33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

A point process approach for analyzing gait variability dynamics

Robert J. Ellis, Luca Citi, *Member, IEEE* and Riccardo Barbieri*, *Senior Member, IEEE*

***Abstract*—We present a novel statistical paradigm for modeling and analysis of gait variability which captures the natural point process structure of gait intervals and allows for definition of new measures instantaneous mean and standard deviation. We validate our model using two existing data sets from physionet.org. Results show an excellent model fit and yield insights into the underlying statistical structure behind human gait. Statistical analyses further corroborate previous findings of increased variability in gait at different speeds, both self-paced and metronome-paced, and reveal a significant increase in gait variability in Parkinson's subjects, as compared to young and elderly healthy subjects. These results indicate the validity of a point process approach to the analysis of gait, and the potential utility of incorporating instantaneous measures of gait into diagnostic or patient monitoring applications.**

I. Introduction

THE temporal dynamics of human gait, like other complex physiological signals (e.g., cardiac and respiratory chronotropism), is characterized by moment-to-moment fluctuations [1]. Variability is small (roughly 2% around the mean) in healthy adults, but increases dramatically in patients with Parkinson's disease (PD) and Huntington's disease [2]. Increased gait variability, particularly in the time domain (i.e., between successive heel strikes), has a serious effect on patient well-being: as PD progresses, up to 50% of PD patients will experience multiple falls per year [3]. A number of prospective studies have directly linked increased stride time variability with increased fall risk in PD (reviewed in [4]).

Over the years, a number of time domain, frequency domain, time–frequency analyses, and non-linear methods of gait have been utilized to explore the temporal aspects of gait (for reviews, see [1], [4-6]). None of these methods, however, has focused on a highly desirable and clinically relevant parameter: a real-time, instantaneous measure of stride variability. Such a measure could provide a window into the underlying dynamics during brief periods of impairment in PD, such as turning, entering a doorway, or transitioning over different floor surfaces (e.g., [7]).

Furthermore, all previous analyses have treated stride time variability data as a continuous-valued, stationary signal that is sampled as a time series. However, the time between successive strides is more accurately characterized as a point process: that is, a sequence of discrete occurrences in continuous time [8]. The recent application of point-process statistics to heartbeat dynamics [9-11] is particularly relevant to the present topic, as heartbeats (i.e., successive electrical R-waves) and footfalls (i.e., heel strikes) exist in the same time scale (e.g., 60–90 per minute). Specifically, a history-dependent inverse Gaussian (HDIG) point-process model of time interval dynamics yields precise probabilistic definitions of interval variability that can be updated at any desired time resolution [9]. Model goodness-of-fit is also provided as a standard component of the analysis framework, via a Kolmogorov–Smirnov (KS) test derived from the time-rescaling theorem [12].

The present paper serves as a proof-of-concept illustration of the application of a point-process analysis framework to stride time data using two sample datasets from PhysioNet [13]. The first dataset [14], from 5 healthy young adults (age 18–29), comprised self-paced and metronome-cued walks (~1700 strides) at each subject's natural walking tempo. The second dataset [15] comprised a single walking excerpt (~300 strides) from 15 subjects: five healthy young (age 23–29), five healthy elderly (age 60–77), and five PD patients (age 71–77). In both datasets, gait data (i.e., stride time intervals) was recorded via an ultra-thin force-sensitive switch placed in the insole of the right shoe while subjects walked around a large oval track (>160 m), and logged via a recorder strapped to the right ankle.

II. Methods

A. Gait Interval Probability Model

Suppose we are given a set of gait events (i.e., successive heel strike onsets) $\{g_j\}_{j=1}^{J}$. Let $GG_j = g_j - g_{j-1} > 0$ denote the jth gait interval (i.e., stride time interval), or equivalently, the waiting time until the next gait event. By treating the gait intervals as discrete events, we may develop a probabilistic point process model in continuous-time [9-11].

Assuming history dependence, we assume that the waiting time $t - g_t$ until the next gait event follows a HDIG model, where for any time $t > g_t$,:

$$p(t-g_t) = \left(\frac{\theta}{2\pi(t-g_t)^3}\right)^{\frac{1}{2}} \exp\left(-\frac{\theta(t-g_t-\mu_t)^2}{2\mu_t^2(t-g_t)}\right), \quad (1)$$

Manuscript received 15 April 2011. This paper was presented in part at the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society. R. J. Ellis is supported by National Institutes of Health (NIH) under grants R01DC009823 and R01DC008796 to Gottfried Schlaug. R. B. and L.C. are supported by NIH grant R01HL084502 to R. B. *Asterisk indicates corresponding author.*

R. J. Ellis is with the Music, Stroke Recovery, and Neuroimaging Laboratory in the Department of Neurology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02215 USA (phone: 617-632-8978; fax: 617-632-8920; e-mail: rellis@bidmc.harvard.edu).

L. Citi is with the Neuroscience Statistics Research Laboratory Massachusetts Institute of Technology, Cambridge, MA 02139 USA, and the Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114 USA (e-mail: lciti@neurostat.mit.edu).

*R. Barbieri is with the Neuroscience Statistics Research Laboratory Massachusetts Institute of Technology, Cambridge, MA 02139 USA, and the Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114 USA (e-mail: barbieri@neurostat.mgh.harvard.edu).

Where g_t denotes the previous gait event occurred before time t, $\theta > 0$ denotes the shape parameter, and μ_t denotes the instantaneous stride mean that can be modeled as a generic function of the past (finite) gait events $\mu_t = f\left(GG_{t-1}, GG_{t-2}, \ldots, GG_{t-h}\right)$ where GG_{t-j} denotes the previous jth gait interval (or stride) occurred prior to the present time t. The history dependence is defined by expressing the instantaneous mean $\mu_{GG}(t)$ as a linear combination of present and past gait intervals (in terms of an AR model), i.e., function f is linear. Here, we propose to include the nonlinear terms of past gait intervals in an attempt to improve the model fit. Specifically, the instantaneous mean μ_{GG} and standard deviation σ_{GG} are defined as

$$\mu_t \equiv \mu_{GG}(t) = a_0 + \sum_{i=1}^{p} a_i GG_{t-i} \tag{2}$$

$$\sigma_{GG}(t) = \left(\frac{\mu_{GG}(t)^3}{\theta} \right)^{1/2}. \tag{3}$$

Given the proposed parametric model, the nonlinear indices of the gait interval (i.e., stride time, ST) and the gait variability (i.e., stride time variability, STV) will be defined as a time-varying function of the parameters $\boldsymbol{\theta} = \left[a_0, a_1, \ldots, a_p, \theta \right]$.

B. Instantaneous Indices of Rate and Rate Variability

Cadence, or stride rate (SR), may be defined as the reciprocal of the gait intervals. For *GG* measured in seconds, $r = c\left(t - g_t\right)^{-1}$ (where c = 60 s/min) is a physiological measurement in strides per minute (spm). By the *change-of-variables* formula [16], the SR probability $p(r) = p\left(c\left(t - u_t\right)^{-1}\right)$ is given by

$$p(r) = \left| \frac{dt}{dr} \right| p(t), \tag{4}$$

and the mean and the standard deviation of stride rate r (i.e., stride rate variability, SRV) can be derived. Essentially, the instantaneous indices of SR and SRV are characterized by the mean μ_{SR} and standard deviation σ_{SR}, respectively:

$$\mu_{SR}(t) = c\left(\frac{1}{\mu_{GG}(t)} + \frac{1}{\theta(t)} \right), \tag{5}$$

$$\sigma_{SR}(t) = c\left[\frac{2\mu_{GG}(t) + \theta(t)}{\mu_{GG}(t)\theta^2(t)} \right]^{\frac{1}{2}}. \tag{6}$$

C. Local Maximum Likelihood Estimation

It is known from point process theory [9], [10], [12] that the *conditional intensity function* $\lambda(t)$ is related to the inter-event probability $p(t)$ by a one-to-one transformation:

$$\lambda(t) = \frac{p(t)}{1 - \int_{u_t}^{t} p(\tau)\, d\tau}. \tag{7}$$

A local maximum likelihood method [17] is used to estimate the unknown time-varying parameter set $\boldsymbol{\theta}$. In estimating $\boldsymbol{\theta}$ at time t, we take a local likelihood interval $(t-l,\ t]$, where l is the length of the local likelihood observation interval. Within $(t-l,\ t]$, we may observe n pulses, $t-l < g_1 < g_2 < \ldots < g_n \le t$. Then, we consider the local joint probability density of $g_{t-l:t}$, where $g_{t-l:t} = \{ g_1, \ldots, g_n \}$. The log likelihood of the joint probability density is given by

$$\log p(g_{t-l:t}) = \sum_{j=2}^{n} w(t - g_j) \log p(g_j - g_{j-1}) + w(t - g_j) \log \int_{t-g_n}^{\infty} p(\upsilon) d\upsilon, \tag{8}$$

where $w(t - g_j) = \alpha^{t-gj}$, $0 < \alpha < 1$, is a weighting function for the local likelihood estimation. The weighting time constant α governs the degree of influence of a previous event observation g_j on the local likelihood at time t. The second term of (8) represents the likelihood of partially observed interval since the last observed pulse g_n (right censoring). To maximize the local log likelihood in (8) we use a Newton–Raphson method, and obtain the local maximum likelihood estimate of $\boldsymbol{\theta}$. Of note, the time increment Δ for computing the next $\boldsymbol{\theta}$ from t to $t+\Delta$ can be chosen arbitrarily small, thus yielding instantaneous estimates of mean (ST, SR), and standard deviation (STV, SRV).

Goodness-of-fit of the model can be performed based on the *Kolmogorov–Smirnov* (KS) test and the time-rescaling theorem [12]. The KS distance, defined as the maximum distance between the KS plot and the 45° line, is used to measure the lack-of-fit between the model and the data. We also compute the autocorrelation function of the transformed quantiles to check independence of the transformed *GG* intervals [9].

III. RESULTS

The results below illustrate the following features of our point-process model: (1) the ability to detect an underlying probability structure (i.e., history dependence) during self-paced walks that is absent during metronome-cued walks, (2) the provision of instantaneous measures of stride time, stride time variability, stride rate, and stride rate variability, and (3) the automatic detection of outliers in a stride series.

A. Analysis of Probability Structure

Figure 1 presents the results of this test using data from a single subject from Dataset 1 under self-paced (a.) versus metronome-paced (b.) walking; specifically, stride series data (points) and the first moment of the inverse Gaussian distribution (trace) for the two stride events series (a.1, b.1), KS goodness-of-fit plots comparing order 1 (a.2, b.2) versus order 9 (a.3, b.3) models, and correlation functions of time-rescaled *GG* intervals, z_j, from the HDIG model fit for lags 1–60 (order 1: a.4, b.4; order 9: a.5, b.5). These results demonstrate the presence of correlation in the self-paced walking condition (order 1) that disappears under the metronome walking condition (order 1), and under both walking conditions when a higher model order is used (order 9). The KS plots confirm goodness-of-fit for the higher model order. These findings illustrate the ability of our point

process model to capture the history-dependent structure of gait under different walking conditions [18], [19].

B. Assessment of Instantaneous Measures

Figure 2 presents instantaneous ST, STV, SR, and SRV (4.) from three exemplar subjects from Dataset 2: young (a.), healthy elderly (b.), and PD (c.). The instantaneous series, in particular the variability series, are able to track changes at high time resolution, revealing new dynamic trends. For example, note the evident presence of an oscillation with a ~100-s period, particularly marked in the Young and PD subject. Trends such as these may be speculatively associated with feedback control system compensations. Note that these variability changes are independent from large changes in mean gait interval due to the ability of our algorithm to discard outliers, as we discuss below.

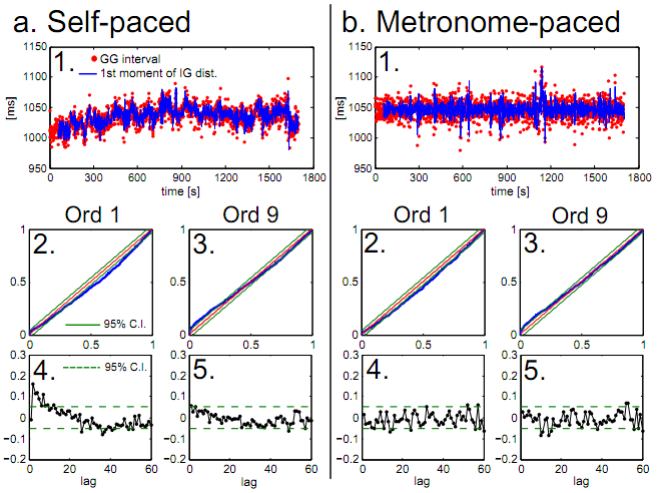

Figure 1. Self-paced versus metronome-paced walks in a single healthy subject. KS plots and correlation plots evaluate model properties with order 1 versus order 9.

C. Detection of Outliers

Figure 2d presents ST and STV data from the PD patient in its original form. The clear presence of outlying strides (i.e., > 1200 ms) drastically affects any estimate of mean or variability, including instantaneous measures. (Here, we mean outliers due to experimental error than gait events that become systematically more or less variable as a function of walking conditions or task demands.) The ability to automatically detect and remove outliers, based on the likelihood equation defined in (8), permits increased sensitivity in the assessment of instantaneous measures. The bottom panel in Fig. 2d illustrates the model's automatic detection of outliers; when an outlier is detected (black circles on *x*-axis), the model suspends the analysis and retains the last estimate for 5 s.

D. Statistical Analyses

Figure 3 presents a statistical analysis (one-way analysis of variance) for the effect of Group (Young, Elderly, PD) on the average value of each instantaneous measure within the ~300-second walk, for each subject. (Standard deviations are shown around the group means). The groups did not significantly differ for instantaneous ST or SR, but both showed a significant effect of Group on instantaneous STV (3b) and SRV (3d). Specifically, the PD patients showed significantly greater variability during their walks than both Young and Elderly subjects, consistent with traditional analyses (i.e., coefficient of variation) (e.g., [20]).

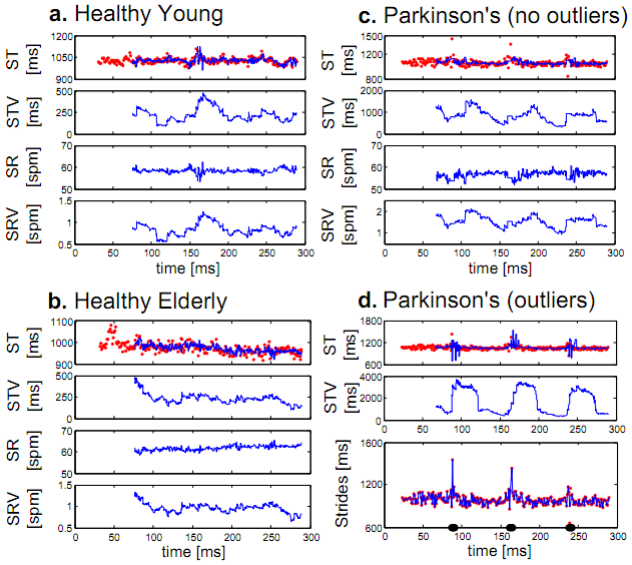

Figure 2. Comparison of instantaneous mean (ST, SR) and standard deviation (STV, SRV) for three exemplar subjects in Dataset 2. The effect of outliers (d.) is illustrated for the PD subject.

Furthermore, the SRV measure was more sensitive in differentiating between Elderly and PD subjects; the effect size (partial η^2) for explained group variance between Elderly and PD was 58.7% for STV and 88.4% for SRV. Finally, we took the ratio of the standard, window-based index of variability to our new estimate of total *GG* variability for each subject. As shown in Figure 3e, the ratios were greater than 1.0 (indicating larger variability estimates in the window-based estimate) for Young (1.15), Elderly (1.35), and PD (2.04). We interpret our lower estimates as reflective of the model's ability to capture nonstationary dynamics and eliminate high-variance outliers.

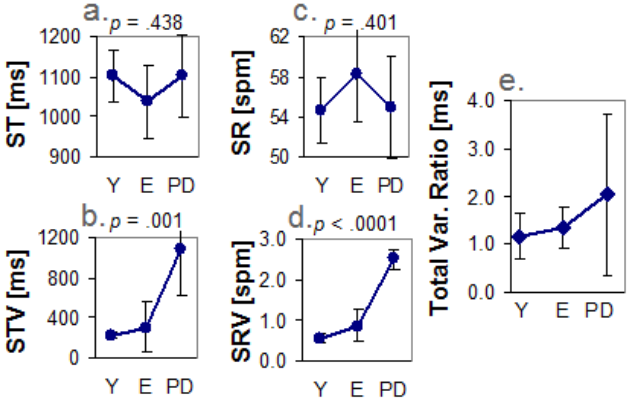

Figure 3. One-way ANOVAs comparing group differences. All tests have 12 degrees of freedom. Error bars are SDs.

IV. Discussion

Our results can be summarized by three points. First, we demonstrated that our model is capable of capturing the underlying structure present in healthy, un-paced walking that disappears under metronome walking conditions (i.e., when gait is synchronized to an external stimulus). Previous analyses of self-paced versus metronome-paced walks in healthy subjects have revealed the presence of an underlying fractal or self-similar structure during self-paced walks [19] that disappears when subjects synchronize to an external auditory stimulus [18]. This striking change is hypothesized to be driven by a shift from an "externally-paced motor network" to an "internally-paced motor network," which involve different brain regions (for a review, see [21]).

Second, our model provides instantaneous measures of ST and STV as well as SR and SRV. These measures yield novel insights into the temporal dynamics of gait, including temporal profiles of gait that change as a function of age and disease. Furthermore, as SR and SRV are derived using the estimated instantaneous mean (μ_{GG}), they capture a distinct dynamic component of gait variability that permits and merits further exploration.

Third, our model detects outliers based on a maximum likelihood estimation of the event series' underlying probability structure, and suspends calculations for 5 s. This prevents an artificial inflation of variance due to a few large aberrant gait events, leading to increased sensitivity in detecting more subtle differences between, for example, healthy elderly versus PD patients. Previous time domain analyses of gait variability have used an arbitrary cutoff of >3 SD from the median gait interval [2], [20].

Finally, we note that the ability to generate instantaneous measures of mean and variability opens up new avenues to explore the real time dynamics of human gait, including correlations with other behavioral or neural measures [22], analyses of gait patterns during clinical assessment (e.g., changing floor surfaces, moving through doorways, and turning) [7], and patient monitoring applications that could trigger rhythmic sensory cues or electronic compensatory devices (cf. [23]).

Further work will seek to explore whether our measures change as a function of disease severity, including the potential of our model to detect gait abnormalities prior to detection using traditional clinical measures (e.g., [24]).

V. Conclusion

We present a new model of human gait which is based on the treatment of gait events (i.e., heel strikes) as a point process rather than a time series, modeling gait intervals using a history-dependent inverse Gaussian process with local maximum likelihood estimates of the model's time-varying parameters. The model provides a robust and coherent statistical model of two measures of gait, permits goodness-of-fit assessments, and yields instantaneous measures of stride time and stride rate, useful in future diagnostic or monitoring applications serving the growing millions affected by with movement disorders, such as Parkinson's disease.

References

[1] J. Hausdorff, "Gait variability: methods, modeling and meaning," *J. NeuroEngineering and Rehabilitation*, vol. 2, no. 1, p. 19, 2005.

[2] J. M. Hausdorff et al., "Gait variability and basal ganglia disorders," *Movement Disorders*, vol. 13, no. 3, pp. 428-437, 1998.

[3] B. H. Wood et al., "Incidence and prediction of falls in Parkinson's disease: a prospective multidisciplinary study," *J. Neurology, Neurosurgery & Psychiatry*, vol. 72, no. 6, pp. 721 -725, 2002.

[4] J. M. Hausdorff, "Gait dynamics in Parkinson's disease," *Chaos*, vol. 19, no. 2, p. 026113, 2009.

[5] T. Chau, "A review of analytical techniques for gait data. Part 1: fuzzy, statistical and fractal methods," *Gait & Posture*, vol. 13, no. 1, p. 49–66, 2001.

[6] T. Chau, "A review of analytical techniques for gait data. Part 2: neural network and wavelet methods," *Gait & Posture*, vol. 13, no. 2, pp. 102-120, 2001.

[7] M. E. Morris, "Locomotor Training in People With Parkinson Disease," *Physical Therapy*, vol. 86, no. 10, pp. 1426-1435, 2006.

[8] D. J. Daley and D. D. Vere-Jones, *An introduction to the theory of point processes*, 2nd ed. New York: Springer, 2007.

[9] R. Barbieri et al., "A point-process model of human heartbeat intervals," *Am J Physiol Heart Circ Physiol*, vol. 288, no. 1, pp. H424-435, 2005.

[10] R. Barbieri and E. N. Brown, "Analysis of heartbeat dynamics by point process adaptive filtering," *IEEE Transactions on Bio-Medical Engineering*, vol. 53, no. 1, pp. 4-12, 2006.

[11] Z. Chen et al., "Characterizing Nonlinear Heartbeat Dynamics within a Point Process Framework," *IEEE Transactions on Bio-Medical Engineering*, Feb. 2010.

[12] E. N. Brown et al., "The time-rescaling theorem and its application to neural spike train data analysis," *Neural Computation*, vol. 14, no. 2, p. 325–346, 2002.

[13] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. e215-220, Jun. 2000.

[14] http://www.physionet.org/physiobank/database/umwdb/

[15] http://www.physionet.org/physiobank/database/gaitdb/

[16] S. M. Ross, *Introduction to Probability Models, Ninth Edition*, 9th ed. Academic Press, 2006.

[17] C. Loader, *Local Regression and Likelihood*, 1st ed. Springer, 1999.

[18] J. M. Hausdorff et al., "Fractal dynamics of human gait: stability of long-range correlations in stride interval fluctuations," *J. Applied Physiology*, vol. 80, no. 5, pp. 1448-1457, 1996.

[19] J. M. Hausdorff et al., "Is walking a random walk? Evidence for long-range correlations in stride interval of human gait," *J. Applied Physiology*, vol. 78, no. 1, pp. 349-358, 1995.

[20] J. M. Hausdorff et al., "Altered fractal dynamics of gait: reduced stride-interval correlations with aging and Huntington's disease," *J Appl Physiol*, vol. 82, no. 1, pp. 262-269, 1997.

[21] D. J. Brooks, "Functional imaging studies on dopamine and motor control," *Journal of Neural Transmission*, vol. 108, no. 11, pp. 1283-1298, 2001.

[22] V. Napadow et al., "Brain correlates of autonomic modulation: combining heart rate variability with fMRI.," *Neuroimage*, vol. 42, no. 1, pp. 169-177, 2008.

[23] M. Bächlin et al., "Wearable assistant for Parkinson's disease patients with the freezing of gait symptom," *IEEE Transactions on Information Technology in Biomedicine: A Publication of the IEEE Engineering in Medicine and Biology Society*, vol. 14, no. 2, pp. 436-446, 2010.

[24] C. G. Goetz et al., "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results," *Movement Disorders*, vol. 23, no. 15, pp. 2129-2170, 2008.