ , according to the equation:  $y_{k,j}^\varepsilon = 1/\varepsilon \sum_{i=j}^{j+\varepsilon-1} x_{k,i}$ ,  $j = 1, 2, \dots, N - \varepsilon + 1$ . For scale one, the time series  $\{y_{k,j}^1\}$  is simply the original time series. Depending on the above data, the moving-averaging multivariate time series is transformed into symbolization sequence of “1” and “0” with respect to a given threshold. Therefore, quantization level 2 (symbols 0 and 1) is applied for symbolization and following is the criteria

$$Y_k^{\varepsilon,\delta} = \begin{cases} 1, & |y_{k,j}^\varepsilon - \bar{y}_k| \geq \theta \\ 0, & |y_{k,j}^\varepsilon - \bar{y}_k| < \theta \end{cases} \quad (1)$$

where  $\bar{y}_k$  is the mean of  $p$ -variate time series,  $\delta$  is quantization levels, and  $\theta$  is the threshold. There are two main methods to define the threshold value  $\theta$ : a fixed number and the threshold value  $\theta = \zeta \times \text{SD}(y)$  (SD is the standard deviation) normalized to the standard deviation of the time series [7].

The embedding vectors  $\mathbf{Y}_{k,m,i}^{\varepsilon,\delta} = \{Y_{k,i}^{\varepsilon,\delta}, Y_{k,i+\tau}^{\varepsilon,\delta}, \dots, Y_{k,i+(m-1)\tau}^{\varepsilon,\delta}\}$ ,  $i = 1, 2, \dots, n - (m-1)\tau$  can be obtained based on  $\{Y_k^{\varepsilon,\delta}\}$  using the embedding dimension  $m$  and the time delay  $\tau$ . The embedding vectors series are generated decimal numbers as

$$w_{k,i}^{\varepsilon} = Y_{k,i}^{\varepsilon,\delta} \cdot \delta^{m-1} + Y_{k,i+\tau}^{\varepsilon,\delta} \cdot \delta^{m-2} + \dots + Y_{k,i+(m-1)\tau}^{\varepsilon,\delta} \cdot \delta^0 \quad (2)$$

For word code series (embedding vectors series) having embedding dimension  $m$  and quantization level  $\delta$ , the total number of all possible words is  $\delta^m$ . Here, eight different types of words can be obtained by using the probability distribution of embedding dimension 3 (a word consisting of three symbols) and quantization level 2. Accordingly, the probability of each type of words is  $p(w_{k,m}^{\varepsilon,\delta})$ . Then, Shannon entropy is calculated as

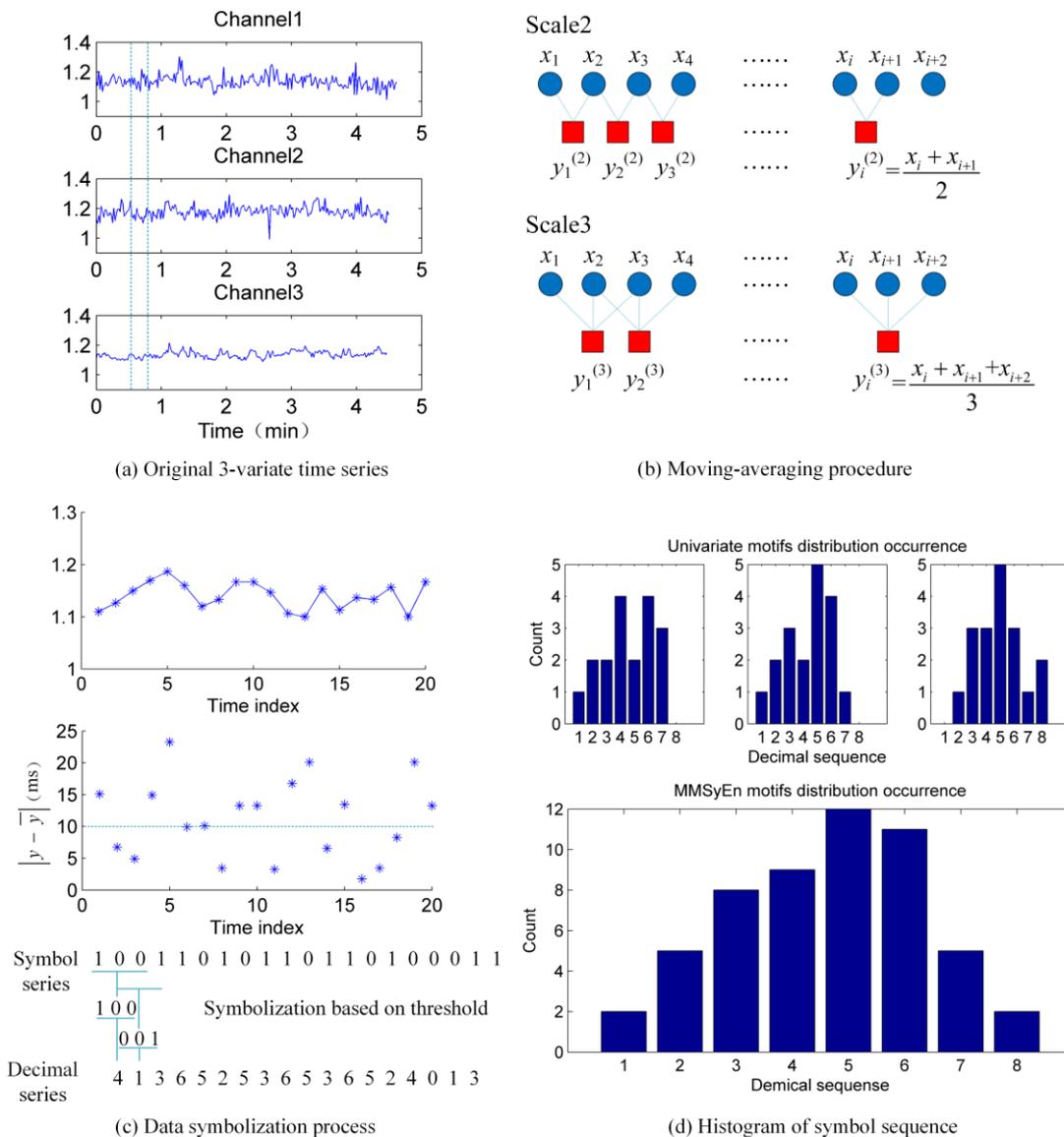
$$\text{SE}(k, \varepsilon) = -\sum p(w_{k,m}^{\varepsilon,\delta}) \cdot \log_2 p(w_{k,m}^{\varepsilon,\delta}) \quad (3)$$

In order to avoid the impact of random error in numbers and a systematic error or bias, Eguia et al. [30] proposed correction terms for the Shannon entropy (CSE)  $\text{CSE}(k, \varepsilon) = \text{SE}(k, \varepsilon) + (C - 1) / (2M \ln 2)$ , where  $M$  (equals  $\delta^m$ ) is the total number of words and  $C$  is the number of occurring words among the possible words. Unfortunately, there is still the problem to compare two values of CSE for two different embedding dimension  $m$  at the same threshold  $\theta$  and same quantization level  $\delta$ . To overcome the problem, normalized corrected Shannon entropy is employed to define MMSyEn as:

$$\text{MMSyEn}(k, \varepsilon) = \text{CSE}(k, \varepsilon) / \text{CSE}_{\max} \quad (4)$$

where the maximum value  $\text{CSE}_{\max} = -\log_2(1/M) + (M - 1) / (2M \ln 2)$  is obtained when all  $M$  words occurs with uniform distribution in a data series. It can be known that MMSyEn will vary from 0 to 1 for any parameters. A bigger MMSyEn value implies that the time series is more complex and irregular. On the contrary, a smaller MMSyEn value indicates that the time series is more regular and periodic.

In order to more clearly demonstrate the algorithm, three small segments of human gait time series are selected in Figure 1a. Then, the 3 channels time series  $\{x_{k,i}\}_{i=1}^N$  ( $k = 1, 2, 3$ ) are respectively calculated to obtain the moving-averaging time series at different scales. Figure 1c demonstrates the symbolization process. For a quantization level of 2, when the absolute values of the difference between all data values and their mean are above the threshold, the symbol series is labeled as 1 and the rest as 0. Subsequently, the symbol series is generated. After defining the embedding dimension  $m = 3$  and the time delay  $\tau = 1$ , the symbol series is converted into decimal series as shown in the last of Figure 1c. The histograms generated from the decimal series are plotted in Figure 1d. Finally, the last histogram is used to calculate the MMSyEn.



**Figure 1.** A pictorial representation of the multivariate multiscale symbolic entropy (MMSyEn) algorithm.

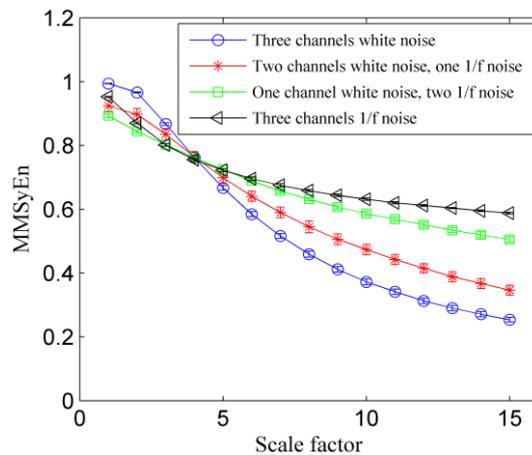
### 3. Results and Discussions

The MMSyEn analysis is evaluated for multichannel stochastic data and real-world multivariate human stride interval recordings. All computation and analysis were run on a computer with following specifications: operating system (Windows 7 Professional 64-bit), processor (AMD Athlon X2 245 @ 2.9 GHz), memory (4 GB RAM). To evaluate the statistical significant difference of the MMSyEn values for signals, the Mann-Whitney *U* test (also known as Wilcoxon rank sum test) was applied to calculate the *p* ( $p < 0.01$ ) values.

#### 3.1. Validation on Synthetic Stochastic Data

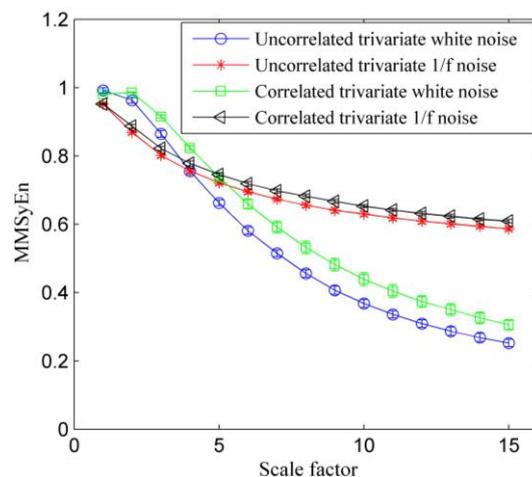
To illustrate the corresponding behavior of numerical simulations for the method of MMSyEn, it is necessary to generate four types of trivariate time series simultaneously containing white noise (the number of variables from 3–0) and independent 1/*f* noise (the number of variables from 0–3). Furthermore, trivariate white and 1/*f* noise and the corresponding correlated time series are generated to illustrate that the proposed MMSyEn fully caters for both within- and cross-channel correlations. The values of the parameters applied to calculate MMSyEn in this section are  $\tau = 1$ ,  $\theta = 0.2 \times \text{sum}(\text{SD})$  (the sum of standard deviation of the trivariate time series) and other determinate parameters for each

data channel. Figure 2 shows the MMSyEn curves for the cases considered; notice that the MMSyEn values of each type of trivariate time series monotonically decrease with the scale factor and MMSyEn is larger at higher scales as the number of variates containing 1/f noises increases, and when all the three data channels contain 1/f noise, the complexity at larger scales is the highest. Consequently, it is worth noting that the analysis of complexity is consistent with the previous research [6,10,19] that 1/f noise (long-range correlated) is structurally more complex than uncorrelated random signals.



**Figure 2.** MMSyEn analysis of 10 simulated three-channel data containing different variables of white and 1/f noise with 5000 data points. In each case, values are shown as means  $\pm$  standard deviations (SDs).

Figure 3 shows that the proposed MMSyEn approach accounts for both within- and cross-channel correlations and is able to distinguish between uncorrelated and correlated trivariate white and 1/f noises. Specifically, the MMSyEn values of four types of trivariate time series monotonically decrease with the scale factor and the MMSyEn values of the correlated trivariate 1/f noise at large scales are the largest, followed by the uncorrelated 1/f noise, correlated and uncorrelated white noise. In other words, the MMSyEn values of the correlated trivariate white and 1/f noise are larger than that of corresponding uncorrelated time series. Therefore, MMSyEn demonstrates that the complexity of correlated multivariate white noise and 1/f noise is higher, which conforms with the underlying physics [19].



**Figure 3.** MMSyEn analysis of 10 uncorrelated and correlated simulated data containing trivariate white and 1/f noise with 5000 data points. In each case, values are shown as means  $\pm$  standard deviations (SDs).

### 3.2. Complexity Analysis of Healthy Human under Different Walking Conditions

To demonstrate how MMSyEn applies to real data about a healthy human stride interval recording under different walking conditions, the MMSyEn algorithm is used to calculate the different entropy by considering three walking paces (slow, normal, fast) as multivariate from the same system. Ten young, healthy men whose mean age was 21.7 years (range: 18–29 years), height was  $1.77 \pm 0.08$  meters (mean  $\pm$  S.D.) and weight was  $71.8 \pm 10.7$  kg, walked continuously on level ground around an obstacle free, long (either 225 or 400 m), approximately oval path, and the stride interval was measured using ultra-thin, force sensitive switches taped inside one shoe [3]. Additionally, the subjects walked for one hour on a metronome with the same average speed and time to unconstrained walking state.

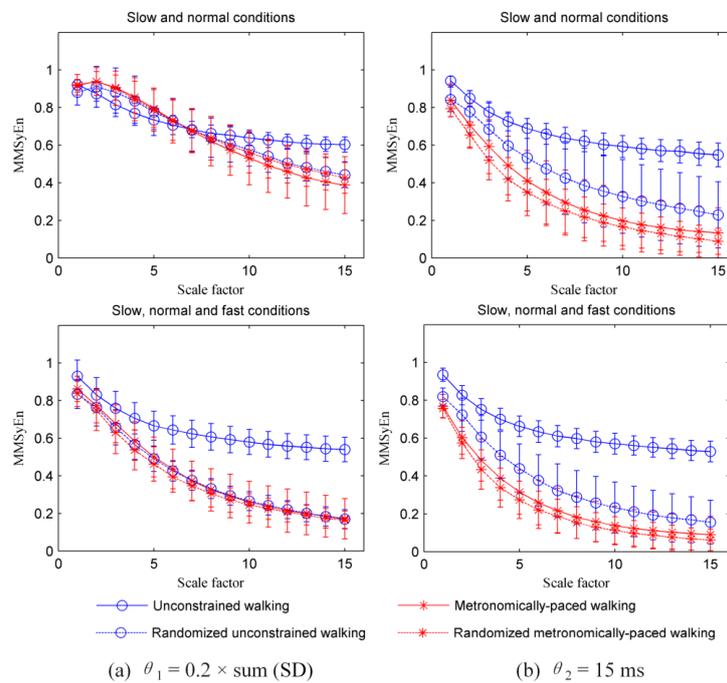
In order to discriminate the relative differences of complexity between the unconstrained and the corresponding metronomically-paced conditions, three bivariate time series (slow and normal, slow and fast, normal and fast) and one trivariate time series (slow, normal, and fast) are generated to illustrate the complexity behavior. In addition, the corresponding surrogate time series is obtained by shuffling (randomly reordering) the sequence of each stride interval time series for reasons of investigation that the correlations among the shuffled stride intervals are destroyed, while the statistical properties of the distribution are preserved.

The values of the parameters applied to calculate MMSyEn in this section are  $\tau = 1$ ,  $\theta = 0.2 \times$  (the sum of standard deviation of the multivariate time series) or  $\theta = 15$  ms ( $\theta$  can take other values, such as 8~35 ms), and other determinate parameters for each data channel. In Figure 4, the calculation results of MMSyEn show that when the walking conditions are considered within the multivariate approach (bivariate for any two walking speeds or trivariate for all the three walking speeds), the proposed algorithm can effectively discriminate between the unconstrained and metronomically-paced walking conditions at larger scales. More specifically, for bivariate time series (the top panels in Figure 4), the values of MMSyEn for unconstrained walking are larger at higher scales than that for metronomically paced walking, indicating unconstrained walking has more complex dynamics. The bottom panels in Figure 4 show that the entropy for all stride interval time series monotonically decreases with increasing scale factor and the MMSyEn curves under unconstrained walking are above those under metronomically-paced walking without the error bars overlapped at larger scales. More obviously, the distinction of the trivariate time series is more significant than that of bivariate time series. These analyses fully exhibit underlying correlations, since the MMSyEn method considers all the walking conditions within one unifying model, directly benefiting from the multivariate and multiscale approach.

The MMSyEn results for unconstrained or metronomically paced walking stride interval and their corresponding shuffled time series are also presented. In Figure 4, the values of MMSyEn for all unconstrained walking are larger at higher scales than that for the corresponding shuffled time series. The Mann-Whitney  $U$  test shows the statistical difference ( $p < 0.01$ ) of the entropy between original and shuffled time series. But for metronomically-paced walking, there is no qualitative difference ( $p > 0.01$ ) between MMSyEn curves corresponding to original stride interval and surrogate time series in contrast to the results for unconstrained walking. With regard to unconstrained walking, the results indicate that the persistent correlations or long-range dependent are presented, while the correlations would decrease for metronomically-paced walking, similar to white noise.

Table 1 shows the  $p$  values over a range of scales between unconstrained and metronomically-paced walking conditions according to the statistical test. For larger scales, all the multivariate stride interval time series can evaluate the statistical significance of the entropy statistics between unconstrained and metronomically-paced conditions. Moreover, metronomically paced walking time series share uncorrelated random underlying dynamics both within and cross-channel. On the contrary, unconstrained walking time series are correlated both within and cross-channel. Therefore, at larger scales, the output of the human locomotor system under unconstrained walking is more complex than walking under metronomically paced protocol, and the difference is statistically significant over more scales when all the available walking conditions (multivariate measurements)

are considered. Furthermore, the MMSyEn supports the general views of MSE (complexity) loss with aging and disease or the adaptive capacity reduction of biological organization at all levels when a system is under constraints (metronomically paced).



**Figure 4.** MMSyEn analysis of trials from 10 subjects containing self-paced (solid blue circle) vs metronomically paced (solid red asterisk) stride interval time series and their corresponding randomized surrogates (dashed line) with 1000 data points. Top: bivariate MMSyEn analysis; Bottom: trivariate MMSyEn analysis. In each case, values are shown as means  $\pm$  standard deviations (SDs).

**Table 1.** Mann-Whitney U test for the bivariate and trivariate human gait stride interval under different walking conditions.  $p$ -values, the comparison of unconstraint and metronomically paced walking at various speeds (slow, normal, and fast) with MMSyEn, show statistical significance at different scales. “No” represents the statistical difference is not obvious. The threshold  $\theta_1 = 0.2 \times$  (the sum of standard deviation of the multivariate time series) and  $\theta_2 = 15$  ms.

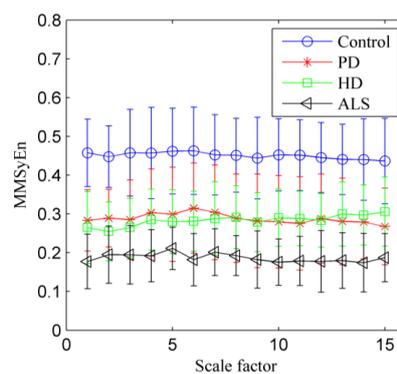
Scale	$p$ -Values (Mann-Whitney $U$ test)							
	Slow and Normal		Slow and Fast		Normal and Fast		Slow, Normal, Fast	
	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
1	No	No	No	$1.05 \times 10^{-3}$	No	$2.17 \times 10^{-5}$	No	$7.58 \times 10^{-5}$
2	No	No	No	$1.05 \times 10^{-3}$	No	$2.17 \times 10^{-5}$	No	$4.33 \times 10^{-5}$
3	No	$5.20 \times 10^{-3}$	No	$4.87 \times 10^{-4}$	No	$1.08 \times 10^{-5}$	No	$4.33 \times 10^{-5}$
4	No	$1.51 \times 10^{-3}$	No	$7.58 \times 10^{-5}$	No	$1.08 \times 10^{-5}$	No	$1.08 \times 10^{-5}$
5	No	$1.30 \times 10^{-4}$	No	$2.17 \times 10^{-5}$	No	$1.08 \times 10^{-5}$	$1.51 \times 10^{-3}$	$1.08 \times 10^{-5}$
6	No	$1.30 \times 10^{-4}$	No	$2.17 \times 10^{-5}$	No	$1.08 \times 10^{-5}$	$3.25 \times 10^{-4}$	$1.08 \times 10^{-5}$
7	No	$4.33 \times 10^{-5}$	No	$1.08 \times 10^{-5}$	No	$1.08 \times 10^{-5}$	$7.58 \times 10^{-5}$	$1.08 \times 10^{-5}$
8	No	$2.17 \times 10^{-5}$	No	$1.08 \times 10^{-5}$	No	$1.08 \times 10^{-5}$	$4.33 \times 10^{-5}$	$1.08 \times 10^{-5}$
9	No	$2.17 \times 10^{-5}$	No	$1.08 \times 10^{-5}$	No	$1.08 \times 10^{-5}$	$2.17 \times 10^{-5}$	$1.08 \times 10^{-5}$
10	No	$1.08 \times 10^{-5}$	No	$1.08 \times 10^{-5}$	$6.84 \times 10^{-3}$	$1.08 \times 10^{-5}$	$2.17 \times 10^{-5}$	$1.08 \times 10^{-5}$
11	No	$1.08 \times 10^{-5}$	$1.51 \times 10^{-3}$	$1.08 \times 10^{-5}$	$1.30 \times 10^{-4}$	$1.08 \times 10^{-5}$	$2.17 \times 10^{-5}$	$1.08 \times 10^{-5}$
12	$3.89 \times 10^{-3}$	$2.17 \times 10^{-5}$	$7.25 \times 10^{-4}$	$1.08 \times 10^{-5}$	$1.08 \times 10^{-5}$	$1.08 \times 10^{-5}$	$2.17 \times 10^{-5}$	$1.08 \times 10^{-5}$
13	$2.09 \times 10^{-3}$	$2.17 \times 10^{-5}$	$2.06 \times 10^{-4}$	$1.08 \times 10^{-5}$	$1.08 \times 10^{-5}$	$1.08 \times 10^{-5}$	$2.17 \times 10^{-5}$	$1.08 \times 10^{-5}$
14	$4.87 \times 10^{-4}$	$2.17 \times 10^{-5}$	$1.30 \times 10^{-4}$	$1.08 \times 10^{-5}$	$1.08 \times 10^{-5}$	$1.08 \times 10^{-5}$	$2.17 \times 10^{-5}$	$1.08 \times 10^{-5}$
15	$3.25 \times 10^{-4}$	$2.17 \times 10^{-5}$	$2.06 \times 10^{-4}$	$1.08 \times 10^{-5}$				

### 3.3. Complexity Analysis of Diseased Human Stride Interval

To evaluate the differences in relative complexity between the healthy subjects and the neuro-degenerative subjects (with PD, HD, and ALS), the stride interval time series of left and right foot is considered as different variables from the same system, and MMSyEn is applied to

distinguish between the healthy and diseased subjects. The subjects were instructed to walk up and down a 77-m-long, straight hallway at their self-determined rate for 5 min on level ground [4]. PD ( $n = 12$ ), HD ( $n = 15$ ), ALS ( $n = 11$ ) and 14 healthy control subjects are respectively selected to calculate their complexity with the proposed MMSyEn. Before doing that, the singular values of these signals are removed.

In reference [7], when the coefficient  $\zeta$  of threshold  $\theta$  was normalized to a unique number, it cannot realize the expectation that SyEn statistically discriminates between the control and all diseased (PD, HD, and ALS) subjects. In other words, to discriminate the healthy controls from all diseased subjects, different coefficients  $\zeta$  are selected to calculate the entropy for respectively discriminating between the healthy controls and subjects with PD, HD, and ALS. Consequently, the parameters used to calculate MMSyEn are  $\tau = 1$ , a fixed threshold  $\theta = 4$  ms (other thresholds are valid, such as 1~12 ms), and other determinate parameters. Figure 5 shows the MMSyEn curves for the cases considered; the entropy of all subjects is approximately different constants and the complexity of control subjects is the largest at all scales. Moreover, it can be observed from Table 2 that MMSyEn can discriminate between the control and diseased subjects, and the degree of distinction is good. This result also indicates lower complexity of gait responses of diseased subjects with PD, HD, and ALS than the healthy ones, thus reducing the adaptive capacity of biological organization, conforming with the complexity loss theory with disease.



**Figure 5.** MMSyEn analysis at a fixed threshold value  $\theta$ : Control versus Parkinson’s disease (PD), Huntington’s disease (HD), and amyotrophic lateral sclerosis (ALS). The values are shown as means  $\pm$  standard deviations (SDs).

**Table 2.** Mann-Whitney  $U$  test for the bivariate human gait stride interval from spontaneous output of the human locomotor system during usual walking.  $p$ -values, the comparison of healthy controls and subjects with PD, HD, and ALS with MMSyEn, show statistical significance at different scales. “No” represents the statistical difference is not obvious.

Scale	$p$ -Values (Mann-Whitney $U$ test)		
	Control vs. PD	Control vs. HD	Control vs. ALS
1	$1.86 \times 10^{-4}$	$7.74 \times 10^{-7}$	$2.04 \times 10^{-6}$
2	$2.88 \times 10^{-5}$	$8.73 \times 10^{-5}$	$3.06 \times 10^{-5}$
3	$3.05 \times 10^{-3}$	$8.73 \times 10^{-5}$	$3.06 \times 10^{-5}$
4	No	$9.12 \times 10^{-3}$	$6.83 \times 10^{-5}$
5	No	$1.77 \times 10^{-4}$	$4.08 \times 10^{-6}$
6	No	$4.21 \times 10^{-4}$	$1.22 \times 10^{-5}$
7	No	$6.83 \times 10^{-3}$	$3.06 \times 10^{-5}$
8	$6.72 \times 10^{-4}$	$2.68 \times 10^{-3}$	$1.22 \times 10^{-5}$
9	$2.50 \times 10^{-3}$	$7.65 \times 10^{-4}$	$1.22 \times 10^{-5}$
10	$1.65 \times 10^{-3}$	$4.21 \times 10^{-4}$	$1.02 \times 10^{-6}$
11	$8.50 \times 10^{-4}$	$2.35 \times 10^{-5}$	$1.02 \times 10^{-6}$
12	$1.07 \times 10^{-3}$	$9.25 \times 10^{-4}$	$1.94 \times 10^{-5}$
13	$4.48 \times 10^{-3}$	$3.70 \times 10^{-3}$	$4.59 \times 10^{-5}$
14	$6.72 \times 10^{-4}$	$1.60 \times 10^{-3}$	$4.08 \times 10^{-6}$
15	$8.50 \times 10^{-4}$	No	$3.06 \times 10^{-5}$

#### 4. Conclusions

An improved symbolic entropy is proposed to accurately quantify the complexity measure considering both within- and cross-channel dependencies and coupling in multiple channels complex signals over a range of scales. The calculation method of multivariate multiscale symbolic entropy is introduced to obtain the values of MMSyEn with normalized corrected method, selective embedding dimension, and time delay. The values of MMSyEn will vary from 0 to 1. A bigger value implies that multivariate time series is more complex and irregular.

The proposed entropy analysis of multiple channels' time series from synthetic stochastic signal is performed to verify the effectiveness of MMSyEn. It is consistent with the fact that  $1/f$  noise is structurally more complex than white noise. Moreover, human gait signals under different walking conditions and various subjects with different diseases are employed to investigate their multivariate multiscale entropy characteristics. The results of MMSyEn demonstrate that the complexity of healthy and normal gait is more distinct than that of disease and constrained walking conditions. More importantly, the proposed method possesses the advantages of symbolic entropy in term of convenience, robustness, and fast computation. It will be helpful for human wearable devices to monitoring of physiologic signals and personal healthcare in the future.

**Acknowledgments:** All authors contributed equally to the paper. We gratefully acknowledge support from the National Natural Science Foundation of China (Grant No. 51575426, 51421004, 51611530547), the Fundamental Research Funds for the Central Universities of China (Grant No. xjj2016002).

**Author Contributions:** Jian Yu, Junyi Cao and Jing Lin designed the algorithm and drafted the manuscript. Junyi Cao, Wei-Hsin Liao and Yangquan Chen conceived the study and corrected the manuscript. Jian Yu and Rong Liu collected the data and performed the experiment analysis.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Taborri, J.; Palermo, E.; Rossi, S.; Cappa, P. Gait partitioning methods: A systematic review. *Sensors* **2016**, *16*, 66. [[CrossRef](#)] [[PubMed](#)]
2. Prakash, C.; Kumar, R.; Mittal, N. Recent developments in human gait research: Parameters, approaches, applications, machine learning techniques, datasets and challenges. *Artif. Intell. Rev.* **2016**, *2016*, 1–40. [[CrossRef](#)]
3. Hausdorff, J.M.; Purdon, P.L.; Peng, C.; Ladin, Z.; Wei, J.Y.; Goldberger, A.L. Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations. *J. Appl. Physiol.* **1996**, *80*, 1448–1457. [[PubMed](#)]
4. Hausdorff, J.M.; Mitchell, S.L.; Firtion, R.; Peng, C.K.; Cudkowicz, M.E.; Wei, J.Y.; Goldberger, A.L. Altered fractal dynamics of gait: Reduced stride-interval correlations with aging and Huntington's disease. *J. Appl. Physiol.* **1997**, *82*, 262–269. [[PubMed](#)]
5. Hausdorff, J.M.; Cudkowicz, M.E.; Firtion, R.; Wei, J.Y.; Goldberger, A.L. Gait variability and basal ganglia disorders: Stride-to-stride variations of gait cycle timing in Parkinson's disease and Huntington's disease. *Mov. Disord.* **1998**, *13*, 428–437. [[CrossRef](#)] [[PubMed](#)]
6. Costa, M.; Peng, C.K.; Goldberger, A.L.; Hausdorff, J.M. Multiscale entropy analysis of human gait dynamics. *Physica A* **2003**, *330*, 53–60. [[CrossRef](#)]
7. Aziz, W.; Arif, M. Complexity analysis of stride interval time series by threshold dependent symbolic entropy. *Eur. J. Appl. Physiol.* **2006**, *98*, 30–40. [[CrossRef](#)] [[PubMed](#)]
8. Abbasi, A.Q.; Loun, W.A. Symbolic time series analysis of temporal gait dynamics. *J. Signal. Process. Syst.* **2013**, *74*, 417–422. [[CrossRef](#)]
9. Goshvarpour, A.; Goshvarpour, A. Nonlinear analysis of human gait signals. *IJIEEB* **2012**, *4*, 15–21. [[CrossRef](#)]
10. Costa, M.; Goldberger, A.L.; Peng, C.K. Multiscale entropy analysis of complex physiologic time series. *Phys. Rev. Lett.* **2002**, *89*, 068102. [[CrossRef](#)] [[PubMed](#)]
11. Goldberger, A.L.; Peng, C.K.; Lipsitz, L.A. What is physiologic complexity and how does it change with aging and disease? *Neurobiol. Aging* **2002**, *23*, 23–26. [[CrossRef](#)]

12. Goldberger, A.L.; Amaral, L.A.; Hausdorff, J.M.; Ivanov, P.; Peng, C.-K.; Stanley, H.E. Fractal dynamics in physiology: Alterations with disease and aging. *Proc. Natl. Acad. Sci. USA* **2002**, *99* (Suppl. 1), 2466–2472. [[CrossRef](#)] [[PubMed](#)]
13. Shannon, C.E. A mathematical theory of communication, Part I, Part II. *Bell Syst. Tech. J.* **1948**, *27*, 623–656. [[CrossRef](#)]
14. Kolmogorov, A.N. Entropy per unit time as a metric invariant of automorphism. *Dokl. Russ. Acad. Sci.* **1959**, *124*, 754–755.
15. Powell, G.E.; Percival, I.C. Spectral entropy method for distinguishing regular and irregular motion of Hamiltonian-systems. *J. Phys. A: Math. Gen.* **1979**, *12*, 2053–2071. [[CrossRef](#)]
16. Rosso, O.A.; Blanco, S.; Yordanova, J.; Kolev, V.; Figliola, A.; Schurmann, M.; Basar, E. Wavelet entropy: A new tool for analysis of short duration brain electrical signals. *J. Neurosci. Methods* **2001**, *105*, 65–75. [[CrossRef](#)]
17. Pincus, S.M. Approximate entropy as a measure of system-complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [[CrossRef](#)] [[PubMed](#)]
18. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [[PubMed](#)]
19. Ahmed, M.U.; Mandic, D.P. Multivariate multiscale entropy: A tool for complexity analysis of multichannel data. *Phys. Rev. E* **2011**, *84*. [[CrossRef](#)] [[PubMed](#)]
20. Wu, S.-D.; Wu, C.-W.; Lee, K.-Y.; Lin, S.-G. Modified multiscale entropy for short-term time series analysis. *Physica A* **2013**, *392*, 5865–5873. [[CrossRef](#)]
21. Humeau-Heurtier, A. Multivariate refined composite multiscale entropy analysis. *Phys. Lett. A* **2016**, *380*, 1426–1431. [[CrossRef](#)]
22. Humeau-Heurtier, A. The multiscale entropy algorithm and its variants: A review. *Entropy* **2015**, *17*, 3110–3123. [[CrossRef](#)]
23. Bandt, C.; Pompe, B. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [[CrossRef](#)] [[PubMed](#)]
24. Li, D.; Li, X.; Liang, Z.; Voss, L.J.; Sleigh, J.W. Multiscale permutation entropy analysis of EEG recordings during sevoflurane anesthesia. *J. Neural Eng.* **2010**, *7*. [[CrossRef](#)] [[PubMed](#)]
25. Morabito, F.C.; Labate, D.; La Foresta, F.; Bramanti, A.; Morabito, G.; Palamara, I. Multivariate multi-scale permutation entropy for complexity analysis of Alzheimer’s disease EEG. *Entropy* **2012**, *14*, 1186–1202. [[CrossRef](#)]
26. Zanin, M.; Zunino, L.; Rosso, O.A.; Papo, D. Permutation entropy and its main biomedical and econophysics applications: A review. *Entropy* **2012**, *14*, 1553–1577. [[CrossRef](#)]
27. Kumar, A.; Aziz, W.; Saeed, S.; Ahmed, I.; Hussain, L. Comparative Study of Multiscale Entropy Analysis and Symbolic Time Series Analysis when Applied to Human Gait Dynamics. In Proceedings of the Seventh International Conference on Open Source Systems and Technologies, ICOSST 2013, Lahore, Pakistan, 16–18 December 2013.
28. Li, Y.; Yang, Y.; Li, G.; Xu, M.; Huang, W. A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection. *Mech. Syst. Signal Pr.* **2017**, *91*, 295–312. [[CrossRef](#)]
29. Daw, C.S.; Finney, C.E.A.; Tracy, E.R. A review of symbolic analysis of experimental data. *Rev. Sci. Instrum.* **2003**, *74*, 915–930. [[CrossRef](#)]
30. Eguia, M.C.; Rabinovich, M.I.; Abarbanel, H.D. Information transmission and recovery in neural communications channels. *Phys. Rev. E* **2000**, *62*, 7111–7122. [[CrossRef](#)]

