Available online at www.sciencedirect.com
ScienceDirect
Medical Engineering & Physics 30 (2008) 299–310
Medical Engineering & Physics
www.elsevier.com/locate/medengphy

Multi-resolution entropy analysis of gait symmetry in neurological degenerative diseases and amyotrophic lateral sclerosis

Fuyuan Liao [a,*], Jue Wang [a], Ping He [b]

[a] *Key Laboratory of Biomedical Information Engineering of Education of Ministry, Xi'an Jiaotong University, Xi'an, China*
[b] *Department of Biomedical, Industrial and Human Factors Engineering, Wright State University, Dayton, OH 45435, USA*

Received 29 August 2006; received in revised form 13 April 2007; accepted 19 April 2007

Abstract

Gait rhythm of patients with Parkinson's disease (PD), Huntington's disease (HD) and amyotrophic lateral sclerosis (ALS) has been studied focusing on the fractal and correlation properties of stride time fluctuations. In this study, we investigated gait asymmetry in these diseases using the multi-resolution entropy analysis of stance time fluctuations. Since stance time is likely to exhibit fluctuations across multiple spatial and temporal scales, the data series were decomposed into appropriate levels by applying stationary wavelet transform. The similarity between two corresponding wavelet coefficient series in terms of their regularities at each level was quantified based on a modified sample entropy method and a weighted sum was then used as gait symmetry index. We found that gait symmetry in subjects with PD and HD, especially with ALS is significantly disturbed. This method may be useful in characterizing certain pathologies of motor control and, possibly, in monitoring disease progression and evaluating the effect of an individual treatment.

Keywords: Multi-resolution; Entropy; Gait symmetry

1. Introduction

Stride interval, stance interval or swing interval reflects gait rhythm. It has been shown that many neurological diseases or aging may alter the gait rhythm [1–3]. Hausdorff et al. [1,2] studied gait rhythm in Parkinson's disease (PD), Huntington's disease (HD) and amyotrophic lateral sclerosis (ALS) by applying detrended fluctuation analysis (DFA) on stride interval time series. They observed that as compared with healthy subjects, the gait rhythm of patients with these diseases is altered in several ways: (i) average stride time is longer and average walking speed is lower, (ii) the magnitude of stride-to-stride variability is increased and (iii) the fluctuation dynamics are perturbed. Aziz and Arif [3] introduced a threshold dependent symbolic entropy measure to study the stride interval complexity and found significant difference between the control subjects and the patients with the above three types of neurodegenerative diseases at wide range of thresholds. Based on their observation that the dynamics of control subjects are more complex than pathological subjects within short range of threshold, they suggested that loss of complexity is a generic feature of pathological dynamics.

Our primary interest is gait symmetry between the left and right legs. Specifically, we would like to know whether the three types of neurological disease (PD, HD, ALS) would impair the patient's ability to regulate the locomotion of the two legs and whether the degree of impairment is related to the particular type of disease that the patient has. During walking, the extension of a lower limb is accompanied by the flexion of the other. It has been suggested that there is a circuitry, probably located in the spinal cord, which regulates the alternant locomotion of the two lower limbs [4]. The knowledge of the effects of these diseases on gait symmetry would be beneficial in understanding of motor control and might lead to the development of a clinical tool for monitoring disease progression and assessing therapeutic effect. In addition, since more than one neural pathway are involved in producing a particular pattern of the lower extremity movement [5], separate studies of the degree of regularity of stride time,

* Corresponding author. Tel.: +86 29 82663497.
E-mail address: liaofuyuan1024@163.com (F. Liao).

1350-4533/$ – see front matter © 2007 IPEM. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.medengphy.2007.04.014

stance time or swing time fluctuations may reveal problems of particular neural pathways.

Gait dynamics are regulated by a complex nervous system. It is believed that the regulatory feedback loops of a physiologic system need to operate across multiple spatial and temporal scales to be able to adapt to an ever changing environment [6–8]. Thus, the time series of stride, stance or swing are likely to exhibit fluctuations across multiple spatial and temporal scales. Recently, an entropy-based method, termed multiscale entropy (MSE), was introduced to compare the relative complexity of normalized time series [7] and has been applied to stride interval time series [9]. According to the authors, if for the majority of the scales the regularity values of one time series is higher than that of the other, the former is considered more complex than the latter. Based on their findings, a reasonable approach to investigate the effects of neurological impairments on patients' ability to mediate the locomotion of two lower limbs is to compare the left and right stance-interval series in terms of their regularities at multi-resolution levels. The MSE method however, cannot be directly used for our application. That is because in our study, each experimental session is rather sort to avoid muscle fatigue. As a result, the recorded data series in our study are usually very short. For short time series, the MSE method has been shown not effective [8].

For a physiological time series, the regularity has been measured by approximation entropy (ApEn). ApEn, which was introduced by Pincus [10], is easy to apply to clinical time series but may lead to inconsistent results [11]. As a modification of the ApEn algorithm, sample entropy (SampEn) was proposed which was shown to be more accurate than ApEn [11]. However, SampEn statistics may still deviate from the theoretical values when the time series is short, which is the case in our studies.

To study the regularities of short time series at multi-resolution levels, we propose to use wavelet transformation to decompose the original data into multi levels, and to use a modified sample entropy method to measure the regularity in each level. As an effective tool for signal processing/analysis, wavelet analysis has been widely used in various fields in physics as well as in medicine. By applying wavelet transform, a data series can be decomposed into several levels without losing its information about the underlying dynamics. Specifically, at each level, a lowpass filter and a highpass filter are applied to the original data to produce two new sequences that have the same length as the original sequences. For a pair of stance interval series (one from the left leg and one from the right leg), at each decomposition level, we calculated the regularity values of two coefficient series and defined a similarity index as the ratio of the lower value to the higher value. Finally, a gait symmetry index, which is a weighted sum of all the similarity indexes, is used to reflect the relative regularity of the two original series. Our results show that gait symmetry is significantly disturbed in subjects with PD, HD, and ALS. The degree of disturbance is more prominent in the subjects with ALS.

2. Methods

2.1. Data description

The data used in this study were provided by www.physionet.org, which were recorded from 16 control subjects and subjects with PD (15 individuals), HD (20 individuals), and ALS (13 individuals) as they walked at their norm pace along a 77-m-long hallway for 5 min. The mean age (standard deviation) of the control, PD, HD and ALS subjects were 39 (18.5), 67 (10.9), 47 (13) and 56 (12.8) years, respectively. From the recorded force applied to the ground during walking, the time series of the left (right) stride time, left (right) stance time and left (right) swing time were derived. For each subject, the two time series in each left-right pair (e.g. the left stride time series and the right stride time series) have the same length, ranging from 122 to 310.

2.2. Modified sample entropy method

First, we briefly review SampEn introduced by Richman and Moorman [11].

Given a time series of N points, $\{x(1), x(2), \ldots, x(N)\}$, forms $N-m$ vectors $x_m(i) = \{x_i, x_{i+1}, \ldots, x_{i+m-1}\}$ for $1 \le i \le N-m$. A match occurs if the distance between two vectors $x_m(i)$ and $x_m(j)$ is smaller than a specified tolerance level $r\sigma$, where r is a control parameter, σ is the standard deviation (S.D.) of the original series and the distance is defined as

$$d(x_m(i), x_m(j)) = \max\{|x_{i+k} - x_{j+k}|, 0 \le k \le m-1\}. \quad (1)$$

For each of the $N-m$ vectors $x_m(i)$, which is called the template, the number of vectors $x_m(j)$ that match $x_m(i)$ is determined by measuring their respective distance. Let $n_i^m(r)$ be the number of vectors $x_m(j)$ that match $x_m(i)$ at tolerance level $r\sigma$, where $1 \le j \le N-m, j \ne i$. Then $C_i^m(r) = n_i^m/(N-m-1)$ represents the probability that any vector $x_m(j)$ matches the vector $x_m(i)$. The average of $C_i^m(r)$, $C^m(r) = (N-m)^{-1}\sum_{i=1}^{N-m} C_i^m(r)$ represents the probability that any two vectors match each other. SampEn is defined as

$$\mathrm{SampEn}(m, r) = \lim_{N\to\infty}\left[-\ln\frac{C^{m+1}(r)}{C^m(r)}\right], \quad (2)$$

which is estimated by the statistic

$$\mathrm{SampEn} = -\ln\frac{C^{m+1}(r)}{C^m(r)}. \quad (3)$$

SampEn(m, r) is the negative of the natural logarithm of the conditional probability that two sequences that are close within a tolerance $r\sigma$ for m consecutive points remain close at the next point. Although more consistent than ApEn statistic, SampEn statistic may deviate from the theoretically predicted values for short time series. Richman and Moorman suggested that one source of the bias is the correlation between the templates [11]. To reduce the bias, they proposed to par-

tition the original time series into $m+1$ sets of neighboring, disjoint vectors of length $m+1$. The conditional probability that two vectors are close to each other for m points will remain close at the next point is calculated for each of the sets and the results are averaged. However, if the total number of vectors in each set is small, a small change in the number of matches can have a large effect on the observed conditional probability. In addition, for very small values of N the number of matches might be zero. In this case, no value of conditional probability can be given.

As noted previously, Richman and Moorman used the following criterion to determine a match between two vectors:

$$n(i, j, m, r) = \begin{cases} 1, & d \le r\sigma \\ 0, & d > r\sigma \end{cases}. \tag{4}$$

Eq. (4) represents a binary decision based on a single threshold $r\sigma$. If the true distance between the two vectors is very close to the threshold, the decision can be easily influenced by a small noise and thus leads to a great uncertainty.

In brief, for short time series, SampEn statistic (Eq. (3)) is sensitive to noise and lacks robustness. To address this problem, a continuous membership function $\mu(i,j,m)$ is introduced whose value measures how $x_m(j)$ is close to $x_m(i)$, where j ranges from 1 to $N-m$ and $j \neq i$. Then $p_i^m = (N-m-1)^{-1}\sum_{j\neq i}\mu(i, j, m)$ represents the probability that any vector $x_m(j)$ is close to the vector $x_m(i)$ and $p^m = (N-m)^{-1}\sum_{i=1}^{N-m} p_i^m$ represents the probability that any two sequences are close to each other for consecutive m points. Similarly, $p^m = (N-m)^{-1}(N-m-1)^{-1}\sum_i\sum_{j\neq i}\mu(i, j, m+1)$ represents the probability that any two sequences are close to each other for consecutive $m+1$ points. Thus an entropy-like quantity is defined to measure the information increment over one step from m to $m+1$:

$$H(m, N) = -\ln\left(\frac{p^{m+1}}{p^m}\right). \tag{5}$$

The main difference between $H(m, N)$ and SampEn(m, r, N) is that in calculating $H(m, N)$, a membership function $\mu(i, j, m)$ is used to measure the similarity between two vectors. In our application, the membership function is defined as (Fig. 1)

$$\mu(i, j, m, r_1, r_2, \varepsilon) = \frac{1}{1 + \exp((d - r_1\sigma)/(r_2 - r_1)\sigma\ln((1/\varepsilon) - 1))}, \tag{6}$$

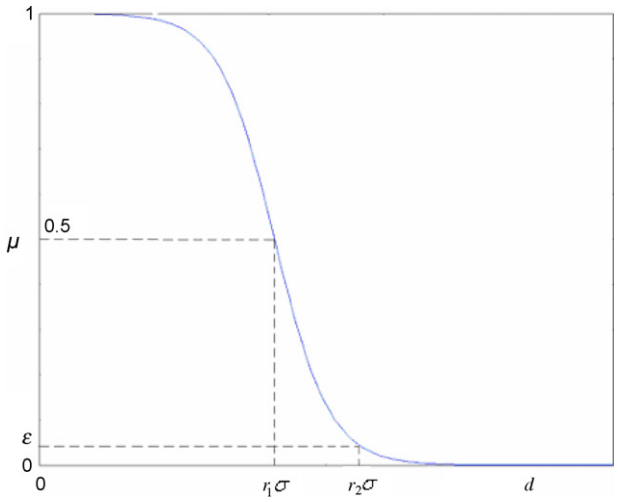

Fig. 1. The membership function μ (Eq. (6)).

where r_1, r_2 and ε are parameters to control the steepness of the transition range and d is the distance between two vectors as defined by Eq. (1). Then Eq. (5) is rewritten as

$$H(m, r_1, r_2, \varepsilon, N) = -\ln\left(\frac{p^{m+1}}{p^m}\right). \tag{7}$$

We tested the performance of $H(m, r_1, r_2, \varepsilon, N)$ on uniform iid random numbers and compared the results with that obtained by SampEn(m, r, N). The various existing rules generally lead to the use of values of r between 0.1 and 0.25 and values of m of 1 or 2 for N ranging from 100 to 5000 (data points). Since the data series are very short, we used $m=1$ in order to calculate the frequency of the m and $(m+1)$-component vectors with sufficient statistical accuracy. The test results showed that, for 500 simulated time series, the mean error (S.D.) of $H(1, 0.148, 0.2, 0.01, 100)$ and SampEn(1, 0.15, 100) is 0.44% (0.14) and 1.05% (0.17), respectively (see Fig. 2 and Table 1). For $N=180$–280, the error of $H(1, 0.148, 0.2, 0.01, N)$ is significantly smaller than that of SampEn(1, 0.15, N) (t-test). We also noted that H approaches to the theoretical value of parameter SampEn(1, 0.15) when sample number increases. Thus for all the cases presented here, we used the following parameter values: $m=1$, $r_1=0.148$, $r_2=0.2$ and $\varepsilon=0.01$. We also tested the general performances of $H(1, 0.148, 0.2, 0.01, N)$ and SampEn(1, 0.15, N) on random numbers with other distributions such as Gaussian distribution. The results were similar to those shown in Fig. 2 and Table 1.

Table 1
Bias of SampEn(1, 0.15, N) and $H(1, 0.148, 0.2, 0.01, N)$ for uniform random numbers

N	100	120	140	160	180	200	220	240	260	280	300	350	400
M_1	1.05	0.93	0.95	0.52	0.65	0.51	0.46	0.51	0.50	0.43	0.51	0.44	0.19
M_2	0.44	0.41	0.41	0.04	0.18	0.05	0.03	0.04	0.04	0.05	0.06	0.02	0.04
p value	0.147	0.149	0.071	0.066	0.049	0.037	0.020	0.007	0.042	0.024	0.081	0.362	0.463

For each N value, SampEn(1, 0.15, N) and $H(1, 0.148, 0.2, 0.01, N)$ were calculated for 500 simulated uniform random numbers. M_1, the mean error of SampEn(1, 0.15, N), expressed as the percentage of the theoretical value. M_2, the mean error of $H(1, 0.148, 0.2, 0.01, N)$. The t-test results showed that, for $N=180$–280, the error of $H(1, 0.148, 0.2, 0.01, N)$ is significantly less than that of SampEn(1, 0.15, N).

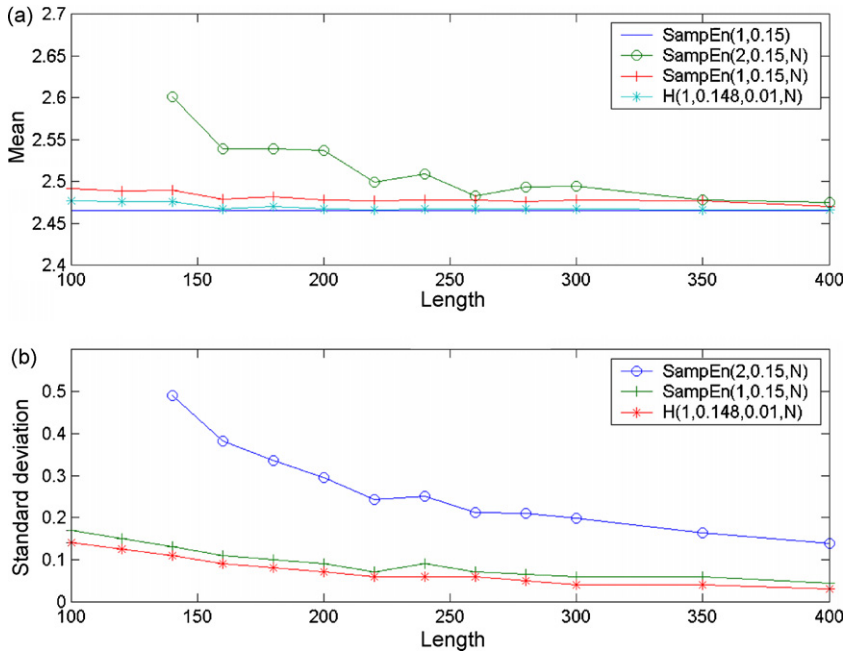

Fig. 2. (a) Mean value of SampEn(2, 0.15, N), SampEn(1, 0.15, N) and H(1, 0.148, 0.2, 0.01, N) for 500 simulated uniform random numbers. The straight line represents the theoretical value of SampEn(1, 0.15). (b) The corresponding S.D.

Next, we compared the degree of self-consistency of SampEn(1, 0.15, N) with that of H(1, 0.148, 0.2, 0.01, N). It is expected that, for two processes, S_1 and S_2, if $H(1, 0.148, 0.2, 0.01, N_1) \leq H(1, 0.148, 0.2, 0.01, N_1)(S_2)$, then $H(1, 0.148, 0.2, 0.01, N_2)(S_1) \leq H(1, 0.148, 0.2, 0.01, N_2)(S_2)$. We tested this expectation using the family of the MIX (p) processes [11]. As shown in Fig. 3, self-consistency is maintained by H(1, 0.148, 0.2, 0.01, N). For $N = 10^2$ to 5×10^4, the mean values of H for MIX (0.4) time series vary less than 0.4% and for MIX (0.6) time series vary less than 0.8%, whereas the mean values of SampEn(1, 0.15, N) for MIX (0.4) and MIX (0.6) time series vary more than 0.8% and 1.6%, respectively.

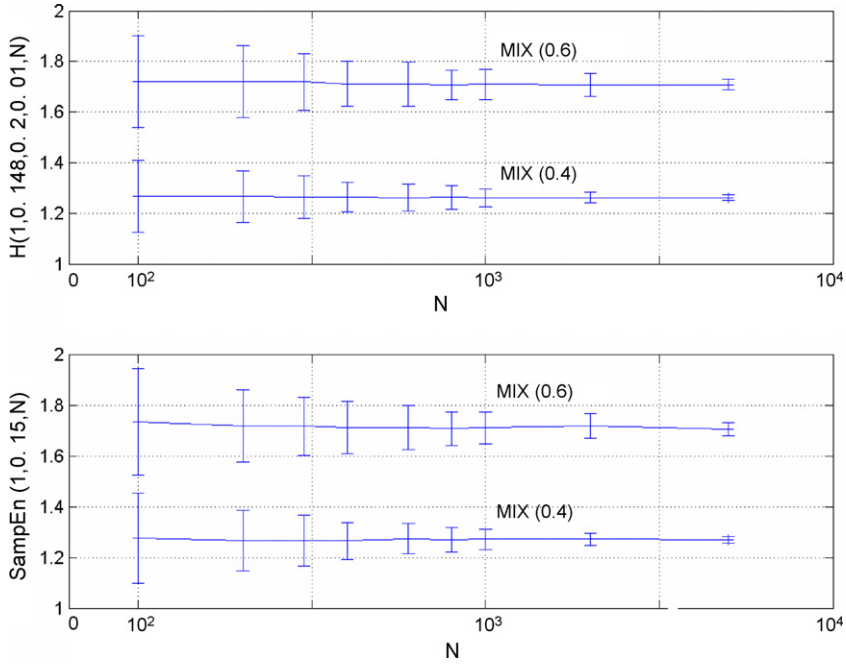

Fig. 3. Mean value (S.D.) of H(1, 0.148, 0.2, 0.01, N) and SampEn(1, 0.15, N) for the MIX (p) time series. For each 50,000-point realization of the MIX (p) process, the entropy values of all N-point nonoverlapping subsequences were calculated and then averaged. Here only the results for MIX (0.4) and MIX (0.6) processes are shown.

2.3. Multi-resolution entropy analysis

Let $L=\{L(n)\}$ be the time series of left stance time and $R=\{R(n)\}$ be the time series of the corresponding right stance time. First, data points that were 3σ greater than or less than the median value were removed. These outliers were largely due to the turns at the end of the hallway [1]. To ensure that a point in L and a point in R with the same index were in the same gait cycle, if a point $L(n)$ was removed, then the point $R(n)$ was also removed; also if a point $R(m)$ was removed, then the point $L(m)$ was also removed, where $1\leq n, m\leq M$ and M is the length of L and R. This process ensures that the lengths of the two processed series remain the same.

After choosing a wavelet basis, L and R were decomposed into J levels by applying the SWT method. For the standard discrete wavelet transform (DWT), the process of decomposition ends when the number of data points after the last subsampling become smaller than the filter length. That is, the maximum level can be estimated by $\log[N/(2k-1)]/\log(2)$, where N is the length of the data series, and k is the order of the wavelet [12]. For the SWT decomposition process, the data at each level is not decimated; however the filters are modified by padding them out with zeros. In our application, we estimated J value according to Ref. [12]. Since the data series have different lengths between 122 and 310 (each data series was processed to have the length $n\times 2^J$, where n is an integer), J value was estimated from the shortest series. We chose $J=4$ given sym4 as the wavelet basis. We used Symlets wavelet because it is a commonly used family of wavelets for discrete analysis. We also chose other kinds of wavelet basis for our application and found that the results were almost independent of the wavelet basis.

If we use $d_j^L(d_j^R)$ to represent the high frequency coefficient series at jth level and use $a_J^L(a_J^R)$ to represent the low frequency coefficient series at jth level transformed from $L(R)$, $j=1,2,\ldots,J$, the similarity between d_j^L and d_j^R in terms of their regularities is defined as

$$S_j=\frac{\min(e_j^L,e_j^R)}{\max(e_j^L,e_j^R)},\quad j=1,2,\ldots,J, \tag{8}$$

where $e_j^L(e_j^R)$ is $H(1,0.148,0.2,0.01,N)$ of $d_j^L(d_j^R)$. Similarly, the similarity between a_J^L and a_J^R is given by

$$S_{J+1}=\frac{\min(e_{J+1}^L,e_{J+1}^R)}{\max(e_{J+1}^L,e_{J+1}^R)}, \tag{9}$$

where $e_{J+1}^L(e_{J+1}^R)$ is $H(1, 0.148, 0.2, 0.01, N)$ of $a_j^L(a_j^R)$.

Given S_j, $j=1, 2, \ldots, J$ and S_{J+1}, the similarity between L and R can be defined as a weighted sum using each level's energy as weight. With energy normalized wavelet, i.e., $\int|\psi_{j,k}(t)|^2\,\mathrm{d}t=1$, an experiential formula is given by [13]

$$S=\frac{2^0S_1+2^{1/2}S_2+\cdots+2^{(J-1)/2}S_J+2^{(J-1)/2}S_{J+1}}{2^0+2^{1/2}+\cdots+2^{(J-1)/2}+2^{(J-1)/2}}. \tag{10}$$

Note that S reflects the relative structural regularity of the left and right stance-interval series, which we called gait symmetry index (GSI).

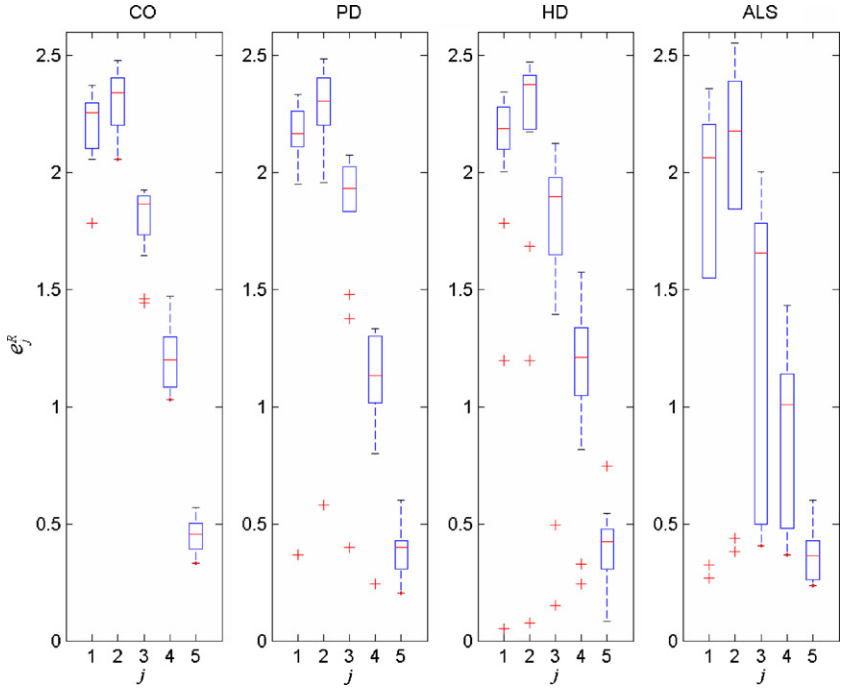

Fig. 4. $H(1, 0.148, 0.2, 0.01, N)$ of the coefficient series transformed from the right stance-interval series, i.e., e_j^R for the four groups. CO, the control group; PD, the PD group; HD, the HD group; ALS, the ALS group.

Table 2
Comparisons of e_j^R among the four groups (rank-sum test)

Resolution level	p value					
	PD vs. CO	HD vs. CO	ALS vs. CO	PD vs. HD	PD vs. ALS	HD vs. ALS
$j=1$	0.41 (0.46)	0.26 (0.43)	0.06 (0.15)	0.98 (0.78)	0.15 (0.32)	0.24 (0.34)
$j=2$	0.98 (0.77)	0.73 (0.81)	0.33 (0.21)	0.58 (0.41)	0.51 (0.35)	0.38 (0.23)
$j=3$	0.10 (0.11)	0.30 (0.23)	0.03 (0.03)	0.64 (0.78)	0.02 (0.02)	0.04 (0.03)
$j=4$	0.20 (0.12)	0.73 (0.76)	0.02 (0.08)	0.33 (0.25)	0.19 (0.50)	0.11 (0.25)
$j=5$	0.11 (0.11)	0.20 (0.16)	0.04 (0.04)	0.66 (0.73)	0.54 (0.60)	0.55 (0.58)

PD, the PD group; HD, the HD group; ALS, the ALS group; CO, the control group. The bracketed values were obtained using the standard sample entropy method.

3. Results

For the left (right) stance-interval time series, the SWT decompositions were carried out over four levels, i.e., $J=4$. We first calculated e_j^R, where $j=1, 2, \ldots, J, J+1$. Compared with the healthy subjects, e_j^R in the subjects with ALS was significantly lower for $j=3, 4, 5$ (Fig. 4 and Table 2). That is, gait rhythm in the ALS patients exhibited more regularity than that in the healthy subjects on the broad scales. However, there was no significant difference between e_j^R in the controls and that in the PD and HD patients. Among the three patient groups, e_j^R in PD and HD exhibited much common features. The analysis of e_j^L gives the similar results.

Next, we quantified gait asymmetry on each scale, i.e., calculated S_j (Eqs. (8) and (9)) for $j=1, 2, \ldots, 5$. As shown in Fig. 5 and Table 3, for $j=1, 2, 3, 5$, S_j was significantly lower in the ALS patients compared with the healthy subjects. Compared with the healthy subjects, S_j in the PD patients was lower mainly for $j=1, 2, 3, 4$, whereas S_j in the HD patients was lower mainly for $j=5$. That is, from signal processing perspective, gait asymmetry in the PD patients mainly exhibited in the high frequency components, whereas for the HD patients it mainly exhibited in the low frequency components.

As far as GSI is concerned, it was significantly reduced in the subjects with PD and HD, especially with ALS as compared with the controls (Fig. 6 and Table 4). Among the three patient groups, GSI tended to be higher in the PD patients than that in the ALS patients and similar to that in the HD patients.

The groups were not matched with respect to age, and the subjects with ALS and PD were, on average, older than other groups. To examine the effect of the age on gait symmetry, we stratified the controls into two subgroups: a young subgroup and an old subgroup (age greater than or equal to 50). The mean age of the latter was 62 years. GSI for the two subgroups

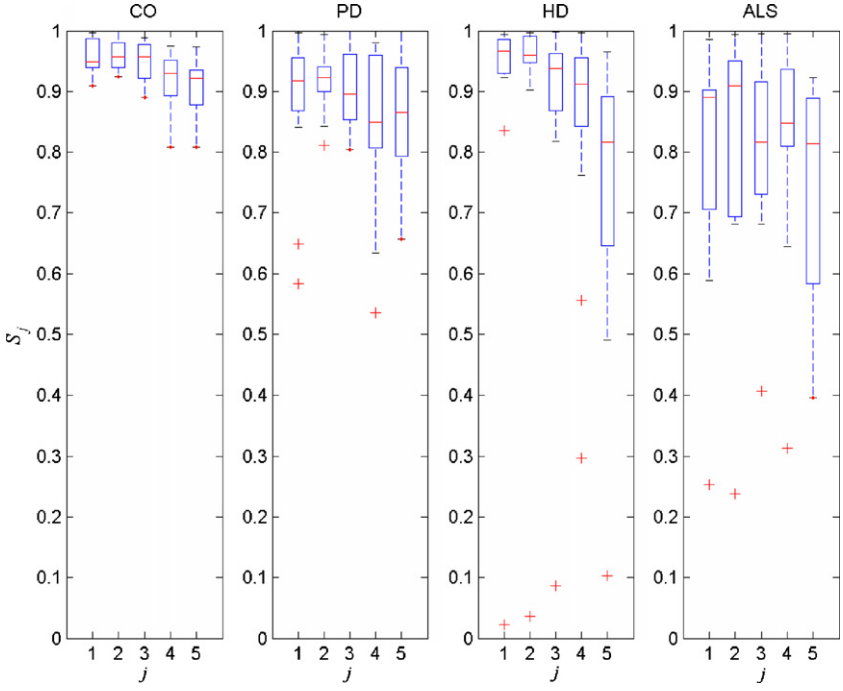

Fig. 5. Similarity at each decomposition level, i.e., S_j for the four groups (obtained using the entropy measure $H(1, 0.148, 0.2, 0.01, N)$). CO, the control group; PD, the PD group; HD, the HD group; ALS, the ALS group.

Table 3
Comparisons of S_j among the four groups (rank-sum test)

Resolution level	p value					
	PD vs. CO	HD vs. CO	ALS vs. CO	PD vs. HD	PD vs. ALS	HD vs. ALS
$j=1$	0.07 (0.08)	0.81 (0.91)	<0.001 (<0.01)	0.04 (0.06)	0.15 (0.08)	<0.01 (0.10)
$j=2$	<0.01 (0.04)	0.90 (0.49)	<0.01 (0.01)	0.01 (0.03)	0.46 (0.38)	0.02 (0.01)
$j=3$	<0.05 (0.21)	0.09 (0.20)	0.01 (0.01)	0.75 (0.89)	0.06 (0.07)	0.07 (0.06)
$j=4$	0.09 (0.08)	0.46 (0.43)	0.09 (0.16)	0.37 (0.39)	0.98 (0.84)	0.40 (0.41)
$j=5$	0.18 (0.30)	<0.01 (<0.01)	<0.01 (0.01)	0.17 (0.13)	0.17 (0.16)	0.76 (0.86)

PD, the PD group; HD, the HD group; ALS, the ALS group; CO, the control group. The bracketed values were obtained using the standard sample entropy method.

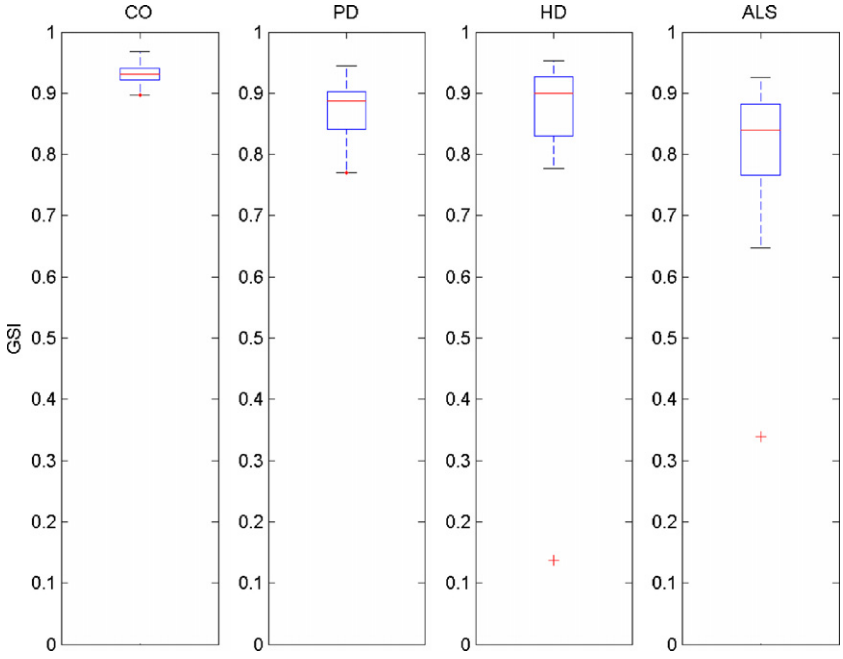

Fig. 6. Gait symmetry index (GSI) obtained based on the modified sample entropy method for the four groups.

had little difference ($p=0.86$) and the averages were almost identical.

To further study the effects of the different neurological impairment on gait symmetry and to examine whether GSI might be useful in augmenting clinical assessment, we compared the mildly impaired patients with the severely impaired ones in the PD and HD group (any specific measure of disease severity for the ALS patients was not available). The PD patients with Hoehn and Yahr scores ≥ 3 and the HD patients with total functional scores ≤ 5 were considered severely impaired [1]. For the PD group, GSI in the severely impaired patients was significantly lower than that in the mildly impaired ones ($p=0.04$) and the averages were 0.86 and 0.91, respectively. For the HD group, GSI was not statistically different in the severely impaired patients and in the mildly impaired ones ($p=0.89$) and the averages were 0.87 and 0.89, respectively.

4. Discussion

In this paper, multi-resolution entropy analysis was applied to left and right stance-interval time series to investigate the effects of PD, HD and ALS on patients' ability to mediate the locomotion of two lower limbs. Our methodological innovation was motivated by two problems. First, the data series are very short, having lengths between 122 and 310. It is well known that as the number of data points decreases, the

Table 4
Comparisons of GSI among the four groups

	PD vs. CO	HD vs. CO	ALS vs. CO	PD vs. HD	PD vs. ALS	HD vs. ALS
p value	<0.001 (<0.01)	<0.01 (0.02)	<0.001 (<0.001)	0.44 (0.58)	0.11 (0.13)	0.07 (0.11)

PD, the PD group; HD, the HD group; ALS, the ALS group; CO, the control group. The bracketed values were obtained using the standard sample entropy method.

consistency of SampEn(m, r, N) results is progressively lost. One important reason for that is that a binary function is used to measure the similarity between two vectors. To address this problem, we introduced a membership function to measure the similarity between two vectors. In our application we used Eq. (6) to define the membership function. On the other hand, m is critical in determining the outcome of either SampEn(m, r, N) or $H(m, r_1, r_2, \varepsilon, N)$ for entropy estimation. To calculate the frequency of the m and $(m+1)$-component vectors with sufficient statistical accuracy, we used $m=1$. It has been shown that, for record lengths $\geq 10^2$, the entropy measure $H(1, 0.148, 0.2, 0.01, N)$ for white noise and the MIX (p) processes is accurate enough.

Secondly, stance time presents fluctuations on multiple spatial and temporal scales. Recently, the MSE method has been applied to stride-interval time series to quantify the information expressed the gait rhythm dynamics over multiscales. However, this method requires an adequate length of data to provide reliable statistics for the entropy measure on each scale. To address this problem, we decomposed the data series into appropriate levels by applying the SWT method. The detail coefficients correspond to the reconstruction of the series after going through a high-pass filter, while the approximation coefficients corresponds to the low-pass filtered series. Since the data series have different lengths, the maximum level J was estimated from the shortest series. With this J value, however, decomposing a longer data series may omit some detail information. If we choose a larger J value, decomposing a shorter data series will introduce ineluctable redundancy in the decomposing process. By comparing different decomposition scales, we chose $J=4$ given sym4 as wavelet basis.

Table 5
Changes in e_j^R and S_j due to the change in number of data points

Resolution level	CO		PD		HD		ALS	
	$d(e_j^R)$	$d(S_j)$	$d(e_j^R)$	$d(S_j)$	$d(e_j^R)$	$d(S_j)$	$d(e_j^R)$	$d(S_j)$
$j=1$	0.1	4.2	1.3	0.3	1.2	2.5	10.9	2.9
$j=2$	2.0	0.3	2.6	1.3	1.0	2.3	11.3	2.0
$j=3$	1.9	4.2	0.5	2.1	1.0	1.9	5.6	3.1
$j=4$	1.1	3.2	9.9	4.7	5.9	3.9	1.8	4.6
$j=5$	9.6	1.0	13.6	3.6	13.6	0.5	5.6	0.5

The original data series were truncated to 112 data point (extracted the left parts). $d(e_j^R)$, the discrepancy between the mean value of e_j^R for the original data series and that for the truncated data series, expressed as the percentage of e_j^R for the original data series. Similarly, $d(S_j)$ denotes the change in the mean value of S_j, expressed as the percentage of S_j for the original data series. These results were calculated using the modified sample entropy method.

To verify the results reported in Section 3, we truncated all the data series to 112 data point (extracted the left parts) and then calculated the results using the modified entropy measure. Fig. 7 and Tables 5 and 6 show the differences occurred in the results due to the changes in number of data points. Although the maximum change in the mean values of e_j^R was larger than 10%, the maximum change in the mean values of S_j and GSI was less than 4.5% and 2.5%, respectively. These changes were mainly due to the nonstationarity of the data series. Indeed, for a nonstationary time series, different subsequences with the same length may have different entropy values. We also extracted the central parts of the original data series to calculate the indexes and found that the maximum change in the mean values of S_j and GSI was less than 4.1% and 2.3%, respectively. These results indicated that, compared with the healthy subjects, gait in the subjects with PD, HD and ALS was significantly asymmetry.

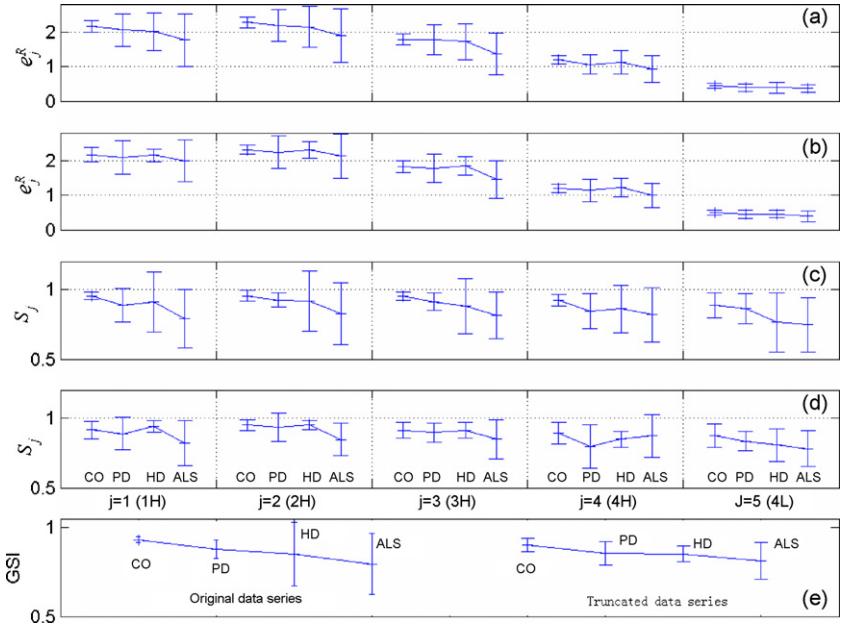

Fig. 7. Illustration of the differences between the results obtained from the original data sets and those obtained from the truncated data sets. The original data series were truncated to 112 points (extracted the left parts). The entropy values were calculated using the modified sample entropy method. (a, c) Mean value (S.D.) of e_j^R and S_j for the original data sets. (b, d) Mean value (S.D.) of e_j^R and S_j for the truncated data sets.

Table 6
Comparisons of GSI among the four groups for the case of truncated time series

	PD vs. CO	HD vs. CO	ALS vs. CO	PD vs. HD	PD vs. ALS	HD vs. ALS
p value	<0.01 (0.02)	0.04 (0.04)	<0.02 (<0.02)	0.53 (0.55)	0.26 (0.29)	0.18 (0.21)

The original data series were truncated to 112 data point (extracted the left parts). PD, the PD group; HD, the HD group; ALS, the ALS group; CO, the control group. The bracketed values were obtained using the standard sample entropy method.

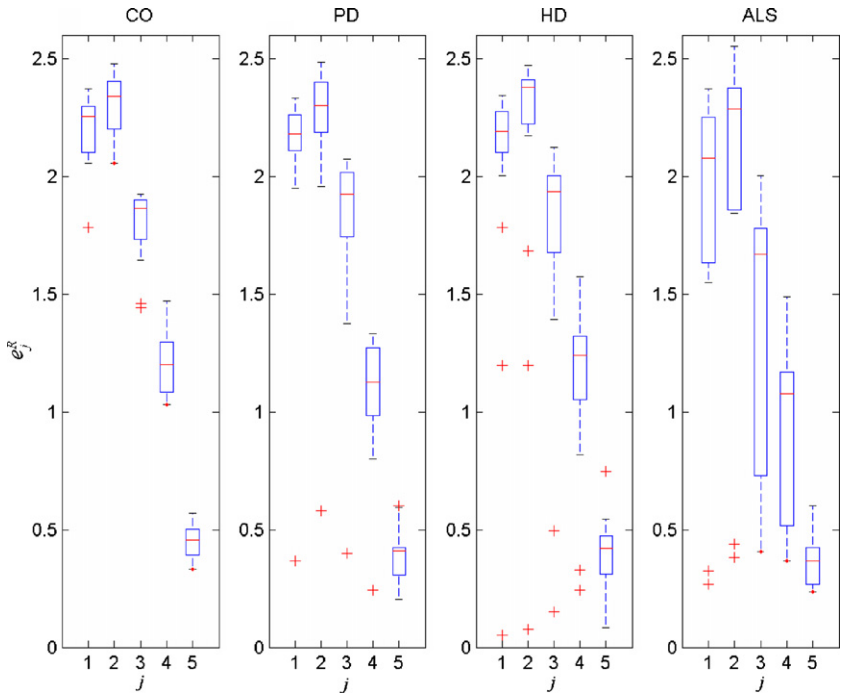

Fig. 8. SampEn(1, 0.15, *N*) of the coefficient series transformed from the right stance-interval series, i.e., e_j^R for the four groups. CO, the control group; PD, the PD group; HD, the HD group; ALS, the ALS group.

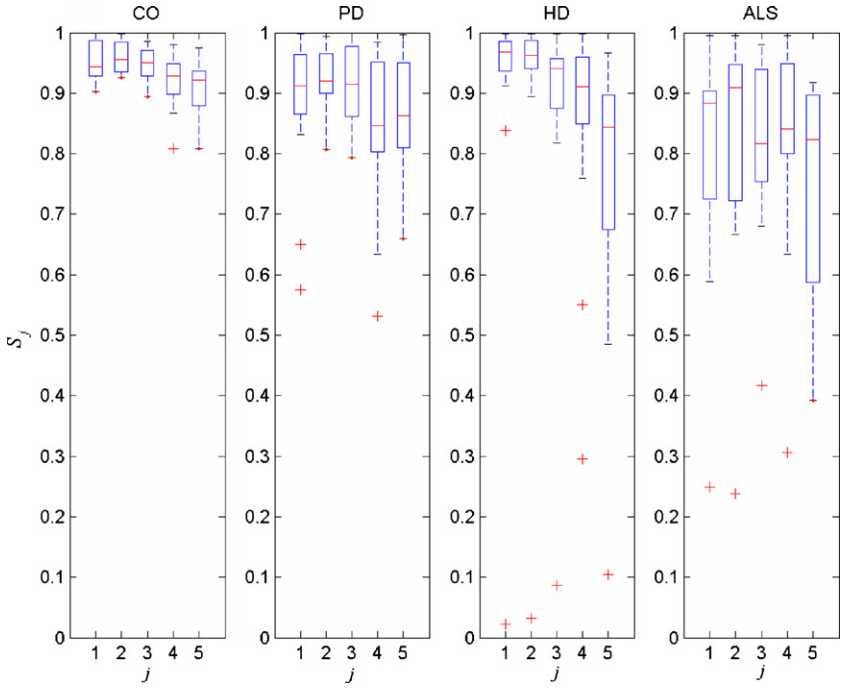

Fig. 9. Similarity at each decomposition level, i.e., S_j for the four groups (obtained using the entropy measure SampEn(1, 0.15, *N*)). CO, the control group; PD, the PD group; HD, the HD group; ALS, the ALS group.

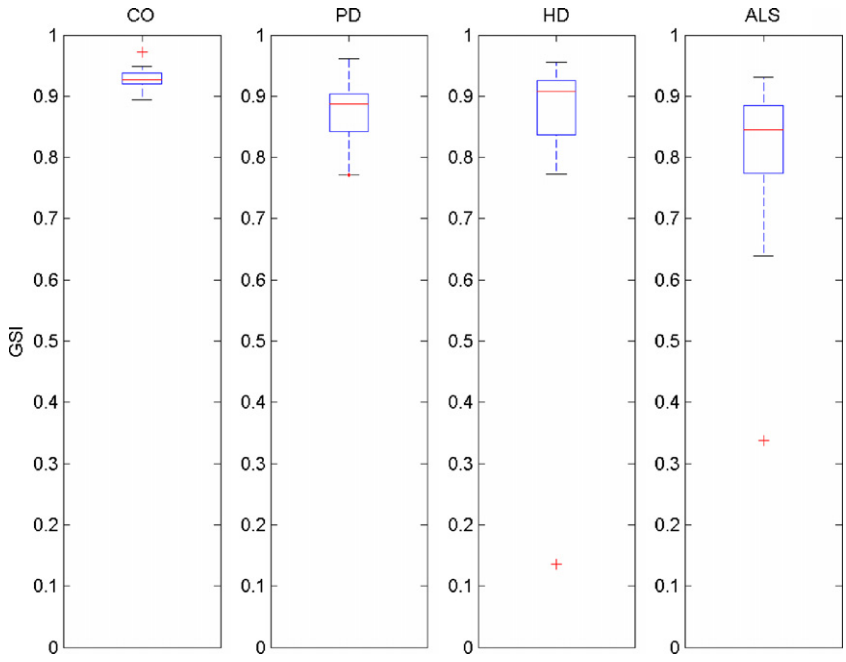

Fig. 10. Gait symmetry index (GSI) obtained based on the standard sample entropy method for the four groups.

The results reported in Section 3 were based on the modified entropy method. By applying the standard entropy method, the obtained results were similar to but slightly different from those reported earlier (see Figs. 4–6, 8–10 and Tables 2–4). For instance, using the measure $H(1, 0.148, 0.2, 0.01, N)$, the obtained mean GSI (S.D.) of the control, PD, HD and ALS subjects was 0.93 (0.017), 0.87 (0.048), 0.85 (0.184) and 0.78 (0.174), respectively, while using SampEn(1, 0.15, N), the mean GSI (S.D.) for the four groups was 0.93 (0.018), 0.88 (0.053), 0.85 (0.181) and 0.79 (0.172), respectively. However, by comparing the p values in Tables 2–4 and 6, it can be seen that the results obtained using the modified entropy measure are more reliable for reflecting the differences among the four groups.

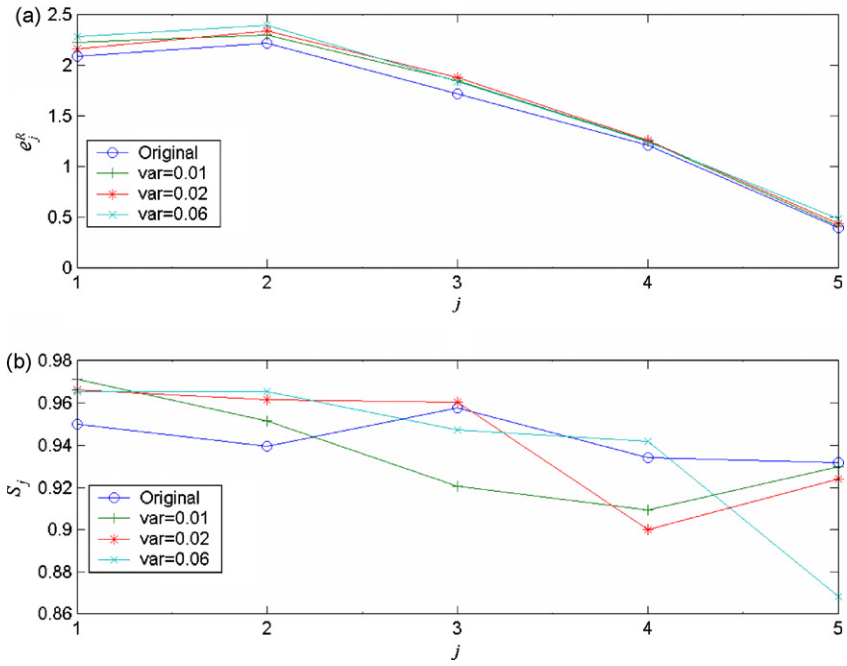

Fig. 11. Effects of Gaussian white noise with different energy levels on e_j^R and S_j. For each noise variance, e_j^R and S_j were calculated 30 times and then averaged. The curves labeled "original" corresponds to the results for the data series from a HD patient.

The original data series are noisy. Here, we discuss the effects of the multi-resolution entropy analysis of superimposing white noise on a data set. Fig. 11 shows that (i) superimposing noise affects mainly the entropy values on small scales (Fig. 11(a)); (ii) the S_j values assigned to the data series with superimposed Gaussian noise increase on small scales ($j = 1, 2$) but decrease on the largest scale (Fig. 11(b)). These effects become more prominent as the noise energy increases. However, noise has little effects on GSI. The GSI of the original data series is 0.94. For the noise variance of 0.01, 0.02, and 0.06, the mean value of GSI is 0.93,0.93 and 0.92, respectively.

The index S_j, which measures the similarity between left and right stance-interval series in terms of their regularities at a specific resolution level, may be useful in discriminating various neurological disorders. Several neurological disorders, including PD, HD and ALS, have been associated with alteration in the central and peripheral mechanisms resulting in specific patterns of gait and balance disturbance. Since gait rhythm presents fluctuations on multiple spatial and temporal scales, monoscale analysis of the data sets may be incapable of indicating the differences among the patterns of gait disturbance. We have applied the detrended fluctuation analysis (DFA) to the median-filtered left and right stance-interval time series to calculate the scaling exponents. However, no significant difference was found between the control group and PD group ($p = 0.62$), also between the control group and HD group ($p = 0.28$). As shown in Fig. 5 and Table 3, compared with the control subjects, gait asymmetry in the PD patients mainly exhibited in high frequency components, whereas for HD patients, it mainly exhibited in low frequency components. For the ALS patients, gait asymmetry exhibited in all frequency components. Furthermore, the indexes S_j and GSI may also be useful in quantitatively monitoring disease progression and evaluating the effect of an individual treatment. For instance, gait in the advanced PD patients was more asymmetry than that in the mildly impaired ones. In this study, each data set was derived from the recorded force applied to the ground during a 5-min walk. Since the force applied to the ground can be measured by a lightweight, portable foot-switch system [1], this technique can be implemented easily for practical applications. Although some highly sophisticated gait analysis techniques including electromyography (EMG) analysis have been used in gait laboratories as research tools, but the utility of these techniques is limited by their inaccessibility and the time and expense involved in performing the testing.

As noted above, prominent gait asymmetry was found in ALS patients. ALS primarily affects the motorneuros of the cerebral cortex, brain stem, and spinal cord and leads to muscle weakness and muscle fatigability [4]. In Ref. [1], it was noted that muscle weakness and muscle fatigability by themselves do not appear to be responsible for changes in gait rhythm. Accordingly, if these factors have some effects on gait dynamics, they would not necessarily have different effects on two lower limbs. HD is marked by loss of neural projection in the stratum and PD is marked by the death of the substantia nigra of the basal ganglia [4]. The mechanisms through which these changes affect the ability of the basal ganglia to regulate motor control are largely unknown. Theoretically, the basal ganglia network could influence movement amplitude by mediating on-line feedback control process and/or feedback planning [14]. A recent study [15] suggested that the basal ganglia is directly involved in the planning of movement amplitude. On small temporal scales, gait asymmetry in subjects with PD is significantly different from that in subjects with HD, whereas on large scales gait asymmetry in these two groups has much common features. These evidences indicate that, although the primary pathology of PD and HD are all in the basal ganglia, it is unlikely that the same mechanisms are affecting subjects with PD and HD. Perhaps the affected regions within the basal ganglia are different.

To further understand the mechanisms of these disorders, it might be necessary to study other aspects of walking. For example, studying changes in kinetics may provide useful information about the alterations of gait in diseases. On the other hand, to track disease progression and assess therapeutic effect, great efforts are needed to develop better diagnostic methods. Finally, we note that, in this study, the analysis of the swing time fluctuations gave results similar to those shown in Section 3. However, for stride time, the results were not meaningful. Perhaps the alterations of gait rhythm are more likely to exhibit in the stance time and swing time fluctuation dynamics.

Acknowledgement

This work was partly supported by the National Natural Science Foundation of China (No. 60271025).

Conflict of interest

There is no conflict of interest.

References

[1] Hausdorff JM, Lertratanakul A, Cudkowicz ME, Peterson AL, Kaliton D, Goldberger AL. Dynamic markers of altered gait rhythm in amyotrophyic lateral sclerosis. J Appl Physiol 2000;88(6): 2045–53.

[2] Hausdorff JM, Mitchell SL, Firtion R, Peng C-K, Cudkowicz ME, Wei JY, et al. Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease. J Appl Physiol 1997;82(1):262–9.

[3] Aziz W, Arif M. Complexity analysis of stride interval time series by threshold dependent symbolic entropy. Eur J Appl Physiol 2006;98(1):30–40.

[4] Bear MF, Connors BW, Paradiso MA. Neuroscience: exploring the brain. 2nd ed. Baltimore: Lippincott Williams & Ilkins; 2001.

[5] Sporns O, Edelman GM. Solving Bernstein's problem: a proposal for the development of coordinated movement by selection. Child Dev 1993;64(4):960–81.
[6] Goldbergera AL, Peng C-K, Lipsitz LA. What is physiologic complexity and how does it change with aging and disease? Neurobiol Aging 2002;23(1):23–6.
[7] Costa M, Goldberger AL, Peng C-K. Multiscale entropy analysis of complex physiological time series. Phys Rev Lett 2002;89:068102.
[8] Costa M, Goldberger AL, Peng C-K. Multiscale entropy analysis of biological signals. Phys Rev E 2005;71:021906.
[9] Costa M, Peng C-K, Goldberger AL, Hausdorff JM. Multiscale entropy analysis of human gait dynamics. Physica A 2003;330(1):53–60.
[10] Pincus SM. Approximation entropy as a measure of system complexity. Proc Natl Acad Sci USA 1991;88(7):2291–301.
[11] Richman JS, Moorman JR. Physiological time series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol 2000;278(6):H2039–49.
[12] Kaiser G. A friendly guide to wavelets. Boston: Birkhäuser; 1994. p. 44–5.
[13] Wen Z-N, Wang K-L, Li M-L, Nie F-S, Yang Y. Analyzing functional similarity of protein sequences with discrete wavelet transform. Comput Biol Chem 2005;29(3):220–8.
[14] Desmurget M, Grafton ST, Vindras P, Grea H, Robert ST. Basal ganglia network mediates the control of movement amplitude. Exp Brain Res 2003;153(2):197–209.
[15] Desmurget M, Grafton ST, Vindras P, Grea H, Robert ST. The basal ganglia network mediates the planning of movement amplitude. Eur J Neurosci 2004;19(10):2871–80.