A Symbol-Based Approach to Gait Analysis From Acceleration Signals: Identification and Detection of Gait Events and a New Measure of Gait Symmetry

Anita Sant'Anna and Nicholas Wickström

Sant'Anna A, Wickström N. A Symbol-Based Approach to Gait Analysis From Acceleration Signals : Identification and Detection of Gait Events and a New Measure of Gait Symmetry. New York: IEEE; IEEE transactions on information technology in biomedicine. 2010;14(5):1180-1187.

DOI: http://dx.doi.org/10.1109/TITB.2010.2047402
Copyright: IEEE

Post-Print available at: Halmstad University DiVA
http://urn.kb.se/resolve?urn=urn:nbn:se:hh:diva-5439

A symbol-based approach to gait analysis from acceleration signals: identification and detection of gait events and a new measure of gait symmetry

Anita Sant'Anna, Nicholas Wickström
School of Information Science, Computer and Electrical Engineering
Halmstad University - Sweden
contact: Anita.Santanna@hh.se

***Abstract*—Gait analysis can convey important information about one's physical and cognitive condition. Wearable inertial sensor systems can be used to continuously and unobtrusively assess gait during everyday activities in uncontrolled environments. An important step in the development of such systems is the processing and analysis of the sensor data. This paper presents a symbol-based method used to detect the phases of gait and convey important dynamic information from accelerometer signals. The addition of expert knowledge substitutes the need for supervised learning techniques, rendering the system easy to interpret and easy to improve incrementally. The proposed method is compared to an approach based on peak-detection. A new symbol-based symmetry index is created and compared to a traditional temporal symmetry index and a symmetry measure based on cross-correlation. The symbol-based symmetry index exemplifies how the proposed method can extract more information from the acceleration signal than previous approaches.**

I. Introduction

Gait analysis has been investigated as an indicator of both physical and cognitive condition. Gait analysis can be used to help diagnose and assess the severity of neurological conditions such as Parkinson's disease [1] and stroke [2]. Measures of gait speed and gait variability have been associated with the risk of developing dementia and mild cognitive impairment [3]. In addition, measures of gait symmetry are important when assessing the quality of gait of amputees [4] and stroke patients [5].

Gait analysis can be performed by a trained clinician through observation, or in a gait and balance laboratory with the help of motion capture (mocap) systems. The benefits of using mocap systems include great accuracy and having complete trajectory information. On the other hand, some of the drawbacks are that the tests are expensive, patients are assessed infrequently and only under controlled situations. Therefore, this clinical gait analysis does not reflect the patient's everyday activities and, as such, may overlook important information. In addition, the frequency of the assessments may not be high enough to detect short term variations. Continuous gait monitoring in uncontrolled environments is, therefore, an important complement to traditional clinical gait assessment.

Studies of mobile (wearable) motion analysis systems have employed a number of different sensing technologies e.g. accelerometers, gyroscopes, pressure sensitive insoles and magnetometers. Accelerometers are a common choice of sensor given their small size, low cost, and low power consumption [6]. In addition, accelerometers can be easily embedded into household objects and clothing items, creating pervasive and unobtrusive systems. Previous gait analysis studies using only accelerometers have mainly focused on either the identification of gait phases, or the classification of walking patterns. Most works have made use of supervised learning techniques and manually tuned parameters which relate to the raw data but not to gait. These approaches strongly depend on training data. The trained model usually resembles a "black box" which clinicians may find difficult to interpret.

This paper is concerned with the processing and analysis of acceleration data for use in unobtrusive gait analysis systems. The contributions of this work include a fully automatic method for extracting gait measurements from accelerometer data. This method is not dependent on supervised learning techniques and conveys more information than previous methods. The method uses symbolization in order to abstract the data to a higher level representation and facilitate the inclusion of expert knowledge. This is a general approach which can be used to extract several clinically interesting measures. This paper reports the use of the method for detecting heel-strike and toe-off events, and deriving a new symmetry measure based on dynamic rather than static information.

II. Related Work

Previous gait analysis studies using accelerometers have mainly focused on either the identification of gait phases, or the classification of walking patterns. Identification of gait phases is normally achieved through detection of particular events in time such as heel-strike (HS) and toe-off (TO), e.g. [7]. This is typically done using thresholding or peak detection, e.g. [8], [9]. Supervised machine learning techniques, such as Artificial Neural Networks (ANN), have also been used, e.g. [10]. These methods normally convey only temporal information about the signal and no information about how the subject's feet are moving through space.

The other group of methods aims to classify walking patterns based on certain dynamics of the walk: walking cadence, if the subject is walking up or down stairs, walking or running, e.g. [11]. These studies normally involve methods based on

sliding-window statistics [12], time-frequency analysis [13], wavelet decomposition [14], and/or ANN [15]. This class of methods is able to describe, to a certain extent, how the subject is walking. However, the phases of gait are overlooked and the information obtained from such methods is insufficient for clinical gait analysis applications. In contrast to both groups of studies, the present work aims to extract both temporal events and dynamic characteristics in order to perform clinical gait analysis.

Gait analysis in the clinical setting is normally concerned with quantitative measures such as gait speed, step length, double support time, stride-to-stride variability and symmetry, e.g. [16], [17], [18]. The first step in identifying such measures is to detect HS and TO. Most previous approaches to identify HS and TO using only accelerometers were based on signal filtering and peak detection. Aminian et al. [19] used a top-down approach to detect peaks corresponding to HS and TO. The analysis started with a rough estimate of HS and TO from a low-pass filtered signal and iteratively closed in on a better estimate from the signal filtered at higher frequencies. More recently, Selles et al. [9] used a low-pass filtered signal to estimate the average stride period, then used this estimate to design different low-pass filters for slow and fast walks. The peaks found in the filtered signals guided the search for HS and TO. Both methods are essentially similar, and depend on appropriate tuning of model parameters. Such parameters can be filter cut-off frequencies, size and location of the window where to narrow down on the estimates of HS and TO, etc. The tuning of these low-level parameters optimizes the method to a given data set and hinders the generalization of the method. The proposed approach aims to overcome the parametrization problem by abstracting the data to a higher level representation where the characteristics of the system can be generalized to different data sets.

Once HS and TO have been detected, other temporal measures may be extracted, e.g. [20]. Symmetry is an important estimate of quality of gait, specially when assessing amputees, stroke patients, and other conditions which typically affect one side of the body. Given the data available from the previously mentioned peak detection methods, a commonly used clinical measure of symmetry is: $SI = \frac{T_R - T_L}{\frac{1}{2}(T_R + T_L)} 100$, where T_R is the average stride time for the right foot and T_L is the average stride time for the left foot [4]. According to SI, the closer the absolute value is to zero, the more symmetric the walk. Although a negative value indicates that the left foot is slower than the right foot, a slower stride time does not indicate a more abnormal movement. The value for this index ranges from -200 to 200 and, in practice, a correspondence between this index and quality of gait is unclear. This measure of symmetry only takes into account the average stride time for each foot. If the subject limps but manages to keep the same cadence, the SI index will not consider this to be an asymmetric walk.

A more informative symmetry measure must consider the whole acceleration signal. Examples of techniques used to derive gait symmetry measures from acceleration signals include eigenvectors [21], principal component analysis [22], frequency analysis [23], and cross-correlation [24]. The latter can be used to compare the shape of any two signals. Miller et all [24] averaged EMG signals from consecutive strides creating a Latency Corrected Ensemble Average (LCEA) for the right, $\hat{y}_R$, and left, $\hat{y}_L$ feet. A normalized cross-correlation curve ρ was used to determine how similar the two LCEA were according to: $SI_{xcorr} = max\{\rho_{\hat{y}_L \hat{y}_R}\}100$, where $\rho_{\hat{y}_L \hat{y}_R} = \frac{r_{\hat{y}_L \hat{y}_R}}{\sqrt{r_{\hat{y}_L \hat{y}_L}(0) r_{\hat{y}_R \hat{y}_R}(0)}}$. SI_{xcorr} is the symmetry magnitude and r_{ij} is the cross correlation between signals i and j. The closer SI_{xcorr} is to 100, the more symmetric are the signals.

Symbolization is another approach used to extract information about the underlying dynamics of time-series [25]. An important practical advantage of working with symbols is that the efficiency of numerical computations is greatly improved, i.e. compression. Also, symbolic data is often less sensitive to measurement noise. Many previous works have used symbolic transformation of continuous data with great success [25]. Another advantage of using symbolization is the possibility to use context-based analysis. A classic example of context-based analysis is Optical Character Recognition, where the individually classified characters in a word are checked against a list of possible words. A similar context-based analysis can be performed on motion data if the possible symbol patterns are identified. Some works have already considered the use of symbolization and context analysis on motion capture data [26] or visual data [27]. These works used a linguistic approach to classify different activities. The present work, however, uses a symbol-based approach to extract gait measurements from acceleration data.

III. Method

The method proposed here aims to extract temporal (HS and TO) and dynamic gait measurements from acceleration signals. This is achieved through symbolization of the signal, and analysis of the context and distribution of each symbol. The symbolized data represents a higher abstraction level, which is more easily coupled to expert knowledge. The use of expert knowledge substitutes the need for supervised learning techniques. The proposed method is, therefore, automatic and does not depend on extensive model training.

Throughout the paper, transition ij refers to the moment when symbol i ends and symbol j begins. A stride lasts from HS to the following HS on the same foot. Swing is the period during which the foot is off the ground (from TO to HS). Stance is the period during which the foot is on the ground (from HS to TO).

A. Symbolization

The symbolization of the signal is illustrated in Figure 1. The acceleration signals were filtered with a low-pass filter at 20 Hz. The resultant acceleration was calculated according to: $A_{res} = \sqrt{A_x^2 + A_z^2}$, where A_x and A_z are accelerations in the accelerometer's local coordinate system. The "sideways" acceleration was not considered. It was judged dispensable when the subjects walk a straight line (See Section IV).

The resultant acceleration was segmented according to a bottom-up piecewise linear segmentation algorithm described

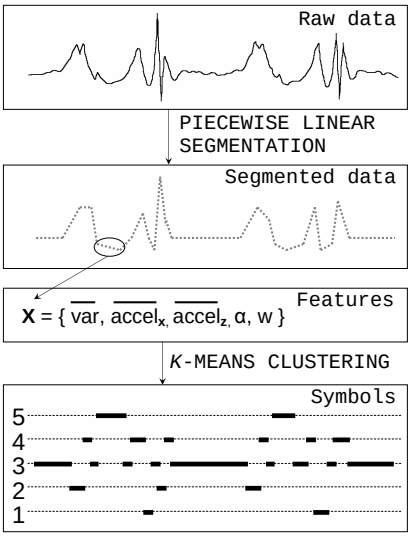

Fig. 1. **Symbolization**. Graphical representation of the steps taken towards symbolization of the signal.

in [28]. The algorithm starts by fitting a small line segment over every consecutive two samples, and iteratively merges neighboring line segments to form longer ones. The merging line segments are chosen so as to minimize the mean square error (MSE) between the original signal and the resulting line segment. The process continues until the total MSE reaches a predetermined threshold. For this study, the MSE threshold used was 0.1. One of the purposes of symbolization is to reduce the amount of information that needs to be processed while preserving the main characteristics of the signal. The piecewise linear segmentation was chosen in order to preserve its overall shape.

The segment features were chosen in order to preserve the shape of the line segment (slope and length), and the main characteristics of the original signal (mean and variance). The features extracted from each segment were: mean segment variance of the resultant acceleration; mean segment acceleration on both axes; the tangent of the angle between the approximated line segment and the horizontal axis; and the number of samples in the segment. Other sets of features are possible, depending on the targeted analysis.

The acceleration signals and the segment features were standardized (zero mean and unit standard deviation). K-means clustering of the features was used to divide the segments into groups. Clustering was performed considering from 2 to 10 clusters. The optimum number of clusters was chosen based on the minimum Davies-Bouldin index [29]. The limit of 10 was chosen because most of the initial trials resulted in less than 10 clusters. A unique symbol (integer between 1 and the number of clusters) was assigned to the segments belonging each cluster. This clustering technique can adequately adapt to very different signals. The technique can produce, when appropriate, different sets of symbols for each signal. This is important because two subjects may have very different walking patterns. Nonetheless, the "context analysis" phase is the same regardless of the set of symbols used.

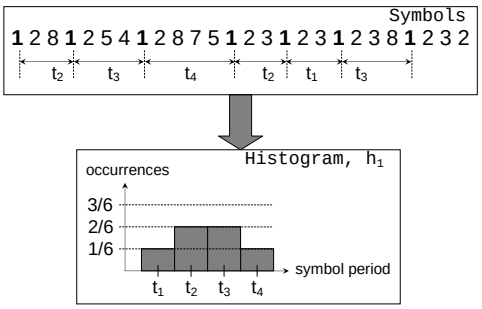

Fig. 2. **Period histogram**. Example of how to compute the histogram of symbol periods for symbol 1. The time elapsed between two consecutive transitions is calculated and the corresponding histogram bin is incremented.

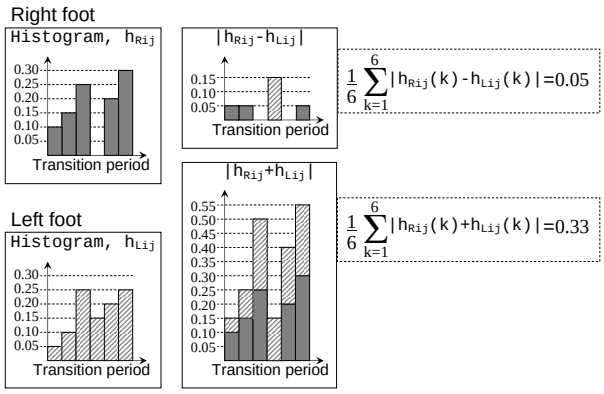

Fig. 3. **Histogram analysis**. Example of histogram analysis for **one** particular pair of transition histograms. The symmetry index is a combination of this analysis for all transitions.

B. Symbol-based Symmetry Index

The symbolized data can be used to compute a measure of gait symmetry. This symbol-based index takes all the acceleration data and its dynamics into account. Other authors have investigated symmetry indices based on acceleration signals, e.g. [24]. However, the proposed index takes advantage of the symbolization technique described previously.

After the segments are clustered into Z symbols, $\boldsymbol{S} = \{S_1, S_2, ...S_Z\}$, the symbolized sequence is analyzed in terms of its symbol periods. The periods between two consecutive appearances of the same symbols are computed for all symbols. The symbol periods are organized into histograms as exemplified in Figure 2. The symmetry index based on the period histograms, SI_{symb}, is computed by:

$$SI_{symb} = \frac{\sum_{i=1}^{Z} \frac{1}{n_i} \sum_{k=1}^{K} |h_{Ri}(k) - h_{Li}(k)|}{\sum_{i=1}^{Z} \frac{1}{n_i} \sum_{k=1}^{K} |h_{Ri}(k) + h_{Li}(k)|} 100, \quad (1)$$

where Z is the number of symbols; K is the number of bins in the histograms; n_i is the number of non-empty histogram bins (for either foot) for symbol i; $h_{Ri}(k)$ is the normalized value for bin k in the period histogram i for the *right* foot; and $h_{Li}(k)$ is the normalized value for bin k in the period histogram i for the *left* foot. The transition histograms are normalized so as to consider the relative number of symbol periods, disregarding the number of steps recorded.

This index ranges from 0 to 100, where 0 means total *symmetry* and 100 means complete *asymmetry*. SI_{symb} takes into account not only the stride times but also the dynamics of

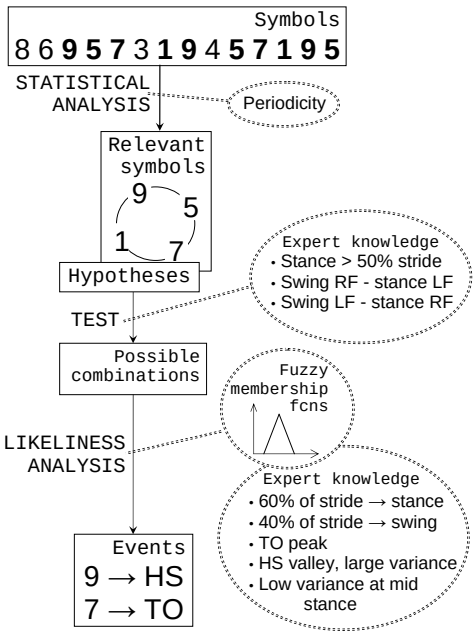

Fig. 4. **Context analysis**. Graphical representation of the steps taken towards context analysis of the signal.

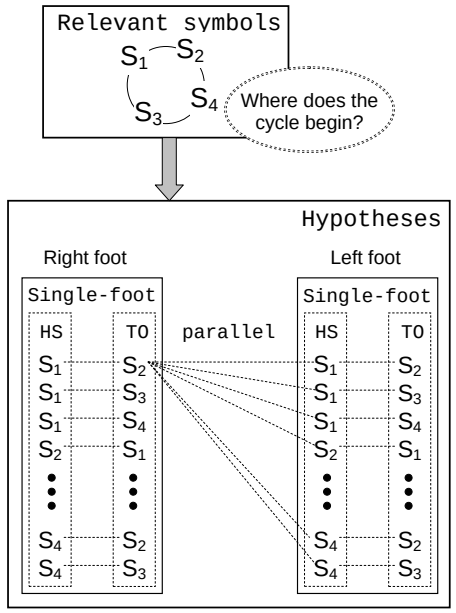

Fig. 5. **Creating hypotheses**. The single step hypotheses are all pairwise combinations of the relevant symbols, and parallel hypotheses are all permutations of the right foot and left foot single step hypotheses.

the movement. This symmetry index can be interpreted as a comparison between the distributions of the transition periods. Although there are several possible ways to analyze the difference between two distributions, this was chosen for its simplicity. Figure 3 illustrates the histogram analysis for **one pair** of transition histograms. The SI_{symb} index demonstrates one way in which the symbol abstraction employed for this method can be directly used to extract meaningful dynamics information from the sensor data.

C. Context Analysis

The physical characteristics of the system are reflected on the symbolic data as certain symbol sequences or symbol distributions. Some of the gait characteristics expected to be found in the signal are described in Table I. This knowledge was incorporated into the model in order to find the symbols corresponding to HS and TO. The algorithm is divided into three main steps: finding relevant symbols, hypotheses testing, and estimating "likeliness", as illustrated in Figure 4.

1	During normal walk, approximately 60% of the total stride time corresponds to stance
2	TO is reflected on the resultant acceleration signal as a peak
3	HS is reflected on the resultant acceleration signal as a valley and large variance
4	The foot moves the least at mid-stance (very small variance of the resultant acceleration)

TABLE I
GAIT CHARACTERISTICS. EXPERT KNOWLEDGE ABOUT SOME PHYSICAL CHARACTERISTICS OF GAIT.

1) Finding relevant symbols: Taking advantage of the cyclic nature of gait, the most common period out of all symbols and symbol transitions is considered the average stride period, $StrP$. Relevant symbols are symbols which are likely to express striking characteristics of the original signal, such as HS and TO. They are expected to appear approximately once every cycle. Symbols (or transitions) with period similar to or half of the estimated stride period are considered relevant. From this point on, relevant transitions are represented as extra relevant symbols and analyzed similarly.

The original acceleration signal can now be represented as a cyclic sequence composed of relevant symbols (see Figure 4). It is important to determine where the cycle begins, i.e. which relevant symbol corresponds to HS. In order to determine which symbols should be associated with HS (and TO), all relevant symbols are considered and evaluated according to certain assumptions. For each foot, the N relevant symbols $\boldsymbol{S} = \{S_1, S_2, ..., S_N\}$ are organized into all possible pairwise combinations. These combinations are hypotheses of which symbols could correspond to HS and TO. The combinations for the right foot (Eq. 2) and left foot (Eq. 3), are then recombined to express all possible permutations, considering both feet in parallel (Eq. 4). The single-foot and parallel hypotheses are exemplified in Figure 5.

$$C^{RF} = \{(S_1, S_2), (S_1, S_3), ...(S_N, S_{(N-1)})\} = \{C^{RF}_{1,2}, C^{RF}_{1,3}, ...C^{RF}_{N,(N-1)}\} \quad (2)$$

$$C^{LF} = \{(S_1, S_2), (S_1, S_3), ...(S_N, S_{(N-1)})\} = \{C^{LF}_{(1,2)}, C^{LF}_{(1,3)}, ...C^{LF}_{(N,(N-1))}\} \quad (3)$$

$$C_{parallel} = \{(C^{RF}_{1,2}, C^{LF}_{1,2}), (C^{RF}_{1,2}, C^{LF}_{(1,3)}), ...(C^{RF}_{N,(N-1)}, C^{LF}_{(N,(N-1))})\} \quad (4)$$

2) Hypotheses Testing: The single-foot hypotheses created in the previous step are tested according to the assumption

Observations	Input variable for each hypothesis k	Partial likeliness value for each hypothesis k
Approximately 60% of the stride time corresponds to stance	relative stance time, $StnT_{relative}(k) = \frac{StnT(k)}{StrP(k)}$, where $StrP$ is the stride period and $StnT$ is the stance time	$L_1(k) = F_C(StnT_{relative}(k))$
TO is reflected on the resultant acceleration as a peak	average resultant acceleration at TO, $TO_{accel}(k)$	$L_2(k) = F_B(TO_{accel}(k))$
HS is reflected on the resultant acceleration signal as a valley and large variance	average resultant acceleration at HS, $HS_{accel}(k)$, and the maximum acceleration variance between HS and HS + 10 samples, $HS_{var}(k)$	$L_3(k) = F_A(HS_{accel}(k))$ $L_4(k) = F_A(1\text{-}HS_{var}(k))$
The foot moves the least at mid-stance	average variance of resultant acceleration at mid-stance, $MidStance_{var}$	$L_5(k) = F_A(MidStance_{var}(k))$
Detected number of strides must be approximately 40% of the maximun number of strides which could fit within the recorded data	relative number of detected strides, $N_{relative}(k) = \frac{N(k)StrP(k)}{M}$, where, N is the number of detected strides and M is the length of the recorded data set	$L_6(k) = F_D(N_{relative}(k))$

TABLE II
PARTIAL LIKELINESS VALUES CALCULATED FROM HYPOTHESIS ESTIMATES AND MEMBERSHIP FUNCTIONS. EACH LIKELINESS VALUE CORRESPONDS TO A PIECE OF EXPERT KNOWLEDGE.

that stance lasts over 50% of the stride time. The average stride period is calculated considering the corresponding HS symbol. For each hypothesis, average stride period and stance time are compared. If the average stance time is shorter than half the average stride time, the hypothesis is discarded. Then, the remaining parallel hypotheses are tested based on the assumption that swing in one foot can only take place during stance in the other foot. The parallel hypotheses that fail this test are discarded.

3) Estimating likeliness: The remaining hypotheses are considered possible, and a measure of "likeliness" tries to estimate which hypothesis is more likely to be true. The likeliness of each hypothesis k is estimated according to the four observations stated in Table I. An extra assumption is added to ensure that the most likely hypothesis is able to detect an adequate number of strides, given the size of the data. Each observation is represented by a fuzzy membership function which maps measures such as average symbol acceleration and variance to a likeliness value. The four membership functions used are illustrated in Figure 6. The calculus of the partial likeliness value associated with each assumption is shown in Table II. For each hypothesis k, the final likeliness value, $L(k)$, is obtained from the product of the corresponding six partial likeliness values for each foot, $L(k) = \prod_{i=1}^{6} L_i^{RF}(k).\prod_{i=1}^{6} L_i^{LF}(k)$. The hypothesis with the largest final likeliness value is chosen as true and the corresponding symbols are used to identify HS and TO instances in the data.

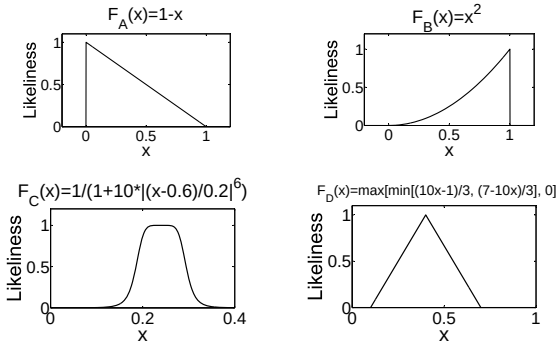

Fig. 6. **Membership functions** used to estimate the likeliness of each hypothesis. The inputs to each of these functions is explained in Table II.

IV. EXPERIMENTS

Gait acceleration data was collected in order to investigate the use of the proposed method for gait analysis. The hardware used for the experiments consisted of:

- Two SHIMMER sensor nodes (shimmer-research.com) each equipped with a tri-axial accelerometer, sampling at 50Hz. The data was streamed continuously via Bluetooth to a nearby computer.
- A six-meter-long Gold Gait Rite pressure sensitive mat [30], sampling at 60Hz. The Gait Rite has its own software for detecting HS, TO and other temporal gait measurements from the pressure sensitive mat. The Gait Rite was only used to provide a reference signal, it is not part of the proposed method.

Six volunteers participated in the experiments. The subjects had the SHIMMER nodes attached to both shins, close to the ankles. When the subject was standing still, the z axis of the accelerometer corresponds to the vertical axis, and the x axis corresponds to the horizontal axis tangential to the subject's sagittal plane.

The subjects were asked to walk a straight line on the pressure mat according to three different instructions: 1) to walk at a comfortable self-paced speed, *normal walk data set*; 2) to walk at a very slow speed taking shorter steps, *slow walk data set*; and 3) to walk while having the right knee immobilized with a brace in order to simulate limping, *limp walk data set*.

The data obtained from the pressure mat was used as ground truth for HS and TO. The proposed method was applied so as to create subject-independent "normal" symbols. The "normal walk" data sets of 3 randomly chosen individuals were

combined, segmented and clustered, resulting in 7 symbols. The centers of these clusters were used to symbolize the remaining data sets into equivalent symbols. All data sets were analyzed in the exact same way.

In order to evaluate the performance of the proposed method with regards to previous approaches, a peak-detection method for detecting HS and TO was implemented [19]. This method was chosen over a more recent work [9]. Both works are based on the same principles but [19] seems to be less dependent on the tuning of optimizing parameters, and therefore more robust to different data sets. The method presented in [19] was slightly modified in order to cope with the different sensor placement. At each iteration, the interval used to narrow down on the location of the peaks was changed to a window ten-sample-wide, centered at the peak location from the previous iteration. All data sets, were submitted to the peak-detection analysis in the exact same way.

The symbol-based symmetry index was computed for all three data sets. In order to evaluate the information conveyed by this index, two other symmetry indices were calculated as reported in [4] and [24]. The traditional measure of symmetry explained in [4] was computed from the ground truth (GT) data, i.e. data from the pressure sensitive mat. The cross-correlation method stated in [24] was computed from the acceleration signals, using the GT for stride segmentation.

V. Results

A. Event Detection

The detected HS and TO instances, both from the proposed method (symbolic approach) and from the peak-detection method, were compared to the GT data. The resulting mean absolute errors and standard deviations for each data set, in seconds, are presented in Tables III, IV, and V. Table III shows that both the symbolic approach and the peak-detection methods may be used to detect HS and TO instances from gait acceleration data during "normal walking" with small errors. Considering the sampling frequency, 50Hz, the mean errors correspond to approximately 2 samples. TO instances are detected more accurately than HS instances. Table IV indicates that, for the slow walk data set, mean errors correspond to approximately 4% and 15% of the average stride time for the symbolic approach and for the peak-detection method respectively. Table V shows that, for the limp walk data set, mean errors correspond to approximately 10% and 11% of the average stride time for the symbolic approach and for the peak detection method respectively.

The data set encompasses 131 "normal walk" steps, 167 "slow walk" steps and 114 "limp walk" steps. Each step generates 2 error samples (HS and TO). Since the subjects were random and all healthy, it is reasonable to assume the error samples are identically distributed (for each type of walk). T-tests were used to determine if the mean absolute errors for each method were statistically different for each type of walk:

$H_0 : \bar{\mu}_{symb} - \bar{\mu}_{peak} = 0;$

$H_1 : \bar{\mu}_{symb} - \bar{\mu}_{peak} \neq 0;$

where $\bar{\mu}_{symb}$ is the mean absolute error in detecting HS and TO with the symbolic approach, and $\bar{\mu}_{peak}$ is the mean absolute error using the peak detection method. Results are shown in Table VI. The tests indicate that the symbolic approach performs equally well to the peak-detection method for the "normal walk" and "limp walk" data sets. The "slow walk" data set is more accurately analyzed by the symbolic approach.

Normal walk data		
Method	Event	mean (standard deviation)
Symbolic approach	HS	0.05 (0.04)
	TO	0.03 (0.04)
Peak detection	HS	0.07 (0.10)
	TO	0.03 (0.03)

TABLE III
MEAN ABSOLUTE ERROR IN SECONDS AND STANDARD DEVIATION FOR THE SYMBOLIC AND PEAK DETECTION METHODS ON THE NORMAL WALK DATA SET. THE AVERAGE STRIDE TIME FOR THIS DATA SET IS 1.01S.

Slow walk data		
Method	Event	mean (standard deviation)
Symbolic approach	HS	0.06 (0.10)
	TO	0.05 (0.13)
Peak detection	HS	0.22 (0.17)
	TO	0.16 (0.26)

TABLE IV
MEAN ABSOLUTE ERROR IN SECONDS AND STANDARD DEVIATION FOR THE SYMBOLIC AND PEAK DETECTION METHODS ON THE SLOW WALK DATA SET. THE AVERAGE STRIDE TIME FOR THIS DATA SET IS 1.46S.

Limp walk data		
Method	Event	mean (standard deviation)
Symbolic approach	HS	0.10 (0.10)
	TO	0.09 (0.12)
Peak detection	HS	0.12 (0.11)
	TO	0.08 (0.08)

TABLE V
MEAN ABSOLUTE ERROR IN SECONDS AND STANDARD DEVIATION FOR THE SYMBOLIC AND PEAK DETECTION METHODS ON THE LIMP WALK DATA SET. THE AVERAGE STRIDE TIME FOR THIS DATA SET IS 1.04S.

	Normal Walk	Slow Walk	Limp Walk
α	0.05	0.05	0.05
Reject H_0?	NO	YES	NO
p-value	0.12	$\sim$0	0.72
95% conf. interval	[-0.02, 0]	[-0.16, -0.11]	[-0.02, 0.01]
n. of samples per group	262	334	228
post-hoc power[1]	82%	77%	78%

TABLE VI
RESULTS OF STATISTICAL TESTS.

B. Symmetry Indices

The symmetry indices were calculated for each subject in each data set. The distributions of the symmetry indices within each data set are shown as box-plots in Figure 7.

[1]post-hoc power calculated using "Cohen's d" for effect size [31].

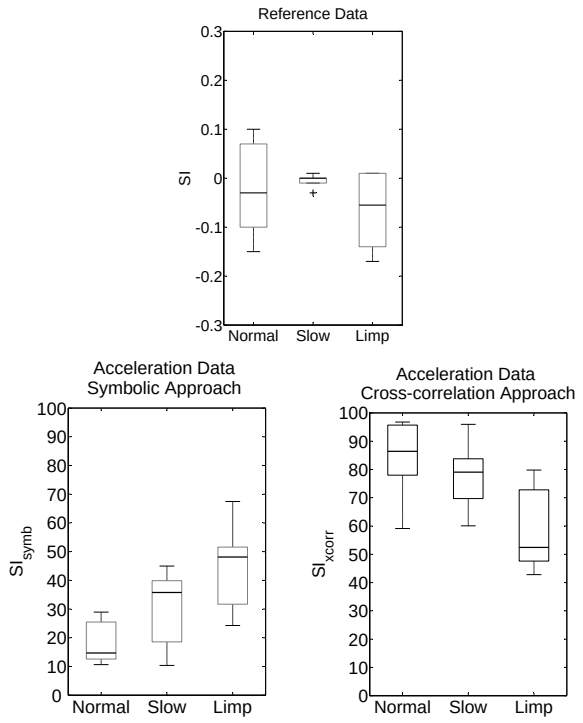

Fig. 7. Symmetry indices. The top plot shows the traditional symmetry measure. The bottom-right plot shows the symbol-based symmetry index. The bottom-left plot shows the cross-correlation symmetry measure. The thick horizontal line corresponds to the median; the lower edge of the box corresponds to the lower quartile; the upper edge of the box corresponds to the upper quartile; the lower "whisker" corresponds to the smallest non-outlier value; the upper "whisker" corresponds to the largest non-outlier value; and outlier values are represented with crosses.

Figure 7 illustrates that the SI index judged "limp walk" data as symmetric as the "normal walk" data. Note that SI may take values between -200 and 200. Given that the SI results are all in the interval [-0.2, 0.2], 0.1% of the index full range, they cannot be considered significantly different. In contrast, both the symbol-based symmetry index SI_{symb} and the cross-correlation measure, SI_{xcorr}, consistently differentiate the data sets in terms of symmetry. Normal walk is, on average, more symmetric than slow walk, which is more symmetric than limp walk. For the SI_{symb} values close to zero indicate symmetry, whereas for SI_{xcorr}, values close to 100 indicate symmetry.

VI. Discussion

The peak detection results are slightly poorer than the results reported by [9], which had an algorithm optimized to the observed data set. For all the three types of walk investigated, the symbolic approach performs equally well or better than the peak detection algorithm.

The proposed symbolic approach is shown to be robust to the different data sets. Although it was designed to analyze normal walking, it provided reasonable estimates of HS and TO for both slow and limp walk data sets. This illustrates how the proposed method is not dependent on training data. According to Table IV the peak detection method is a lot less robust to certain data sets. This can be explained by the fact that slow walk generates much smaller accelerations and milder peaks, which were harder to detect without altering the filtering frequencies.

The differences between normal and limp walk for the traditional symmetry index, SI, were not significant. On the other hand, SI_{symb} and SI_{xcorr} were able to detect certain dynamic information from the signal. In addition, both SI_{symb} and SI_{xcorr} are in agreement regarding the overall symmetry of each type of walk. Slow walk was probably less symmetric than normal walk because the subjects needed to make a conscious effort to change their walking pattern, sometimes oscillating between their normal speed and a slower speed. Although the information conveyed by SI and that conveyed by SI_{symb} and SI_{xcorr} are not equivalent, this analysis aims to exemplify how the peak detection approach fails to extract dynamic information by only taking into account temporal measurements of HS and TO.

The method presented here is totally automatic and does not depend on labeled training data, unlike most previous works based on supervised learning techniques. The use of expert rules for context analysis allows the system to be gradually improved by altering or adding new rules. This process avoids the need to retrain the whole system. This expert system depends on careful tuning of the rules. However, these rules may be coded to reflect the reasoning made by clinicians when assessing gait, rendering the system easy to interpret. In addition, the symbolic representation of the signal allows for direct comparison of symbol sequences between subjects, or the evolution of symbol sequences for the same subject over time.

Only two accelerometers were used for the experiments because the envisioned system is to be embedded into the patient's shoes. Although the use of more sensors could improve the accuracy of the analysis, the final system would become too cumbersome for the intended application. When considering other applications, the method can support as many sensors as desired with small modifications to the "context analysis" phase. The intended application should also be taken into account when comparing the proposed method to previous works. The amount of information this method extracts from the signal is minimum compared to motion capture systems. However, it is as precise and more informative than other currently employed gait analysis systems such as the Gait Rite.

The symbolic approach explored here can be taken further and developed into a Human Activity Language (HAL) [26]. Previous approaches to HAL have used motion capture data or video images, which must be processed differently from accelerometer data. The proposed method narrows the gap between accelerometer based motion analysis and the HAL framework. In addition, most symbol-based methods are used for classification, whereas this work has achieved the detection of temporal events, and the characterization of gait symmetry.

VII. Conclusion

This paper presented a symbol-based method for analyzing gait from accelerometer signals. Compared to previous works, the proposed method performs equally well or better, and can also extract dynamic information from the signal.

The characterization of the signal in terms of its dynamics was here exemplified with a novel symbol-based symmetry index. Which can help identify gait disturbances that do not necessarily interfere with temporal measurements, such as compensatory adaptations in stroke patients. Future investigations include a clinical study which will validate the use of the proposed method for gait analysis, and the symbol-based symmetry index, on patient data.

Other contributions of this work include the use of expert knowledge in order to analyze symbolic context information. Which avoids the need for supervised learning techniques and large sets of training data. The addition of expert knowledge also contributes to creating an intuitive model, which can be easily interpreted and adapted to different applications.

The proposed method was here applied to acceleration data, however, this approach may be extended to other types of continuous sensor data. Once the characteristics of the system are understood, rules can be designed to detect relevant events in any type of time series. This method may also be seen as a compression technique. The symbolization of the signal together with its rule coding can greatly decrease the amount of data needed to express the information held in the original signal.

ACKNOWLEDGMENTS

The author would like to thank: Dr. Misha Pavel and Dr. Holly Jimison, for making the data collection possible; Dr. Fay Horak, for welcoming us into her lab; and Dr. Arash Salarian, for technical support.

REFERENCES

[1] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger, and J. Hausdorff, "Effect of gait speed on gait rhythmicity in Parkinson's disease: variability of stride time and swing time respond differently," *Journal of NeuroEngineering and Rehabilitation*, vol. 2, no. 1, p. 23, 2005.

[2] H. Cruz *et al.*, "Evidence of Abnormal Lower-Limb Torque Coupling After Stroke: An Isometric Study* Supplemental Materials and Methods," *Stroke*, vol. 39, no. 1, p. 139, 2008.

[3] O. Beauchet, G. Allali, G. Berrut, C. Hommet, V. Dubost, and F. Assal, "Gait analysis in demented subjects: interests and perspectives," *Neuropsychiatric Disease and Treatment*, vol. 4, no. 1, pp. 155–160, 2008.

[4] L. Nolan, A. Wit, K. Dudziñski, A. Lees, M. Lake, and M. Wychowañski, "Adjustments in gait symmetry with walking speed in trans-femoral and trans-tibial amputees," *Gait & Posture*, vol. 17, no. 2, pp. 142–151, 2003.

[5] K. Silver, R. Macko, L. Forrester, A. Goldberg, and G. Smith, "Effects of aerobic treadmill training on gait velocity, cadence, and gait symmetry in chronic hemiparetic stroke: A preliminary report," *Neurorehabilitation and Neural Repair*, vol. 14, no. 1, pp. 65–71, 200.

[6] M. J. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler, "Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement," *Physiological Measurement*, vol. 25, no. 2, pp. R1–R20, 2004.

[7] J.-A. Lee, S.-H. Cho, J.-W. Lee, K.-H. Lee, and H.-K. Yang, "Wearable accelerometer system for measuring the temporal parameters of gait," *29th Annual International Conference of IEEE-EMBS, Engineering in Medicine and Biology Society, EMBC'07*, pp. 483 – 486, 2007.

[8] J. B. J. Bussmann, L. Damen, and H. J. Stam, "Analysis and decomposition of signals obtained by thigh-fixed uni-axial accelerometry during normal walking," *Medical and Biological Engineering and Computing*, vol. 38, pp. 632–638, 2000.

[9] R. W. Selles, M. A. G. Formanoy, J. B. J. Bussmann, P. J. Janssens, and H. J. Stam, "Automated estimation of initial and terminal contact timing using accelerometers; development and validation in transtibial amputees and controls," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 13, no. 1, pp. 81 – 88, 2005.

[10] R. Williamson and B. J. Andrews, "Gait event detection for FES using accelerometer and supervised machine learning," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, no. 3, pp. 312–319, 2000.

[11] M. Sekine, T. Tamura, T. Fujimoto, and Y. Fukui, "Classification of walking pattern using acceleration waveform in elderly people," in *Engineering in Medicine and Biology Society, 2000. Proceedings of the 22nd Annual International Conference of the IEEE*, vol. 2, 2000.

[12] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," *Pervasive 2004*, pp. 1–17, April 2004.

[13] R. K. Ibrahim, E. Ambikairajah, B. G. Celler, and N. H. Lovell, "Time-frequency based features for classification of walking patterns," *Digital Signal Processing, 2007 15th International Conference on*, pp. 187–190, 2007.

[14] M. N. Nyan, F. E. H. Tay, K. H. W. Seah, and Y. Y. Sitoh, "Classification of gait patterns in the time-frequency domain," *Journal of Biomechanics*, vol. 39, pp. 2647–2656, 2006.

[15] A. Miller, "Gait event detection using a multilayer neural network," *Gait & Posture*, vol. 29, pp. 542–545, 2009.

[16] M. Thomas, J. Jankovic, M. Suteerawattananon, S. Wankadia, K. Caroline, K. Vuong, and E. Protas, "Clinical gait and balance scale (GABS): validation and utilization," *Journal of the Neurological Sciences*, vol. 217, no. 1, pp. 89–99, 2004.

[17] C. Toulotte, A. Thevenon, E. Watelain, and C. Fabre, "Identification of healthy elderly fallers and non-fallers by gait analysis under dual-task conditions," *Clinical Rehabilitation*, vol. 20, no. 3, p. 269, 2006.

[18] G. Yogev, M. Plotnik, C. Peretz, N. Giladi, and J. Hausdorff, "Gait asymmetry in patients with Parkinsons disease and elderly fallers: when does the bilateral coordination of gait require attention?," *Experimental Brain Research*, vol. 177, no. 3, pp. 336–346, 2007.

[19] K. Aminian, K. Rezakhanlou, E. De Andres, C. Fritsch, P. F. Leyvraz, and P. Robert, "Temporal feature estimation during walking using miniature accelerometers: an analysis of gait improvement after hip arthroplasty," *Medical and Biological Engineering and Computing*, vol. 37, pp. 686–691, 1999.

[20] A. Salarian, H. Russmann, F. J. G. Vingerhoets, C. Dehollain, Y. Blanc, P. R. Burkhard, and K. Aminian, "Gait assessment in parkinsons disease: Toward an ambulatory system for long-term monitoring," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 8, pp. 1434–1443, 2004.

[21] S. J. Crenshaw and J. G. Richards, "A method for analyzing joint symmetry and normalcy, with an application to analyzing gait," *Gait & Posture*, vol. 24, no. 4, pp. 515 – 521, 2006.

[22] H. Sadeghi, P. Allard, and M. Duhaime, "Functional gait asymmetry in able-bodied subjects," *Human Movement Science*, vol. 16, no. 2-3, pp. 243 – 258, 1997. 3-D Analysis of Human Movement - II.

[23] H. Sadeghi, P. Allard, F. Prince, and H. Labelle, "Symmetry and limb dominance in able-bodied gait: a review," *Gait & Posture*, vol. 12, no. 1, pp. 34 – 45, 2000.

[24] R. A. Miller, M. H. Thaut, G. C. McIntosh, and R. R. Rice, "Components of emg symmetry and variability in parkinsonian and healthy elderly gait," *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*, vol. 101, no. 1, pp. 1–7, 1996.

[25] C. S. Daw, C. E. A. Finney, and E. R. Tracy, "A review of symbolic analysis of experimental data," *Review of Scientific Instruments*, vol. 74, no. 2, p. 915, 2003.

[26] G. Guerra-Filho and Y. Aloimonos, "A language for human action," *Computer*, vol. 40, no. 5, pp. 42–51, 2007.

[27] P. Fihl, M. B. Holte, T. B. Moeslund, and L. Reng, "Action recognition using motion primitives and probabilistic edit distance," in *4th international Conference on Articulated Motion and Deformable Objects*, pp. 375–384, 2006.

[28] E. Keogh, "Data Mining in Time Series Databases Tutorial," in *Proceedings of the IEEE Int. Conference on Data Mining*, 2004.

[29] S. Petrovic, "A Comparison Between the Silhouette Index and the Davies-Bouldin Index in Labelling IDS Clusters," in *Proceedings of the 11th Nordic Workshop of Secure IT Systems*, pp. 53–64, 2006.

[30] A. McDonough, M. Batavia, F. Chen, S. Kwon, and J. Ziai, "The validity and reliability of the GaitRite system's measurements: A preliminary evaluation," *Archives of Physical Medicine and Rehabilitation*, vol. 82, no. 3, pp. 419–425, 2001.

[31] J. Cohan, *Statistical power analysis for the behavioral sciences*. New York: Academic Press, 1969.