British Journal of Applied Science & Technology
3(4): 1097-1105, 2013

SCIENCEDOMAIN *international*
www.sciencedomain.org

Entropy in the Analysis of Gait Complexity: A State of the Art

B. De La Cruz Torres[1], M. D. Sánchez López[2], E. Sarabia Cachadiña[3] and J. Naranjo Orellana[3*]

[1]Department of Physiotherapy, University of Sevilla, Spain.
[2]Andalussian Center for Sports Medicine; Sevilla, Spain.
[3]Department of Sports and Computing, Pablo de Olavide University; Sevilla, Spain.

Authors' contributions

This work was carried out in collaboration between all authors. Author JNO designed the study. Authors BCT and MDSL performed the statistical analysis and wrote the protocol. BCT wrote the first draft of the manuscript. Auhor ESC managed literature searches. Authors BCT, ESC, JNO managed the analyses of the study. All authors read and approved the final manuscript.

Review Article

Received 2nd May 2013
Accepted 20th July 2013
Published 25th July 2013

ABSTRACT

Human gait is a non-linear complex process requiring appropriate mathematical measuring tools. Entropy is a measure that quantifies regularity in time series: the more predictable a series is the lower the entropy value. The mathematical methods used to estimate entropy have evolved over time. At present, three algorithms are the most used to study human gait complexity: the approximate entropy (ApEn), the sample entropy (SampEn), and the multi-scale entropy (MSE). Most studies on human gait complexity have been conducted on elderly subjects or subjects with specific disorders affecting gait patterns and they used ApEn; but, because of a set of conceptual errors, the ApEn is not the most appropriate algorithm for the analysis of biological signals. Very few studies use SampEn or MSE to analyze human gait variability, but they agree that these algorithms might contribute new perspectives in the analysis of human gait and that MSE seems to be the most sensitive algorithm to changes in gait in healthy subjects.

Keywords: Entropy; human gait; complexity; multiscale.

**Corresponding author: E-mail: jnarore@upo.es;*

1. INTRODUCTION

Physiological signals present complex and irregular fluctuations that cannot be analyzed with conventional statistical analysis methods, as these methods yield very limited information on behavior patterns.

There is not any clear definition of “complexity”. Generally, a complex signal accounts for one (normally more than one) of the following characteristics [1]: a) *Non-linearity*; complex systems are composed of multiple subunits that should not be separately analyzed since they may interact with each other; b) *non-stationarity*: the statistical properties of the system input change with time; c) *time irreversibility* or *asymmetry*: the system operation is not balanced; d) *multi-scale variability:* the system exhibits patterns on multiple space-time scales.

Human gait is considered a complex, non-linear process [2-4] by which the locomotor system incorporates input from the cerebellum, the motor cortex, and the basal ganglia, as well as feedback from visual, vestibular, and proprioceptive sensors. Traditionally, under healthy conditions, the locomotor system is thought to produce a remarkably stable walking pattern; the kinetics, kinematics, and muscular activity of gait appear to remain relatively constant from one step to the next, even during unconstrained walking [4-9]. For this reason, most conventional biomechanical studies are based on the thorough analysis of a walking cycle. The data obtained are then extrapolated into the whole walking process. However, a number of studies, conducted using the non-linear dynamic approach, have revealed that gait patterns present fluctuations even under apparently stable conditions [10-13]. Thus, human gait dynamics have a complex behavior that many studies have attempted to elucidate [14] using practical applications mainly focused on aging and pathologies affecting human walking.

There are several devices which analyze human gait complexity using pressure or force sensors [15-16] or hip and knee angles [17]. Also, there are many mathematical methods that have proven useful in examining the complexity of biological signals such as stride intervals [18,19], Detrended Fluctuation Analysis (DFA) [20-22], power law scaling by Fourier’s method (23), the Lyapunov exponent (24-25) or entropy [26]. This paper focus on the entropy as the most common nonlinear tool for human signals analysis.

This review is divided into two sections: the first section develops the concept of entropy as a nonlinear tool for the analysis of physiological signals complexity and the second section shows the application of entropy to the analysis of human gait complexity.

2. CONCEPT OF ENTROPY. ALGORITHMS FOR ESTIMATING ENTROPY

The concept of “entropy” was first developed in classic thermodynamics as a measure of the molecular disorder within a closed system.

In the field of non-linear dynamic systems, entropy quantifies the regularity of a system: the more predictable a series is the lower the entropy value. For example, it is known that aging reduces entropy and impaired systems exhibit lower entropy than healthy systems [27].

The more regular a series, the more predictable and the less complex it will be, which is indicative of a less adaptive system. Therefore, in any time series accounting for a system output variable, entropy is a measure of its uncertainty.

The mathematical methods used to estimate entropy have evolved over time. At present, there are three algorithms: the approximate entropy (ApEn), the sample entropy (SampEn), and the multi-scale entropy (MSE).

In 1991, Pincus introduced ApEn (26) as a tool to measure regularity of a time series under the assumption that a series is regular when it contains repetitive patterns [28,29]. A time series containing many repetitive patterns will have lower values of ApEn, while a complex series (i.e. without repetitive patterns and, consequently, poorly regular and less predictable) will show greater values of ApEn.

The SampEn was proposed by Richmann and Moorman [30] to correct some errors to which the ApEn leads. As the ApEn counts each sequence as matching itself, in a time series it yields a degree of similarity greater than the real one. So, when the calculations above are applied to physiological data, impaired systems sometimes yield greater values of entropy than healthy systems [31]. This is illogical, since an impaired system is less adaptive, less complex and more regular than a healthy system and it should yield a lower value of entropy. Such inconsistencies do not appear with SampEn.

The SampEn is defined as the negative natural average logarithm of the conditional probability that two sequences are similar for *m* points remain similar when the number of points is increased to *m*+1.

The calculation is as follows [32]:

Given a sequence of *N* measures, $U_N = \{u_1, u_2, ..., u_N\}$, where $x_m(i)$ y $x_m(j)$ are two vectors of U_N, both with a length of *m*. In the sequence $x_m(i)$ the vector starts at the u_i element of the series, and in the sequence $x_m(j)$, the vector starts at the u_j element. The $d[x_m(i),x_m(j)]$ distance between two vectors of $x_m(i)$ and $x_m(j)$ is defined as the maximum difference between their respective components. Consequently, the two vectors will be similar if $d[x_m(i),x_m(j)] < r$, where *r* is the parameter defining the criterion of similarity.

Let X_m be the set of all vectors with a length of *m* within U_N (i.e. $x_m(1)$, $x_m(2)$, …,$x_m(N-m+1)$). Given a vector $x_m(i)$, count the number of vectors $x_m(j)$, where $1 \le j \le N-m$, so $d[x_m(i),x_m(j)] < r$. Let B_i be the number of vectors. Thus, define the function $1 \le i \le N-m$ as:

$$B_m(i) = \frac{B_i}{N-m} \tag{1}$$

The probability that two vectors of *m* points match will be:

$$B_m = \frac{1}{N-m} \sum_{i=1}^{N-m} B_m(i) \tag{2}$$

The same procedure applies to *m+1* to obtain $A_m(i)$ and A_m

Let *SampEn* be defined as:

$$\mathrm{SampEn(m,r)} = \lim_{N\to\infty}\left\{-\ln\left[\frac{A_m}{B_m}\right]\right\} \tag{3}$$

As N is a finite number:

$$SampEn(m,r) = -ln\left[\frac{A_m}{B_m}\right] \tag{4}$$

The selection of *m* and *r* is key to estimating the SampEn. The default values are m=1 or m=2 (preferably m=2) and *r* between 10 and 25% of the standard deviation of the time series. This is due to the fact that r must be at least greater than the noise contaminating the signal; furthermore, it cannot have a very high value, as we would miss a great part of the signal information content.

As it is said above, SampEn measure the degree of regularity of a time series. However, SampEn is calculated on a single scale where the structure and organization in higher scales of the series is not considered [33] and, also, SampEn is extremely sensitive to parameter choices, especially for very short data sets, N≤200 [34].

To solve this, Madalena Costa introduced the concept of MSE [35]. Given a discrete time series, we generate new series which points are average values of k sequential elements of the original series (where k is the order of the scale) with no overlapping. Thus, for a time scale = 1, we have the original series; for a scale = 2 the new series will be composed of the average value of the elements taken in pairs on each scale. Finally, we calculate the SampEn for each of the new series generated accounting for values obtained against the scale factor. Therefore, entropy is highly dependent on the time scale.

The maximum value of the scale will depend on the length of the series. Previous studies [36] take a maximum scale factor of 20 with time series of a length of 20 000 data points. The MSE method uses the same statistical formula as SampEn. Therefore, the results lose consistency as the number of data points decreases.

Recently, a new algorithm called Composite Multiscale Entropy (CMSE) has been proposed [37]. This concept is introduced to overcome some difficulty of MSE related to a reduction of statistical reliability as a time scale factor is increased.

3. APPLICATIONS OF ENTROPY TO THE ANALYSIS OF HUMAN GAIT COMPLEXITY

Most studies on human gait complexity have been conducted on elderly subjects, patients with disorders affecting walking or after anterior cruciate ligament reconstruction.

Kurz and Stergiou [38] used the ApEn to elucidate whether the neurophysiological changes associated with aging have an impact on the nervous system operation which gives more stability to human gait. They studied elderly and young control groups walking on a treadmill

at a self-selected pace. Joint angles were calculated for the ankle, knee and hip. The yielded results supported the hypothesis that aging is associated with a loss in the ability of the neuromuscular system to move the lower limbs during walking. They hypothesized that such changes might yield inappropriate input from visual, vestibular, and somatosensory sensors (proprioceptive, cutaneous and articular). Therefore, an aged neuromuscular system provides inappropriate input that hinders correct walking. This might be the cause of the frequent falls in the elderly.

Later, Khandoker et al. [39] used the ApEn to study the risk for falls in the elderly by analyzing the variability of gait. On such purpose, they compared the minimum foot clearance (MFC) data during treadmill walking for 14 healthy elderly and 10 elderly participants with balance problems and a history of falls (falls risk). The study demonstrated that the ApEn of elderly subjects with a history of falls was significantly greater than that of the control group (0.18 and 0.13, respectively), which is indicative of an increase in irregularities and randomness in their gait patterns and an indication of loss of gait control mechanisms. Therefore, they concluded that gait variability –as analyzed by ApEn– could be useful for the early diagnosis of at-risk gait and for the detection and prevention of falls in elderly subjects.

In the field of pathology, Moraiti et al. [40] studied gait variability after anterior cruciate ligament reconstruction (ACL). They compared patients with ACL reconstruction, 2 years postoperatively, and 6 healthy control subjects walking on a treadmill at a self-selected pace while 2 minutes of continuous kinematic data were recorded with a 6-camera optoelectronic system. Stride-to-stride variability was calculated from the knee flexion/extension data. The results obtained showed that subjects undergoing ACL surgery presented ApEn values significantly greater than those of the control group (0.24 and 0.30, respectively).

Previously, the same authors [25,41] analyzed gait variability in subjects undergoing surgery for ACL tear. They hypothesized that the ACL deficient knee will exhibit more regular and less variable walking patterns than the contralateral intact knee. Ten subjects with unilateral deficiency walked on a treadmill at their self-selected speed, 20% faster and 20% slower, while kinematics was collected. They analyzed knee joint flexion-extension time series and they found that the ApEn values were significantly less in the injured leg as compared to the healthy leg. So, the ACL deficient knee exhibited more regular and less variable patterns than the contralateral intact knee.

In our opinion, the ApEn is not the most appropriate algorithm for the analysis of biological signals because of some conceptual errors. In fact, a study [41] yield consistent changes in ApEn values from a physiological perspective, but numerical values were characteristic of a period time series.

We only found three studies where either the SampEn or the MSE were used in the analysis of gait variability. Firstly, Tochigi et al. [42] applied the SampEn to analyze the cycle-to-cycle variability in leg acceleration signals during walking in elderly subjects and in adults with symptomatic knee osteoarthritis. The authors found that the study group showed a lower variability than the control group. Secondly, Costa et al. [43] used the MSE to analyze human gait variability by a pressure sensor installed in the footwear in different experimental situations (normal spontaneous walking at higher and lower speed, and walking controlled with a metronome). On such purpose, they asked a sample of ten healthy young subjects aged between 18-29 years to walk for one hour on the floor. The authors observed that

normal, spontaneous free walking has the highest complexity as compared to the other forms of walking.

Sánchez et al studied different forms of gait and running using the three entropy algorithms described above [44]. In particular, they analyzed time series of stride interval in ten healthy young subjects in three experimental situations: walking at a comfortable speed across the floor for 25 minutes in a form chosen by the subject (Situation A); walking on a treadmill at the same speed as in Situation A for 25 minutes (Situation B), and running on a treadmill at a 8 Km/h speed for 20 minutes (Situation C). The authors found (Table 1) that ApEn yielded values significantly lower than those obtained by the SampEn or MSE and no significant changes were observed in ApEn when the experimental situations were compared (A vs B and B vs C). The significant decrease in MSE (all scales) indicated that signal complexity was greater during spontaneous free walking than walking or running on a treadmill.

Table 1. ApEn, SampEn, MSE 2, MSE3 MSE4 and MSE5 values expressed as the mean and standard deviation of the time series for each of the experimental situations

	Situation A		**Situation B**		**Situation C**		**Significance level (p)**	
	MEAN	SD	MEAN	SD	MEAN	SD	A vs B	B vs C
ApEn	0.529	0.109	0.403	0.362	0.377	0.451	0.146	0.449
SampEn	1.666	0.094	1.559	0.182	1.306	0.245	0.070	*0.003*
MSE 2	1.607	0.050	1.517	0.064	1.482	0.082	*0.002*	*0.035*
MSE 3	1.497	0.082	1.390	0.074	1.316	0.081	*0.006*	*0.001*
MSE 4	1.408	0.098	1.305	0.059	1.247	0.068	*0.002*	*0.007*
MSE 5	1.363	0.110	1.253	0.074	1.158	0.063	*0.006*	*0.001*

Walking at a comfortable speed across the floor for 25 minutes in a form chosen by the subject (Situation A); walking on a treadmill at the same speed as in Situation A for 25 minutes (Situation B), and running on a treadmill at a 8 Km/h speed for 20 minutes (Situation C). Significance level $p < 0.05$. From (44) with permission.

4. CONCLUSION

In short, entropy –especially SampEn and MSE algorithms– might contribute new perspectives in the analysis of human gait and might be useful in the evaluation of pathological situations of the locomotor system. The MSE seems to be the most sensitive algorithm to changes in gait in healthy subjects.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Costa M, Peng C-K, Goldberger AL. Multiscale analysis of heart rate dynamics: entropy and time irreversibility measures. Cardiovasc Eng 2008;8(2):88-93.
2. Hausdorf JM, Peng CK, Ladin Z, Wei JY, Goldberger AL. Is walking a random walk? Evidence for long range correlations in the stride interval of human gait. J Appl Physiol. 1995;78:349-58.

3. Hausdorf JM, Purdon PL, Peng CK, Ladin Z, Wei JY, Goldberger AL. Fractal Dynamics of human gait: Stability of long range correlations in stride interval fluctuations. J Appl Physiol. 1996;80:1448-57.
4. Decker LM, Cignetti F, Stergiou N. Complexity and human gait. Rev Andal Med Deporte. 2010;3(1):2-12.
5. Inman VT, Ralston HJ, Todd F. Human Walking. Baltimore, MD: Williams & Wilkins; 1981.
6. Kadaba, MP, Ramakrishnan HK, Wootten ME, Gainey J, Gorton G, Cochran GVB. Repeatibility of kinematic, kinetic and electromyographic data in normal adult gait. J Orthop Res. 1989;7:849-860.
7. Pailhous J, Bonnard M. Steady-state fluctuations of human walking. Behav Brain Res. 1992;47:181-190.
8. Palta AE. Some characteristics of EMG patterns during locomotion: implications for locomotor control processes. J Motor Behav. 1985;17:443-461.
9. Winter DA. Kinematic and kinetic patterns in human gait: variability and compensating effects. Hum Mou Sci. 1984;3:51-76.
10. Gabell A, Nayak USL. The effect of age on variability in gait. J Gerontol 1984;39:662-666.
11. Guimares RM, Isaacs B. Characteristics of the gait in old people who fall. Int Rehab Med 1980; 2: 177-180.
12. Yamasaki M, Sasaki T, Torii M. Sex difference in the pattern pattern of lower limb movement during treadmill walking. Eur J Appl Physiol Occup Physiol. 1991;62:99-103.
13. Yamasaki M, Sasaki T, Tsuzki S, Torii M. Stereotyped pattern of lower limb movement during level and grade walking on treadmill. Ann Physiol Anthropol 1984;3:291-296.
14. Van Emmerik REA, Rosenstein MT, McDermott WJ, Hamill J. Nonlinear Dynamical Approaches to Human Movement. J App Biomech. 2004;20:396-420.
15. Webster KE, Wittwer JE, Feller JA. Validity of the GAITRite® walkway for the measurement of averaged and individual step parameters of gait. Gait posture. 2005;22:317-321.
16. Beauchet O, Hermann FR, Grandjean R, Dubost V, Allali G. Concurrent validity of SMTEC® footswitches system for the measurement of temporal gait parameters. Gait posture. 2008;27:16-159.
17. Davis R, Opunpuu S, Tyurski D, Gage J. A gait analysis data collection and reduction technique. Hum Mov Sci. 1991;10:575-587.
18. Stergiou N, Buzzi UH, Kurz MJ, Heidel J. Nonlinear Tools in Human Movement. In: Stergiou N, editor. Innovative Analyzes of Human Movement. Champaign, IL: Human Kinetics. 2004;63-90.
19. Stergiou N, editor. Innovate Analyzes of Human Movement. Analytical Tools for Human Movement Research. Champaign (Illinois): Human Kinetics Publ; 1999.
20. Peng CK, Mietus J, Hausdorff JM, Havlin S, Stanley HE, Goldberger AL. Long-range anticorrelations and non-Gaussian behavior of the heartbeat. Phys Rev Lett 1993;70:1343–1346.
21. Hausdorff JM, Purdon PL, Peng CK, Ladin Z, Wei JY, Goldberger AL. Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations. J Appl Physiol. 1996;80:1448-1457.
22. Hausdorff JM. Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride fluctuations of human walking. Hum Mov Sci. 2007;26:555-589.
23. Hausdorff JM, Peng CK, Ladin Z, Wei JY, Goldberger AL. Is walking a random walk? Evidence for long-range correlations in stride interval of human gait. J Appl Physiol. 1995;78:349-358.

24. Buzzi UH, Stergiou N, Kurz MJ, Hageman PA, Heidel J. Nonlinear dynamics indicates aging affects variability during gait. Clin Biomech. 2003;18(5):435-443.
25. Moraiti C, Stergiou N, Ristanis S, Georgoulis AD. ACL deficiency affects stride-to-stride variability as measured using nonlinear methodology. Knee Surg Sports Traumatol Arthros. 2007;15(12):1406-1413.
26. Pincus SM. Aproximate entropy as a measure of system complexity. Proc Natl Acad Sci. 1991;88:2297-2301.
27. Goldberger AL, Peng CK, Lipsitz, LA. What is physiologic complexity and how does it change with aging and disease? Neurobiol Aging. 2002;23(1):23-26.
28. Pincus SM, Goldberger AL. Physiological time-series analysis: what does regularity quantify?. Am J Physiol. 1994;266 (Heart Circ Physiol 35):H1643-H1656.
29. Pincus SM. Assessing serial irregularity and its implications for health. Ann NY Acad Sci. 2001;954:245-267.
30. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ Physiol. 2000;278(6):2039-2049.
31. Mäkikallo TH, Seppänen T, Niemelä M, Airaksinen KEJ, Tulppo M, Huikuri HV. Abnormalities in beat to beat complexity of heart rate dynamics in patients with a previous myocardial infarction. J Am Coll Cardiol. 1996;28:1005-1011.
32. Chen W, Zhuang J, Yu W, Wang Z. Measuring complexity using FuzzyEn, ApEn and SampEn. Medical Engineering and Physics 2009;31:61-68.
33. Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of biological signals. Phys Rev E Stat Nonlin Soft Matter Phys 2005;71:021906 Epub.
34. Yentes JM, Hunt N, Schmid KK, Kaipust JP, McGrath D, Stergiou N. The appropriate use of approximate entropy and sample entropy with short data sets. Ann Biomed Eng. 2013;41(2):349-365.
35. Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of complex physiologic time series. Phys Rev Lett. 2002;89:068102 Epub.
36. Costa M, Goldberger Al, Peng CK. Multiscale Entropy Analysis: a new measure of complexity loss in heart failure. J Electrocardiol. 2003;36 Suppl:39-40.
37. Wu SD, Wu CW, Lin SG, Wang CC, Lee KY. Time series analysis using Composite Multiscale Entropy. Entropy. 2013;15:1069-1084.
38. Kurz M, Stergiou N. The aging neuromuscular system expresses less certainty for selecting joint kinematics during gait in humans. Neurosci Lett. 2003;348(3):155-158.
39. Khandoker AH, Palaniswami R, Begg RK. A comparative study on approximate entropy measure and poincaré plot indexes of minimum foot clearance variability in the elderly during walking. J Neuroeng Rehabil. 2008;5:4.
40. Moraiti C, Stergiou N, Ristanis S, Vasiliadis HS, Patras K, Lee A, Georgoulis AD. The effect of anterior cruciate ligament reconstruction on stride-to-stride variability. Arthroscopy. 2009;25(7):742-749.
41. Georgoulis AD, Moraiti C, Ristanis S, Stergiou N. A novel approach to measure variability in the anterior cruciate ligament deficient knee during walking: the use of the approximate entropy in orthopaedics. J Clin Monit Comput. 2006;20(1):11-18.
42. Tochigi Y, Segal NA, Vaseenon T, Brown TD. Entropy analysis of tri-axial leg acceleration signal waveforms for measurement of decrease of physiological variability in human gait. J Orthop Res. 2012;30:879-904.
43. Costa M, Peng C-V, Goldberger AL, Hausdorff JM. Mustiscale entropy analysis of human gait dynamics. Physica A 2003;330:53-60.

44. Sánchez MD, De la Cruz B, Algaba J, Lagos MA, Naranjo J. Influencia del tapiz rodante sobre la variabilidad de la marcha en sujetos jóvenes y sanos. Rev Andal Med Deporte. 2012;5(2):53-56.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sciencedomain.org/review-history.php?iid=226&id=5&aid=1736