Wireless Sensor Network for Low-Complexity Entropy Determination of Human Gait

Mihaela I. Chidean[1], Giancarlo Pastor[1,2], Eduardo Morgado[1], Julio Ramiro-Bargueño[1], and Antonio J. Caamaño[1]
[1]Signal Theory and Communications Departament, Rey Juan Carlos University, Fuenlabrada, Spain
{mihaela.chidean, eduardo.morgado, julio.ramiro, antonio.caamano}@urjc.es
[2]Department of Communications and Networking, Aalto University, Espoo, Finland
giancarlo.pastor@aalto.fi

***Abstract*—In this work we present a Wireless Sensor Network (WSN) system designed for the on-board determination of human gait entropy. The usage of nonlinear entropy-based metrics has proven to be a useful tool for analyzing the complexity of biological systems. The final goal of entropy calculation in this type of biological system is to identify possible causes of future injuries (in order to improve aging) and the early injury detection (ideal for elite athletes). Existing systems for human gait analysis are limited to traditional data gathering, e.g. continuous measurement and wireless transmission to a Data Fusion Center (DFC), due to the computational burden of entropy calculation. In addition, actual systems are likely to interfere the natural movement due to their cumbersome nature. The WSN presented here uses four sensor nodes, located in both ankles and hip sides, and are equipped with triaxial accelerometers. We propose the use of low-complexity algorithms in order to perform on-board entropy determination prior to wireless transmission. The proposed system can be used to reliably determine long-term human gait entropy.**

***Index Terms*—Wireless Sensor Networks, Body Area Networks, Entropy, Wearable Sensors, Human Gait.**

I. Introduction

Usually, biological systems of interest are too complex to be finely characterized from one or two-dimensional signals. To this end, the usage of nonlinear metrics as simplified markers of complexity (or lack thereof) in biological signals dates from a few decades back (ECGs, EEGs, ...). They have been mainly used to tackle the nature of self-organized complexity that arises from competitive interactions. The main idea behind their use is that physiological measures, both complex and with many characteristic times present, tend to turn into a collection of a few regular patterns as disease presents itself.

In the last decade, the analysis of human gait dynamics through nonlinear metrics related to entropy has gained momentum from the seminal work of Costa et al. [1]. Up to that date, previous studies using fractal measures [2] indicated that fluctuations of human gait cycle of free walking healthy individuals do not represent independent or uncorrelated random noise. A power law characterizing the appearance of long-range correlations emerges in such analysis of the stride interval. Thus, the duration of such interval cannot be seen as a Markov process in which it depends on the immediate anterior stride duration, but also on distant intervals. Therefore, complex dynamics are present in the human gait though the characterization of its meaning is still elusive.

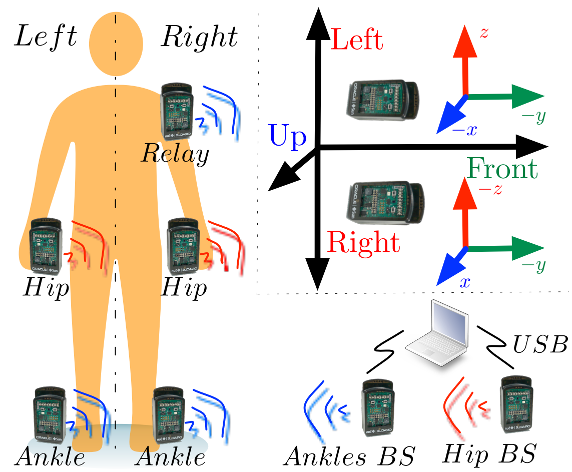


Fig. 1. Wireless Sensor Network system setting.

The use of entropy in the analysis of human gait is sensible from the point of view that characteristic time intervals in a series are closely related to the entropy of the system that originated it [3]. Apart from the Detrended Fluctuation Analysis (DFA) work mentioned above [2], entropy-related metrics have been mainly used in human gait time-series. Approximate Entropy (ApEn) [4] and Sample Entropy (SampEn) [5] are extensively used measures in human gait analysis. They are both based on the Kolmogorov complexity and, therefore, they amount to finding the number of different consecutive patterns (of increasing length) and their abundance, in the stride interval time-series. This poses the first difficulty as it involves solving a NP-complete problem with exponential computational complexity [6]. Furthermore, as identical patterns are highly unlikely to arise, both measures are highly dependent on the relaxation made in the definition of patterns similarity. Further refinements led to the development of the Multiscale Entropy [1] that has been able to explicitly capture the long-range correlations of the stride duration.

Up to now, the complexity of human gait has been mainly limited to uniaxial (in the direction of advance) accelerometric measurements [2], [1]. However, static triaxial measurements in balance boards have shown that equilibrium complexity is a critical marker of healthiness [7]. As equilibrium plays a crucial role in the determination of the human gait, it seems reasonable to apply the entropy analysis of human gait, not only in the direction of advance but in the three spatial dimensions. Previous attempts at estimating the complexity of stride intervals in both legs and with triaxial accelerometers were limited to the gathering of the time-series in both ankles, and then by extrapolation of the measured data to infer behaviour of hip movement [8].

In this work we propose the use of a Wireless Sensor Network (WSN) setup in order to measure the stride interval time-series of running individuals with triaxial accelerometers in both hip and ankles. The proposed setup is done in order to avoid traditional interference of the movement due to the cumbersome nature of the measuring apparatus. The full accelerometric time-series is continuously transmitted to a Data Fusion Center (DFC) and the processing of the twelve time-series is performed offline. We observe that, although the WSN is appropriate and accurate for metronomic running (treadmill running), it is inadequate for free running (athletics track running). This is shown to be due to the unreliability of the wireless channel, with far worse characteristics in the latter case. Therefore, it is desirable to *compress and protect* the time-series in the node (before transmission) or to *process* it in order to obtain the stride duration time-series, much less cumbersome to transmit. Ideally, it should be possible to further process the stride duration time-series to obtain the entropy measure of choice. However, the complexity of such calculation prevents it from being performed in "lightweight" processors such as those in the WSN nodes.

The main result of this work is the use of low-complexity entropy measures [9] in the analysis of the human gait biological systems. These entropies are computed using the strides interval time-series and can be calculated in real time (RT) in the WSN nodes as the subject is running, thus allowing their meaningful estimation in free running conditions. Comparison in terms of computational complexity and performance is done with SampEn for actual time-series obtained from a single individual for a treadmill running setting.

We present the setting of the WSN network for human gait complexity estimation in Section II and the definitions of the used entropy measures in Section III. We then explain the experiments and the measurements obtained and the entropy analysis in Section IV. Finally we make some concluding remarks in Section V.

II. System Setting

In this section we summarize the system requirements for the WSN and present the detailed network configuration used in the experiments. The configuration used satisfies the requirements and gives the chance of circumventing the cases where the radio channel has poor characteristics.

A. Requirements

First, as the goal is the analysis of the human gait, the main condition is that the WSN has to gather accelerometric data with a sufficiently high sampling frequency ($\geqslant$ 50 Hz), in order to capture all the signal variations.

The WSN has to consist of 4 sensor nodes, each located in a specific human body position. There are two sensor nodes for each of the sinister and dexter (left and right) portions, positioned at ankle and hip height. The approximate sensors node positions are plotted in the left side of Figure 1 using as reference the body silhouette, the accelerometer axis orientation is in the upper right side of the same figure.

Once measured, data packets are formed and sent towards the DFC. In addition, the network topology is not set in advance, however the DFC has to be located inside the coverage range of at least one of the sensor nodes. Naturally, the complete WSN has to cause minimal inconvenience to the individual.

B. Network configuration

The devices selected for the WSN deployment are the SunSPOT[1] mote and the SunSPOT Base Station (BS) for the sensor node and the DFC, respectively. As WSN nodes in general, the SunSPOT motes have limited memory long-term measurement storage on board and an offline processing is not possible, so a sense-and-transmit policy is used. Moreover, the only way to recover the data stored in a mote is through the wireless channel, so a error-free transmission of the acceleration data is highly unlikely.

Due to the limited bandwidth, and to enable the maximum sample frequency, two sub-networks are considered: 1) for the ankles and 2) for the hip. Each sub-network uses a different wireless sub-channel, sufficiently separated in the spectrum so that there is no interference between the two sub-networks. In Figure 1 the ankles and the hip sub-networks are indicated with blue and red colors, respectively.

An additional node, acting as relay for the ankles sub-network, is proposed for free running experiments performed in the athletics track. This decision was made during the prototyping stage of the work as the ankle sub-network radio channel showed very poor propagation conditions due to the node height. By situating the relay node at shoulder height, we ensure the best possible wireless channel for the ankle sub-network.

Regarding the transmissions, the client-server model is chosen, being the sensor nodes and the BS the clients and the server, respectively. The communication between the sensor nodes and the BS is unicast, to take advantage of the ACK/retry mechanism of the radiogram protocol available for the SunSPOT devices.

In order to enable the run-time manipulation of the BS's radio properties, the SunSPOT BS mode is set to "dedicated". The transmit power is set to the maximum power available for SunSPOT devices, namely 0 dBm, for coverage throughout the athletics track. The choice made as to the transmit power determines the radio channel that are available. For example, for radio channel 26 the maximum power is -3 dBm due to regulations, so it cannot be used in our WSN.

[1] http://www.sunspotworld.com/

TABLE I
TRANSMISSION CONFIGURATION FOR THE SUNSPOT DEVICES.

	Ankles sub-network	Hip sub-network
Transmission mode	Unicast	Unicast
Wireless channel	25	11
PANID	4	3
Transmit power [dBm]	0	0
Samples / Packet	3	3

The samples obtained from the triaxial accelerometer part of the SunSPOT devices consist of three values, one for each axis. The original data type is DOUBLE, meaning 64-bit. However, to gain efficiency in transmission and, since an adequate resolution should be kept, data is converted to FLOAT, a 32-bit type. Therefore, for each sample an amount of 20 Bytes have to be transmitted; a 64-bit LONG value is considered for the timestamp.

An additional extent for transmission efficiency is the data encapsulation, meaning that each packet contains three samples, meaning that has 60 Bytes size. With this size we ensure no fragmentation and that the whole packet is transmitted in the interval between two consecutive timestamps. Table I summarizes the transmission configuration.

III. LOW-COMPLEXITY SPARSITY-BASED ENTROPY MEASURES

In a recent work by some of the authors, the proposal of two low-complexity measures of entropy have been made [9]. The authors present the properties that sensible sparsity measures have to attain and what are their counterparts with respect to entropy measures. They propose two measures that satisfy some or all of the criteria. These measures are

$$h_{\mathsf{pmf}}(x) = \left\langle \frac{\mathbf{1}_b}{||\mathbf{1}_b||_1}, \frac{p(x)}{||p(x)||_\infty} \right\rangle \tag{1}$$

and

$$h_{\mathsf{ipmf}}(x) = \left\langle \frac{K_N}{||\mathbf{1}_N||_1}, \frac{x^\uparrow}{||x||_\infty} \right\rangle \tag{2}$$

where $x = (x_i) \in \mathbb{R}^N$ is a r.v. signal with probability mass function (pmf) $p(x)$, $\mathbf{1}_N$ is a vector of all ones with length N, $||\cdot||_p$ denotes the p-norm of a given vector and b is the number of bins used in the histogram method. In Eq. (2), K_N is related to $\Phi(x) = \exp(-n^2/2), n \in \{-b/2, \dots, b/2\}$ as the inverse of sorted Φ (increasing order). Also $x^\uparrow$ is the vector of sorted individual realizations of x s.t. $x_1^\uparrow \leqslant \cdots \leqslant x_N^\uparrow$.

What these two simple measures aim to evaluate is the similarity of either the pmf of x or x itself to a given pmf or corresponding vector r.v.. In $h_{\mathsf{pmf}}(x)$ it is the similarity of the pmf of x with the pmf of the most uncertain signal i.e., the pmf of a uniform r.v. is measured. In the case of $h_{\mathsf{ipmf}}(x)$, it is the similarity of sorted x with a sorted realization of a standard normal r.v. of the same length, which is the natural reference signal for entropy.

The complexity of a signal determines its compressibility under a given basis transformation. In [9] Pastor et. al have established that the sparsity of a signal under such transformation can be related to the two measures presented above. The first h_{pmf} is directly related to the changes in the occurrence probability with the uncertainty of the signal. As it may be of use to work directly with signal x instead of its pmf, the second measure h_{ipmf} is defined. It is directly related to the sparsity of the signal but with a relaxation of the compared signal as Φ is used as an approximation to the delta function δ, the sparsest of the signals.

The computational complexity of these measures is extremely low (compared to the exponential complexity of ApEn or SampEn computation with N increasing) as Eq. (1) can be rewritten in a much simpler form as $\langle \mathbf{1}_N, p(x) \rangle = N$, $||\mathbf{1}_b||_1 = b$ and $||p(x)||_\infty = n_{\mathsf{max}}$:

$$h_{\mathsf{pmf}} = \frac{N}{b \cdot n_{\mathsf{max}}} \tag{3}$$

where n_{max} is just the number of elements in the largest bin in the histogram of x. The most involved computation in the aforementioned measures is the sorting of the x and Φ histograms. Therefore, these measures are ideally suited for on-board processing of the complexity of physiological signals in lightweight wireless sensor nodes.

IV. EXPERIMENTS AND RESULTS

This section is dedicated to the analysis of the accelerometric data collected in the experiments performed. Firstly, we compare time-series gathered from both treadmill and athletics track experiments in order to show the significant influence of the wireless channel in this WSN. Next, we focus on the periodical signal shape and analyze it so that the stride interval time-series presence becomes clear. Following, the entropy metrics proposed in this work are applied to the stride interval time-series and compared with the SampEn. Finally, a comparison between the traditional data gathering and the on-board processing proposed in this work is made, in order to show the tradeoffs for the nodes and network resources in both cases.

A. Treadmill vs. athletics track experiments

The WSN proposed in this work allows accelerometric data gathering for both treadmill and athletics track setting, from the radio coverage and node memory point of view. However, the major drawback for the athletics track setting is the wireless channel due to i.e. the obstacles in the form of athletics equipment present in the track (e.g. high jump and pole vault mattress metal covers) that are bound to further degrade the quality of the signal.

In Figure 2 two accelerometric time-series collected for the same node location and accelerometer axis and for both metronomic and free running experiments are plotted. It can be clearly observed the effect of the channel on the signal, in the form of packet losses, thus causing irrecoverable damage to the waveform.

In this work, in order to establish a baseline for the entropy metrics proposed, we further analyze only the treadmill data. Thereby, most of the damage caused by the channel is avoided.

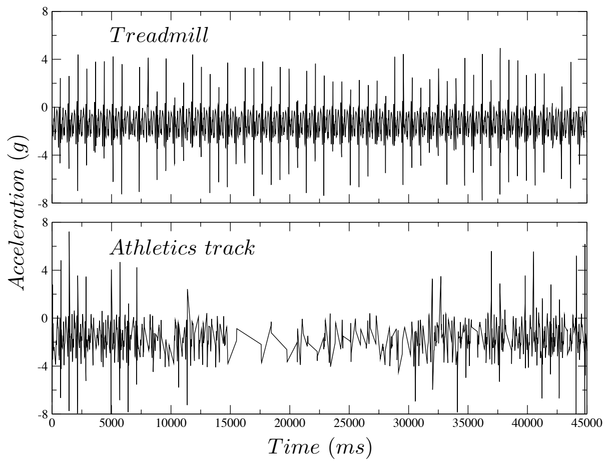


Fig. 2. Acceleration data gathered by the WSN for the X axis of the left ankle during experiments for the treadmill (upper graph) and the athletics track (lower graph). Data loss is evident in the latter case.

B. Stride interval time-series

Once the dataset is selected, the next step is stride detection in order to compute the stride interval time-series. In Figure 3 temporal slices of the accelerometric time-series for all WSN node locations and axis are plotted.

In order to obtain the stride interval time-series, the data peaks were detected using a thresholding method applied to windowed data. In Figure 3 the peaks detected for each of the 12 accelerometric time-series are with red dots. The wanted time-series are the intervals between each two consecutive peaks.

C. Entropy results

In this section we analyze the results obtained by computing h_{pmf} and h_{ipmf} vs. a traditional entropy measure of physiological complexity such as SampEn with measures from a single individual in a treadmill running for approximately a 10 minute period. In Fig. 4 we can observe the online computation of h_{pmf} (dashed blue line) and h_{ipmf} (dash-dotted red line) and the offline computation (as the whole time series is necessary) of SampEn (solid black) with parameter $m = 2$ for the four nodes and three axes. We can observe the evolution of the proposed measures as time advances. If we focus on axes X (up-and-down movement, Fig. 4.a) and Z (outwards left-and-right movement, Fig. 4.c), we may observe similar trends for the wireless node hip pairs as well as those for the ankles. Similar comparison with the SampEn won't reveal the same trends in pairs. We observe that SampEn may be inconclusive for comparison in the same axes and nodes. However if we focus on axis Y (front-and-back movement, Fig. 4.b), we may see an *inversion* in the behavior of trends with respect to wireless node pairs. Left hip h_{pmf} and h_{ipmf} exhibit the same trend as the right ankle and, accordingly, left ankle's complexity measure trends to those in the right hip. Close examination of the Y time series by an experienced physician revealed evidences of hip rotation to compensate for a slight difference in leg lengths as the data gathering progressed, i.e., as the individual becomes stressed and tired. Hip rotation may be cause for future injuries, mainly located in calf muscles. In comparison, examination of SampEn revealed none of these symptoms. This former fact is unsurprising as it is widely accepted that, for SampEn to be a meaningful measure of physiological complexity, time series of at least $N = 2000$ samples are needed.

Therefore, it is shown that h_{pmf} and h_{ipmf} are able to reveal physiological complexity decrease (or increase) in RT and for short-time series. In comparison, computationally cumbersome measures of physiological complexity such as SampEn are unable to be used in RT, for short time intervals for this purpose.

TABLE II
TRADITIONAL DATA GATHERING VS. ON-BOARD PROCESSING REQUIREMENTS.

	On-board processing	Data gathering
Memory	$O(S)$	$O(M)$
Packet size	$O(M)$	$O(M)$
Packet rate	$O(\frac{1}{MK})$	$O(\frac{1}{MT_s})$
Packets for N measurements	$O(\frac{N_m}{MK})$	$O(\frac{N_m}{MT_s})$

D. On-board data processing vs. traditional data gathering

In order to show the benefits of the on-board data processing, in this section we compare both procedures considering, for example, the memory required or the amount of data packets related to the same measurements sent towards the BS.

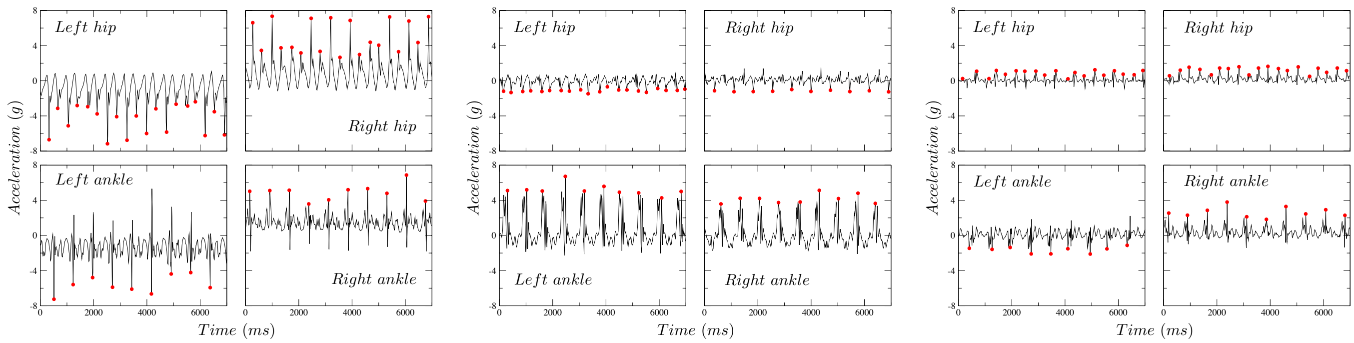
Let F_s be the sample frequency, $T_s = \frac{1}{F_s}$ and, N_m be the total amount of measurement instants during entire experiment. The data encapsulation is considered by the parameter M, which indicates how many measurement instants are included in the same packet, each measurement being identified by a timestamp. Together with the timestamps, the corresponding samples and the on-board computed entropies are transmitted, each for the respective procedure. For the on-board processing, S is the amount of states in the system (histogram bins) and K is the sliding window size. The values used in the comparison are: $F_s = 50$ Hz, $T_s = 20$ ms, $N_m = 90.000$ measurements, $M = 3$ timestamps/packet, $S = 100$ states, $K = 10.000$ measurements.

The parameters taken into consideration for the comparison are: 1) memory storage, 2) packet size, 3) packet transmission rate and, 4) total amount of packets sent for N measurements instants.

As an exact calculation of these parameters is complicated due to the impact of variables such as auxiliary data used in the SunSPOT application, an approximation is shown in Table II.

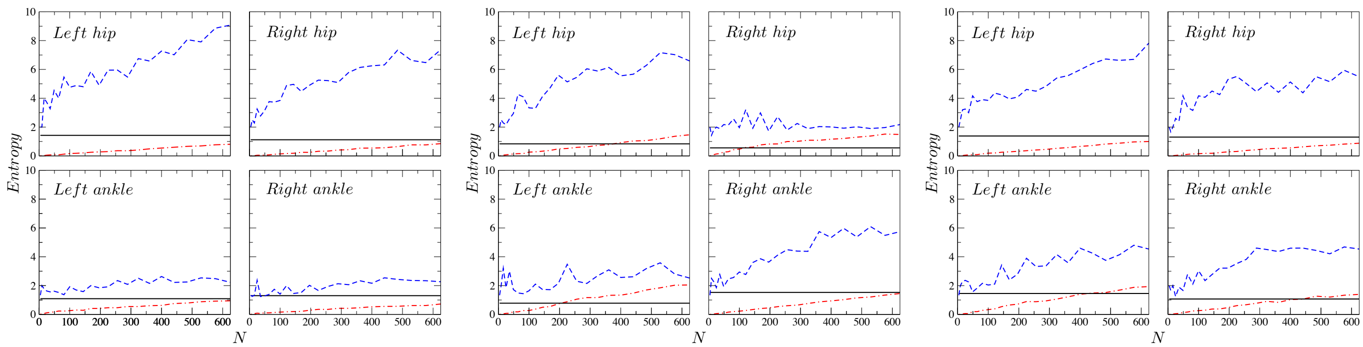
Using the previously defined values, it can be seen that on-board processing uses about two orders of magnitude more memory than a raw storage-and-forward strategy. On the other hand, the packet rate and the total amount of packets sent is lessened in about six orders of magnitude, with the resulting reduction in accesses to the wireless channel, one of the major sources of energy consumption in WSN.

In conclusion, we consider that the benefits of the on-board processing overcome its disadvantages in terms of energy efficiency. Undergoing research is directed toward a WSN



(a) X axis (b) Y axis (c) Z axis

Fig. 3. Acceleration data gathered by the WSN for all positions and axis in the treadmill experiment and their respective peaks (red dots) identifying the strides, detected in the post processing.



(a) X axis (b) Y axis (c) Z axis

Fig. 4. Entropies calculated using the stride duration time-series for the treadmill experiment for all positions and axis. Black line is for SampEn, computed with constant $N = 625$ values. Dashed blue line is for h_{pmf} and dash-dotted red line is for h_{ipmf}.

deployment using the entropy calculation proposed in this work as on-board efficient processing for similar network configuration as in Figure 1.

V. Conclusions

In this work we have presented a WSN system designed for the determination of human gait entropy. In order to make the system useful in free running conditions the calculation of the complexity of human gait with entropy metrics is performed on-board as opposed to traditional data gathering. To avoid computational burdens a new set of low-complexity entropy metrics have been proposed. We have shown that these metrics are able to capture key characteristics of human gait complexity. We have been able to identify possible causes for future injuries (i.e. hip rotation) with short length time-series of stride interval. We have also shown that the on-board proposed method overcomes the limitations of the wireless channel as a much lower transmission rate is needed. The present system will be shortly used to carry out extensive measurements on both healthy and injured individuals for further analysis.

Acknowledgments

The authors are thankful to Dr. Inmaculada Mora-Jiménez for valuable suggestions during the course of this work and to Dr. José Naranjo Orellana for his help in defining physiologically valuable points of measure.

This work was supported in part by the Spanish Government under Projects TEC2009-12098 and TEC2010-19263 and PhD Research Grant AP2012-2981.

References

[1] M. Costa, C. K. Peng, A. L Goldberger, and J. M. Hausdorff, "Multi-scale entropy analysis of human gait dynamics," *Physica A: Statistical Mechanics and its Applications*, vol. 330, no. 1-2, pp. 53–60, Dec. 2003.

[2] J. M. Hausdorff, C. K. Peng, Z. Ladin, J. Y. Wei, and A. L. Goldberger, "Is walking a random walk? Evidence for long-range correlations in stride interval of human gait," *Journal of Applied Physiology*, vol. 78, no. 1, pp. 349–358, 1995.

[3] S.-k. Ma, "Calculation of entropy from data of motion," *Journal of Statistical Physics*, vol. 26, pp. 221–240, Dec. 1981.

[4] S. M. Pincus, "Approximate entropy as a measure of system complexity." *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, Mar. 1991.

[5] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[6] T. Batu, S. Dasgupta, R. Kumar, and R. Rubinfeld, "The Complexity of Approximating the Entropy," *SIAM Journal on Computing*, vol. 35, no. 1, pp. 132–150, Jan. 2005.

[7] A. Priplata, J. Niemi, M. Salen, J. Harry, L. Lipsitz, and J. Collins, "Noise-Enhanced Human Balance Control," *Physical Review Letters*, vol. 89, no. 23, p. 238101, Nov. 2002.

[8] Y. Tochigi, N. A. Segal, T. Vaseenon, and T. D. Brown, "Entropy analysis of tri-axial leg acceleration signal waveforms for measurement of decrease of physiological variability in human gait," *Journal of Orthopaedic Research*, vol. 30, no. 6, pp. 897–904, Dec. 2011.

[9] G. Pastor, I. Mora-Jimenez, R. Jantti, and A. J. Caamano, "Sparsity-Based criteria for entropy measures," in *The Tenth International Symposium on Wireless Communication Systems 2013 (ISWCS 2013)*, Ilmenau, Germany, Aug. 2013.