# Probabilistic modelling of gait for robust passive monitoring in daily life

Yordan P. Raykov, Luc J.W. Evers, Reham Badawy, Bastiaan R. Bloem, Tom M. Heskes, Marjan J. Meinders, Kasper Claes, Max A. Little

***Abstract*— Passive monitoring in daily life provide valuable insights into a person's health throughout the day. Wearable sensor devices are play a key role in enabling such monitoring in a non-obtrusive fashion. However, sensor data collected in daily life reflect multiple health and behavior-related factors together. This creates the need for a structured principled analysis to produce reliable and interpretable predictions that can be used to support clinical diagnosis and treatment. In this work we develop a principled modelling approach for free-living gait (walking) analysis. Gait is a promising target for non-obtrusive monitoring because it is common and indicative of many different movement disorders such as Parkinson's disease (PD), yet its analysis has largely been limited to experimentally controlled lab settings. To locate and characterize stationary gait segments in free-living using accelerometers, we present an unsupervised probabilistic framework designed to segment signals into differing gait and non-gait patterns. We evaluate the approach using a new video-referenced dataset including 25 PD patients with motor fluctuations and 25 age-matched controls, performing unscripted daily living activities in and around their own houses. Using this dataset, we demonstrate the framework's ability to detect gait and predict medication induced fluctuations in PD patients based on free-living gait. We show that our approach is robust to varying sensor locations, including the wrist, ankle, trouser pocket and lower back.**

***Index Terms*— Gait modelling, health monitoring, passive monitoring, gait detection, medication prediction**

The authors wish to thank all participants for their enthusiasm to contribute to this study, and for welcoming us in their own homes. This work was funded by the Michael J. Fox Foundation for Parkinson's Research (grants 10231 and 17369), UCB Biopharma, ZonMw (grant 91215076 "Big Data for Personalised Medicine"), the Dutch Ministry of Economic Affairs (PPP Allowance made available by the Top Sector Life Sciences and Health to stimulate public-private partnerships: TKI-LSH-T2016-LSHM15022), and Sichting ParkinsonFonds.

Yordan P. Raykov is with the Department of Mathematics at Aston University, Birmingham, UK (email: y.raykov@aston.ac.uk).

Luc J.W. Evers is with the Center of Expertise for Parkinson and Movement Disorders, department of Neurology, Donders Institute for Brain, Cognition and Behaviour at Radboud University Medical Center, and with the Institute for Computing and Information Sciences at Radboud University, Nijmegen, Netherlands (email: Luc.Evers@radboudumc.nl).

Reham Badawy is with the School of Computer Science at University of Birmingham, Birmingham, UK (email: rehambadawy@hotmail.com).

Bastiaan R. Bloem leads the Center of Expertise for Parkinson and Movement Disorders, department of Neurology, Donders Institute for Brain, Cognition and Behaviour at Radboud University Medical Center, Nijmegen, Netherlands (email: Bas.Bloem@radboudumc.nl).

Tom M. Heskes is with the Institute for Computing and Information Sciences at Radboud University, Nijmegen, Netherlands (email: Tom.Heskes@ru.nl).

Marjan J. Meinders is with the Scientific Center for Quality of Healthcare (IQ healthcare), Radboud Institute for Health Sciences at Radboud University Medical Center, Nijmegen, Netherlands (email: Marjan.Meinders@radboudumc.nl).

Kasper Claes is with UCB Pharma, Brussels, Belgium (email: kasper.claes@ucb.com)

Max A. Little is with the School of Computer Science at University of Birmingham, Birmingham, UK and also with the Media Lab at MIT, Cambridge, US (email: maxl@mit.edu).

## I. Introduction

Ubiquitous consumer devices such as smartphones and wearables are equipped with low power inertial sensors such as accelerometers and gyroscopes capable of continuously recording their wearer's movements. In controlled laboratory settings, such sensors have been used successfully to measure symptoms of patients with various movement disorders, such as Parkinson's disease (PD) [1]. However, these measurements only provide a snapshot of the patient's condition, and may not be representative of the symptoms experienced in daily living conditions outside the lab, for example because of observer effects [2]. Unobtrusive wearable sensors enable us to monitor patients in daily life, which may provide patients, care providers and researchers with useful insights into the course of symptoms [3].

However, obtaining reliable and interpretable measurements in uncontrolled environments is difficult. One strategy has been to record the patient's ability to perform specific tasks (e.g. walk 10 meters) at different times of the day (*active tests*) [4]. An important limitation of active tests is that patients are interrupted in their daily activities during the tests, which can lead to high attrition in compliance [5]. Additionally, it is practically impossible to obtain a continuous view of symptom fluctuations using short active tests.

Instead of instructing patients to perform specific tasks, we could use daily routine activities that are affected by the patient's condition to measure how someone's symptoms fluctuate throughout the day (i.e *passive monitoring*). An important example of such activity is walking, otherwise known as *gait*. Many movement disorders are associated with alterations in gait patterns, and neurologists often use in-clinic gait examination to establish a diagnosis. PD-related changes in gait patterns consist of continuous impairments involving slowness and reduced arm swing (*bradykinetic gait*) and episodic hesitations to produce effective steps (*freezing of gait*). In many patients, bradykinetic gait is already present early in the disease [6] and is responsive to symptomatic

medication (e.g. levodopa) [7]. Therefore, measuring free-living gait could serve as a marker for disease progression and therapy-related symptom fluctuations in PD patients. This would allow for unobtrusive remote patient monitoring, and can potentially facilitate titration of medication, early diagnosis and evaluation of new drugs [8].

In order to extract reliable information about a patient's free-living gait, we need a principled framework to locate and summarize gait segments. Important challenges are:

- The scarcity of suitable reference datasets: most available methods are trained and evaluated using labelled data from a pre-defined, and sufficiently distinguishable set of scripted activities, often collected in controlled environments. This does not reflect free-living conditions, where much more variation is present due to environmental and behavioural factors. This means that we need labelled training data that better reflects real-life variation.
- On the other hand, we need to acknowledge that it remains infeasible to capture all real-life variation in training datasets. Highly flexible, supervised systems can have unpredictable behavior as data shifts outside the training distribution. Therefore, the free-living setting asks for principled, interpretable models which can produce parsimonious signal representations, while accounting for distributional changes and uncertainty in our data.

In this work, we propose a unified framework for gait detection and gait pattern analysis. We have combined common characteristics used for gait analysis into a principled probabilistic graphical model, which can be directly applied to the accelerometer data. We adopt a flexible *nonparametric* model which can locate different gait and non-gait activities that vary both in terms of their statistical and temporal characteristics. Specifically, we use a set of high order *autoregressive* (AR) *processes*. The AR process is a parametric model of the frequency spectrum, hence it directly captures characteristics derived from the *power spectral density* of the data. At the same time, AR processes are time domain models which allows us to couple them with a *nonparametric* hidden Markov model (HMM) leading to an AR-iHMM also known as a *nonparametric switching AR process* [9] to capture changes in behavior patterns and gait types in free-living conditions.

To demonstrate the applicability of this analytical framework, we used a new, unique dataset consisting of sensor data from various wearables and concurrent reference video annotations, collected during unscripted daily living activities in and around the homes of 25 PD patients with motor fluctuations, and 25 age-matched controls. Using this dataset, we show that the proposed AR-iHMM can be used in free-living conditions to accurately detect healthy and pathological gait across different sensor wear locations. Furthermore, we show that the model can identify changes in gait pattern after intake of dopaminergic medication in individuals with PD.

## II. Related work

In the last two decades, advances in wearable sensors have made it feasible to unobtrusively monitor patients outside controlled laboratory conditions, allowing us to study real-life gait patterns. However, to successfully deliver on that promise, we need tools which can reliably and robustly model data recorded from wearables in this setting. Here we review relevant prior work in terms of wearable sensor devices, gait detection algorithms, and gait characteristics under study.

### A. Wearable sensor devices

Because of its simplicity, robustness and affordability, the 3-axis accelerometer is by far the most widely used sensor for free-living gait analysis. The accelerometer sensor measures the vector sum of all sources of acceleration acting on the device in each spatial direction. The unit of measurement is $m/s^2$ and if the device is not under other sources of acceleration, the only acceleration measured by the device is due to the force of gravity (zero magnitude under free-fall).

Sensor devices can be worn on various body locations, including the trouser pocket, the lower back, the shin or ankle, the shoe, as well as the wrist. The choice of device location is influenced by the expected gait detection accuracy, the type of gait characteristics that can be reliably estimated, patient acceptance, and the commercial availability of devices. An extensive review of widely-used wearable devices and their sensors for gait analysis can be found in Tao et al. [10], and a focused review on sensor placement for monitoring of PD can be found in Brognara et al. [11].

There is no consensus on the best device location to detect and characterize the gait of PD patients, and whether there is added value in combining multiple locations. Therefore, we evaluate our proposed framework on various commonly used sensor locations. Another concern can be the limited commercial availability and high costs of "research-grade" devices. For this reason, we include a consumer smartphone in our comparison, which is widely available and relatively low-cost.

### B. Gait detection algorithms

Most gait detection techniques rely on parametric assumptions about the spectral density, time domain distribution or both [12]. Typically, features are extracted from windows of fixed width, and the decision to classify a window as gait or non-gait behavior is made using pre-defined thresholds or using a trained classifier.

Various features have been been used for gait detection. One of the most widely used methods for identifying gait is based on the standard deviation (STD) of a windowed accelerometer signal [13]. An alternative, and similarly popular approach is the window-based analysis of spectral features [10]. Gait is typically highly periodic with Nyquist bandwidth of 10-15Hz [14]. This has motivated the use of the *short-time Fourier transform* (STFT) to detect gait. For example, Sama et al. [15] studied the energy of the accelerometer signal in 800 different frequency bands. They applied Relief feature selection to identify the energy bands that are most descriptive of gait. Karantonis et al. [16] suggested directly analyzing the Fourier coefficients of the z-axis on the accelerometer to look for sufficient power at the expected range of walking frequencies (0.7–3.0 Hz). The time-frequency resolution issues of STFT-based walking detection have sometimes been

addressed using wavelet transforms. Continuous wavelet transforms often require large computational effort, but discrete wavelet transforms can be used to efficiently estimate high quality features of gait [17], more efficiently even compared to Fourier transform [18, page 254]. We can also encode the power spectrum directly in the time domain if we use windowed auto-correlation [19] and then use the values at a subset of time lags corresponding to the duration of the gait cycle [20]. Alternatively, a *stride template* can be formed offline and online similarity to the template be determined (e.g. via cross-correlation [21] or dynamic time warping [22]).

A problem with these different window-based feature extraction methods is that signals acquired in daily life are highly non-stationary. When these non-stationarities occur within a window, for example, the transition from standing to gait, they may reduce the usefulness of the extracted features, particularly in the case of STFT (as we will further discuss in Section IV).

Gait detection systems not only vary in the features they rely on, but also in the classification algorithm they use. Support vector machines and random forest classifiers are commonly trained on window-based features [23], [24]. In addition, HMMs have also been used to detect gait based on window-based features, which offers the advantage of incorporating the sequential nature of human behavior [25]. Haji et al. [25] demonstrated that, in more challenging settings, a hierarchical HMM significantly improves gait detection compared to, for example, peak detection and dynamic time warping. More recently, generic activity recognition pipelines based on deep learning methods [26], [27] are being introduced, although the scarcity of labelled free-living data currently limits their practical use.

Despite the heterogeneity in gait patterns, gait detection is generally treated as a binary classification problem (gait/non-gait). Whereas this may be appropriate to globally describe how much users walk, problems can emerge when it is used as a starting point for evaluating the quality of gait in medical applications; these systems group all gait together, regardless of changes in the gait pattern that can occur even within the same gait segment (e.g. because of changes in symptoms, pace or environment). As a result, the detected gait segments are likely heterogeneous and non-stationary, which can be problematic for subsequent gait pattern analysis (as we will further discuss in Section IV).

### C. Characterization of the gait pattern

Once gait episodes have been identified, studies have used various approaches to characterize the gait pattern in movement disorders such as PD. Many studies try to identify important events of the gait cycle, including the heel strike or initial contact (IC), and final contact (FC) of both feet. Several variations to peak detection have been used for this, which may benefit from pre-processing the acceleration signal using continuous wavelet transforms (CWT) [28]. The timing of IC and FC events is then used to compute temporal gait features such as step time, swing time, stance time, and double support time. Additionally, based on assumptions about the exact sensor positioning and the biomechanics of gait, location-specific algorithms can be used to estimate spatial gait features. For example, having identified the ICs and FCs, one can use the inverted pendulum model to estimate the step length from the accelerometer signal of a sensor on the lower back [29]. Del Din et al. [30] used this approach and showed that free-living gait analysis discriminated better between PD patients and healthy controls than lab-based gait analysis, which illustrates the potential of free-living gait analysis. Moore et al. [31] suggested that the step length estimated using an ankle sensor could be used to track the free-living gait pattern of PD patients, but only included three PD patients monitored over 24 hours in an apartment-like setting.

Other approaches focus on analyzing the periodicity of the accelerometer signal during gait, either based on the PSD or auto-correlation in the time domain. An advantage of these methods is that they are less dependent on location-specific assumptions, compared to identifying gait cycle events and computing the step length. For example, Weiss et al. [32] computed the width of the dominant frequency in the PSD during free-living gait (based on the accelerometer signal from a lower back sensor), and demonstrated that it could be used to predict future falls in patients with PD. Similarly, Rispens et al. [33] computed the PSD during free-living gait based on a lower back accelerometer, and showed that the spectral power in the lower frequencies, and the amplitude and slope of the dominant frequency, were related to the number of falls in older adults. Pérez-López et al. [34] combined the identification of ICs with analysis of the PSD during individual strides, and showed that the power in the gait range (based on a waist accelerometer) was correlated to changes after medication intake in PD patients. Bellanca et al. [35] suggested that the harmonic ratio (ratio of the sum of the amplitudes of the even and uneven harmonics, computing over the PSD of a single stride) could be used as a measure of step symmetry. Alternatively, the periodicity of free-living gait can also be analyzed in the time domain, for example by estimating the auto-correlation [36].

All analyses mentioned in this paragraph strongly depend on accurate localization of stationary gait segments, which may be sub-optimal given current gait detection algorithms. In this work, we propose that free-living gait analysis can be improved by employing a unified approach to gait detection and gait pattern characterization.

## III. Free-living data collection

To allow for a more realistic understanding of the challenges of modelling free-living gait data, and to evaluate our proposed model, we have used a new reference dataset from the Parkinson@Home validation study [37]. This study includes sensor data and video recordings during uninterrupted and unscripted daily life activities in the participants' natural environment. For a detailed description of the study design and participants, we refer to Evers et al. [37]. The de-identified dataset will be made available to the scientific community in collaboration with the Michael J Fox Foundation. In brief, both patients with Parkinson's disease with motor fluctuations (PD group)

and 25 age-matched participants without PD (non-PD group) were recruited. Inclusion criteria for both groups consisted of: (1) age 30 years or older and (2) in possession of a smartphone running on Android OS version 4.4 or higher. Additional inclusion criteria for participants in the PD group were: (1) diagnosed with PD by a neurologist, (2) receiving treatment with dopaminergic medication (levodopa and/or dopamine agonist), (3) experiencing motor fluctuations (MDS-UPDRS item $4.3 \geq 1$), and (4) known to have PD-related gait abnormalities, i.e. bradykinetic and/or freezing of gait (MDS-UPDRS item $2.12 \geq 1$ and/or item $2.13 \geq 1$). PD patients who received advanced treatment (deep brain stimulation and/or intestinal infusion of levodopa or apomorphine) were excluded.

Participants were visited in their own homes and each visit included a standardized clinical assessment (full MDS-UPDRS [38] and AIMS [39]) and an unscripted free-living assessment of at least one hour. To ensure indicative behaviors such as longer gait episodes were captured, assessors encouraged participants to include these in their routines. Participants in the PD group were asked to skip their morning dose of dopaminergic medication before the visit, so that they were in the OFF medication state at the start of the visit. After the MDS-UPDRS part III (motor examination) and free-living assessment were conducted in the OFF state, participants took their usual medication and the full MDS-UPDRS, AIMS and free-living assessment were performed in the ON state, i.e. with the symptomatic effects of medication present.

During the full visit, participants wore various light-weight sensors on different body locations. In this study, we used the accelerometer data from the smartphone worn in the front trouser hip pocket (collected using the HopkinsPD app [40]; all participants were instructed to wear trousers with a front pocket), and the accelerometer data from Physilog 4 devices worn on both ankles, both wrists and the lower back. To allow for time synchronization, all devices were triggered together (hit ten times against a table) in front of the video camera at the beginning and end of data collection.

The video recordings during the free-living assessments were annotated by a research assistant, who labeled as "gait" any activity that involved at least 5 consecutive steps, with the exception of any running episodes.

## IV. Challenges of modelling free-living gait

Analysis of free-living gait is challenging because accelerometer data simultaneously reflects disease symptoms, behaviour, device orientation, sensor location and environment. This makes it difficult to design a reliable analytical pipeline which untangles these factors and allows us to focus solely on representative aspects of the gait that are relevant for monitoring PD. Before we introduce our proposed model, we first highlight some of the common *estimation* challenges which we aim to address. We use examples from the unscripted free-living assessments of the Parkinson@Home validation study.

### A. Data filtering and accounting for orientation

Accelerometers measure any forces due to accelerations which partly prevent the device from free-fall in the Earth's gravitational field. If we are interested in monitoring gait, however, we first need to remove this field effect from the raw accelerometer data, as irrelevant device rotations (e.g. slight variations in the attachment of the sensor) may otherwise confound any inferences we make about a person's gait. This analytical step is most commonly done using fusion of data from a magnetometer, gyroscope and accelerometer [41], or simply using a digital low pass filter [42] applied to the accelerometer signal. Sensor fusion is well justified and commonly used for estimation of more complex tasks such as sensor positioning and heading, where it outperforms techniques relying on numerical integration of gyroscopes [43]. However, due to the inherent smoothness assumption in the Kalman filter typically used for the fusion, estimates might be biased during abrupt changes [43]. Different methods of sensor fusion, such as one that relies on $l_1$-regularization, can address this problem, but in this work we opt for using a single sensor approach. Low pass filters can be used in an accelerometer only setup, but they are poorly justified, since orientation changes can have a broad bandwidth leading to unwanted distortions in the time domain depending on the cut off frequency of the filter. In this work we opt for a piecewise $l_1$- trend filter as motivated in Badawy et al. [44] which assumes that changes due to orientation are piecewise linear [45].

The accelerometer data we use in any subsequent analysis, is pre-processed by interpolating to a uniform sample rate [1] (i.e. using cubic spline interpolation), applying the $l_1$- trend filter to each individual axis and computing the magnitude of acceleration according to $\sqrt{a_x^2 + a_y^2 + a_z^2}$.

### B. Parsimonious representation of gait data

We have seen in Section II that most pipelines for analysis of gait data involve windowing of sensor data and estimation of statistical or spectral features. The estimated feature values are then used to make inferences about the behaviour monitored at that point in time (i.e. gait vs non-gait) or the gait pattern. However, in free-living the variability of these features is large. For example, in Figure 1 we show how much the window standard deviation differs for both gait and non-gait classes, even within a single individual. To reduce some of this variability, we tend to aggregate feature values (i.e. across time, across individuals, across similar behaviours). The way we make such aggregation will inevitably affect the quality of the inferences we make.

Let us consider the following example: we have 10 minutes of consecutive gait data from a PD patient where the gait varies significantly across different segments. In Figure 2 we plot how much the 1 second window standard deviation varies in time and how this feature variation can be reduced by smoothing through time. The underlying assumption, which is commonly made, is that feature values collected closely in time should be similar (i.e. change smoothly). However,

[1] Smartphone data is sampled at non-uniform rate of 50-150Hz. The wrist-worn, shin-worn and lower back Physilog devices output uniformly sampled data at rate of 200Hz. All accelerometer data is processed in meter per second squared.

behaviour can change abruptly, hence this assumption does not hold. In contrast, we also display the same feature, but now conditioned on stationary segments of variable length. This results in a more compact representation of this type of data that enables simple inference, while still preserving the important signal characteristics.

### C. Spectral estimation challenges

A similar argument can be made for features based on the spectrum. The support at different frequencies also exhibits large variation both within gait and non-gait classes, as shown in Figure 3. Similar to standard deviation, more stable spectral estimation can be done by aggregating across neighbouring windows, e.g. using Welch's overlapped averaging power spectral density (PSD) estimator [18], [46, Section 7.4]. The problem is that this still assumes that the signal is stationary across windows, which is often not the case in free-living data because of the abrupt changes in behaviour. Even within one gait segment, the characteristics of the gait pattern can abruptly shift due to intentional changes (e.g. turning, starting to make gestures), environment (e.g. changing walking terrain), and PD symptoms (e.g. hesitations to walk through doorways). Figure 4 displays an example of this. We show that the Welch PSD associated with each of the two gait patterns, varies significantly from the Welch PSD estimated when grouping both gait patterns together. This underlines that if we condition on piecewise stationary segments, we obtain more useful estimates of the spectrum.

An additional problem that arises with estimating Fourier features in free-living is that accurate estimation of the spectrum rests on the assumption of *periodic continuation* [18]. Because of common non-stationarities in free-living (e.g. mentioned changes within gait episodes, but also the start and end of gait episodes), violations of this assumption are common when using fixed size windows. This can lead to spurious spectral artifacts, for example caused by Gibbs phenomenon [18]. Typically, these issues are ameliorated by using other window functions than the rectangular window, such as the Hanning window. However, while windowing matches samples at window edges (by zeroing), it also distorts the waveform because it causes *amplitude modulation*.

In conclusion, the usefulness of spectral estimates largely depends upon accurately locating stationary segments in time, i.e by accurately detecting the start and end of gait episodes, and by detecting (abrupt) changes within gait episodes. At the same time, doing this depends on having access to spectral estimates. Because of this interdependence, we propose a unified framework that addresses both these problems simultaneously.

## V. Probabilistic Modelling of Gait

Whereas most systems focus on segmenting accelerometer data into gait vs. non-gait classes, we first segment the data into multiple different groups (more than two) and afterwards assign these groups to gait or non-gait class. We do this efficiently by designing a flexible, probabilistic model which is trained directly on the magnitude of acceleration obtained after removing piecewise linear device orientation changes (see Section IV).

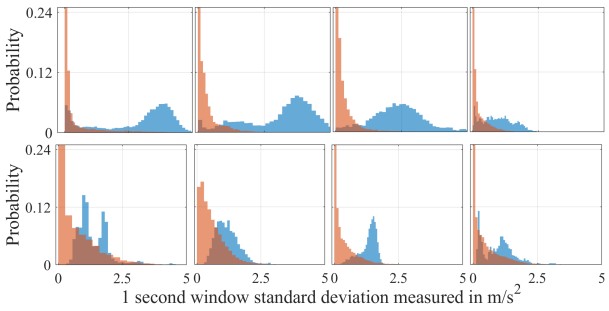


Fig. 1. Histograms of 1 second window standard deviation of the magnitude of acceleration from different PD patients during the unscripted free-living assessment, collected using a smartphone placed in the front trouser pocket (top row) and a wrist-worn device (bottom row). The horizontal axes show the window standard deviation, and the vertical axes show the normalized bin counts. Blue: gait, orange: non-gait (according to video annotations).

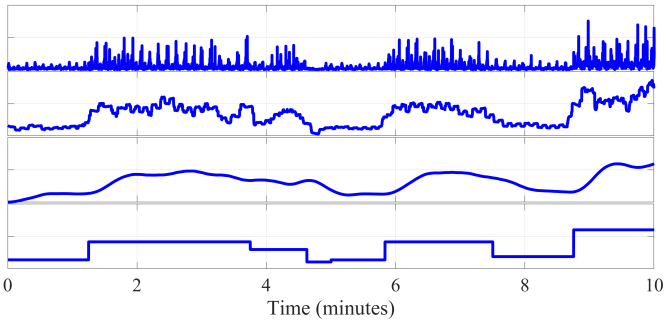


Fig. 2. Illustrative example of feature smoothing over accelerometer data during unscripted gait. The top panel displays the magnitude of 10 minutes of pre-processed accelerometer data collected using a smartphone placed in the front trouser pocket. In the second panel, we display the fixed size (1 sec) window standard deviation. In the third panel, we show the smoothed feature values using standard moving average in time (5 sec). The bottom panel displays the standard deviation, computed over approximately stationary segments of variable length (determined visually).

### Autoregressive modelling of gait

The first assumption we make is that the repetitiveness of the gait cycle (heel strike, midstance, heel off, midswing, heel strike) is one of the key properties that characterize gait episodes. The periodic nature of the accelerometer data during gait [36] makes it efficient to detect and model gait based on the spectrum, for example using the Fourier transform. As discussed, Fourier spectral analysis inherently assumes periodic continuation (see Section IV). So we address this problem by simultaneously estimating the spectrum and the start and end points of the stationary gait episodes. To achieve this, we first model the spectrum of the gait in the time domain, using *autoregressive (AR) processes* [18]. An order $r$ AR model is a random process which describes a sequence $\{x_t\}_{t=1}^{T}$ as a linear combination of previous values in the sequence and a stochastic term:

$$x_t = \sum_{j=1}^{r} A_j x_{t-j} + e_t \qquad e_t \sim \mathcal{N}\left(0, \sigma^2\right) \tag{1}$$

where $A_1, \ldots, A_r$ are the AR coefficients, $T$ denotes the length of the sequence, and $e_t$ is a zero mean random variable, assumed to be an i.i.d. Gaussian sequence (we can trivially

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2020.3037857, IEEE Journal of Biomedical and Health Informatics



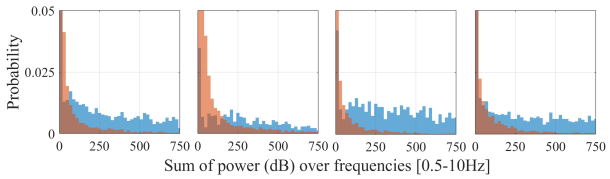


Fig. 3. Histograms of the spectral energy at typical gait frequencies (0.5 - 10Hz) obtained using STFT with window length 1 second. Each subplot displays the feature distribution for the unscripted free-living assessment of a single PD patient, from a smartphone worn in the front trouser pocket. The horizontal axes are different total energy values, and the vertical axes show the normalized bin counts. Blue: gait, orange: non-gait (according to video annotations).

extend the model such that $e_t \sim \mathcal{N}\left(\mu, \sigma^2\right)$ for any real-valued $\mu$). We assume that the AR noise variance $\sigma^2$ is unknown and place a conjugate *inverse-Wishart* prior over it. This essentially means that in addition to modelling the periodicity of the input signals, we also account for changes in the non-periodic components of the signals. We saw in Section IV that the window variance of the acceleration can be a useful discriminator of gait versus non-gait on its own in certain scenarios. If we assume an AR model of order $r = 0$, the variance of $e_t$ is the variance of the window.

AR processes are commonly used as parametric models of the PSD since the power spectrum is determined by the AR parameters [18]:

$$S\left(f\right) = \frac{\sigma^2}{\left|1 - \sum_{j=1}^{r} A_j \exp\left(-i2\pi f j\right)\right|^2} \tag{2}$$

where $f$ is the frequency variable and $i$ denotes the imaginary unit. This means that the number of non-zero AR coefficients determines the complexity of the PSD which the model can represent: there is a peak in the PSD for each complex-conjugate pair of roots of the coefficient polynomial. Parametric spectral estimation is often more stable than non-parametric PSD methods, and can be of high quality using fairly little data, assuming the model is correct. The parametric model of the spectrum will allow us to construct a flexible, non-parametric model of the switching dynamics of different gait and non-gait activities in free-living. More detailed discussion on the relative merits of different spectral estimation methods combined with machine learning, can be found in Little [18].

### High order adaptive autoregressive processes

As mentioned above, the AR order $r$ we use will determine the complexity of this parametric model of the spectrum. The optimal AR model $r$ is likely to vary across different stationary segments of sensor data and choosing fixed $r$ which is too large will lead to problems with parameter estimation (fitting the AR coefficients). At the same time, gait is typically characterized by a low fundamental frequency, with bandwidth of up to 10-15Hz (see Section II). This implies the need for fairly high order $r$ AR processes (together with sufficiently high sample rate) in order to accurately capture the typical range of gait frequencies. To address this conflict, we use a non-conjugate Bayesian prior on the AR coefficients $A_1, \ldots, A_r$ which induces *sparsity* of the coefficients (only a few are non-zero at any one time). This allows us to draw conclusions about the AR coefficients that do not contribute to the underlying dynamics of the gait. In effect, this means that we attempt to learn fewer than $r$ AR coefficients supported by the signal but potentially associated with larger AR time delays. This is done by assuming independent, zero-mean Gaussian priors on the coefficients $A_1, \ldots, A_r$ with unknown precisions, which acts as an *automatic relevance determination prior* (ARD) [47]. The ARD prior was first proposed in the context of neural network models in Mackay [47] and then later adopted for switching AR processes in Fox et al. [48].

### Latent switching behavior dynamics

To analyze free-living data, it is insufficient to define a parametric spectral model for the patients' gait, because participants regularly switch between different gait and non gait episodes, which results in highly non-stationary time series (see Figure 4).

Even within gait episodes, the optimal AR parameters to model the gait might change depending on the speed, amplitude and other characteristics of the walking pattern. In order to group similar gait signals, but also separate gait from non-gait data, we use a *switching AR process* model (AR-HMM). However, one drawback of conventional switching AR processes, is that it requires a fixed number of hidden states and AR order. Since the heterogeneity in both gait and non-gait episodes will increase as more free-living data becomes available, we adapt the more flexible *non-parametric switching AR process* first proposed in Fox et al. [9]. The model can be thought of as an infinite-state extension of the *switching AR process* (hence we refer to it as AR-iHMM). Viewing the switching AR model as a hidden Markov model (HMM) with AR processes used to model the HMM emissions, then in the non-parametric switching AR model the parametric HMM is effectively replaced with an *infinite* HMM [49].

In the AR-iHMM model, we assume that the data is an inhomogeneous stochastic process and that multiple AR models are required to represent the dynamic structure of the signal, i.e.:

$$x_t = \sum_{j=1}^{r} A_j^{z_t} x_{t-j} + e_t^{z_t} \qquad e_t^{z_t} \sim \mathcal{N}\left(0, \sigma_{z_t}^2\right) \tag{3}$$

where $z_t \in \{1, \ldots, K^+\}$ indicates the AR model associated with time index $t$. The latent variables $z_1, \ldots, z_T$ describing the switching process are modelled with a Markov chain. A transition matrix $\pi$ is estimated with $K^+$ rows and $K^+ + 1$ columns indicating the probability of specific transitions from existing state $i$ to existing state $j$, $\pi_{ij}$, or from existing state $i$ to a new state $K^+ + 1$, $\pi_{iK^+ + 1}$. Transitions that are observed more often during the training of the model will have higher probability, represented in the transition term $\pi_{ij}$.

When $K^+ \ll T$, this model clusters together parts of the signal into an, a priori, unknown number $K^+$ of time segments which are best represented with the same AR coefficients. In AR-iHMM, $K^+$ is unknown: instead of being fixed it is inferred from the data and can adapt to new, unseen structure

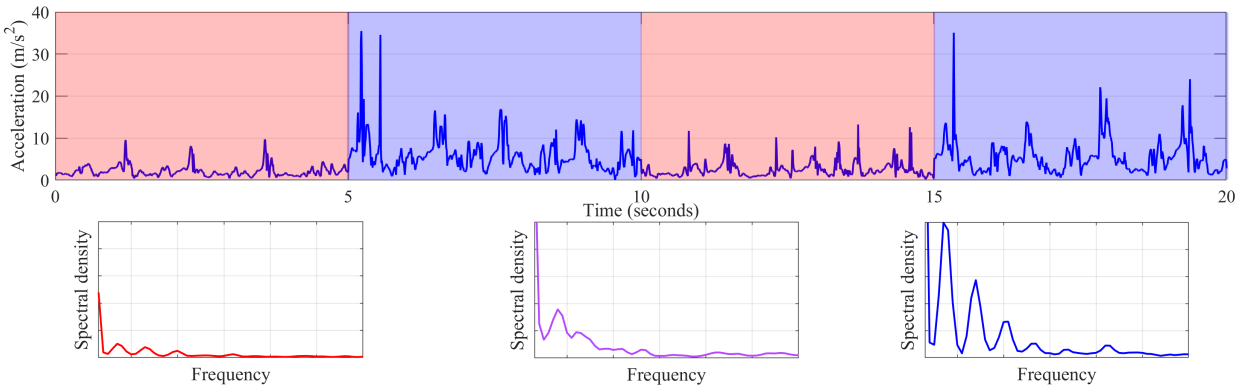


Fig. 4. Illustrative example of estimating the power spectral density over an unscripted gait segment that contains switches between two different gait patterns (i.e. is approximately piece-wise stationary). The top panel displays the signal magnitude of 20 seconds of pre-processed gait data from a PD patient, obtained from a smartphone worn in the trouser pocket. The red and blue shading indicates different gait states. The bottom panels display the Welch's PSD estimates: for data from the red gait state (left); for data from the blue gait state (right); for an equal amount of data from both (middle). To allow for same resolution in the 3 bottom plots, we have used 20 seconds of data for each plot.

in the data. The AR-iHMM is obtained by augmenting the transition matrix of the Markov process $\pi$ underlying the latent variables $z_1, \ldots, z_T$ with a *hierarchical Dirichlet process* (HDP) [50] prior.

## VI. Results: Gait Detection

In order to make inferences about changes in the gait pattern, we first need to verify that our proposed framework is able to accurately identify gait segments. In this section, we evaluate our ability to detect gait as annotated in the video recordings of the Parkinson@Home validation study, using the pre-processed accelerometer data (see Section IV) from the smartphones and Physilog 4 devices placed on various body locations (see Section III). To establish whether our approach achieves satisfactory results, we include a comparison with some of the most widely used gait detection algorithms. It is important to note that our goal was not to perfectly reproduce the manual (imperfect) video annotations, but to locate stationary gait segments in time, that can be used to make inferences about the effect of PD on the gait pattern. For example, since our method exploits the high periodicity of gait, we can expect that it is less suitable to detect short, irregular gait segments. However, this is actually a desirable property because we expect that the longer, more "steady-state" gait segments are most useful to quantify bradykinetic gait in PD patients.

### A. Model based gait detection

We infer the AR-iHMM described in Section V using scalable iterative MAP inference proposed in Raykov et al. [51]. Any hyperparameters associated with the AR state priors or the HDP prior (see Section V) are fixed across patients and are selected using standard Bayesian model selection. For each point $x_t$, we consider it is associated with its most likely state $z_t = k*$ to enable direct comparison, i.e. we ignore the estimated uncertainty associated with the segmentation indicators.

To determine if the identified hidden Markov states should be classified as gait or non-gait, we consider the AR-based PSD estimates associated with each state. Specifically, we compute the total energy at frequencies in the range [0.5 - 10Hz], and select a threshold of minimal spectral energy that maximizes the balanced accuracy (average of sensitivity and specificity) averaged across participants (measured against the manual video annotations for the presence of gait). We evaluate the performance of selecting the threshold using leave-one-subject-out cross-validation. Thresholding using a shared PSD range across participants is done only to enable a fair and intuitive comparison with the other commonly used techniques for detection of gait in smartphones and wearables; in principle, once the AR-iHMM model is trained we can derive multiple features related to the distribution of the sensor data and train a supervised classifier on these features.

### B. Implementation of existing gait detection algorithms

For the comparison with existing algorithms, we implemented methods that rely on leveraging one or two intuitive, window-based features from the time and frequency domain, to separate gait from non-gait classes: STD-thresholding [12], [13]; STFT-thresholding; normalized autocorrelation step detection and counting (NASC) [19] and continuous wavelet transform (CWT) thresholding [52]. We evaluate the performance of the original formulations of the algorithms, and the performance after applying our pre-processing pipeline and adjusting thresholds to maximize the balanced accuracy across participants using leave-one-subject-out cross-validation[2]:

- STD-thresholding: we set a threshold based on the 1 second window standard deviation to maximize the balanced accuracy averaged across participants;
- STFT-thresholding: we set a threshold based on the 1 second window total energy at frequencies in the range [0.5 - 10Hz] to maximize the balanced accuracy averaged across participants;
- NASC algorithm: the NASC involves first applying STD-thresholding and then evaluating the auto-correlation of the remaining data over 2 second windows, specifically

[2]Different thresholds are used for PD and controls cohorts.

TABLE I

GAIT DETECTION PERFORMANCE OF THE PROPOSED AR-IHMM AND OF COMMON GAIT DETECTION ALGORITHMS (USING THE THRESHOLDS REPORTED IN THE LITERATURE, AND AFTER PRE-PROCESSING AND OPTIMIZING THRESHOLDS). WE HAVE COMPUTED THE AVERAGE PERFORMANCE AND STANDARD DEVIATION USING LEAVE-ONE-SUBJECT-OUT CROSS-VALIDATION. FOR PD PATIENTS, WE SHOW THE PERFORMANCE OF THE COMPLETE FREE-LIVING ASSESSMENTS, AND THE DIFFERENCE IN BALANCED ACCURACY BETWEEN THE PARTS BEFORE AND AFTER MEDICATION INTAKE.

| Method | PD | | | Control | |
|---|---|---|---|---|---|
| | Sensitivity | Specificity | Accuracy difference before/after medication | Sensitivity | Specificity |
| *With fixed thresholds* | | | | | |
| STD-thresholding | 77% (15%) | 93% (4%) | 5% | 90% (6%) | 96% (4%) |
| STFT-thresholding | 85% (12%) | 94% (4%) | 3% | 86% (6%) | 85% (6%) |
| NASC | 87% (9%) | 96% (4%) | 3% | 88% (6%) | 97% (4%) |
| CWT-thresholding | 66% (10%) | 97% (5%) | 5% | 70% (4%) | 97% (4%) |
| *With optimized thresholds and pre-processing* | | | | | |
| STD-thresholding | 90% (14%) | 91% (4%) | 5% | 93% (5%) | 92% (3%) |
| STFT-thresholding | 91% (11%) | 89% (4%) | 3% | 92% (5%) | 91% (4%) |
| NASC | 93% (10%) | 87% (5%) | 3% | 94% (5%) | 91% (3%) |
| CWT-thresholding | 92% (8%) | 85% (5%) | 5% | 94% (3%) | 85% (6%) |
| **Probabilistic modelling (based on our AR-iHMM)** | **91% (8%)** | **91% (3%)** | **1%** | **94% (5%)** | **93% (3%)** |

looking at the time delays representative of gait. We set a modified STD threshold, a range of delays, and an auto-correlation threshold to maximize the balanced accuracy averaged across participants (iteratively, one at a time);

- CWT-thresholding: we compute the ratio between the energy in the band of walking frequencies and the total energy across all frequencies, and set a threshold to maximize the balanced accuracy averaged across participants.

Our comparison omits some previously proposed detection algorithms because (1) they are based on heuristics which could not be trivially adapted for detection of pathological gait; (2) they had strong conceptual overlap with the techniques included in the comparison; (3) they demonstrated very poor performance on our dataset. In addition, our comparison does not include an evaluation of deep learning activity recognition and gait recognition pipelines, or deep autoencoder features which can approximate arbitrarily complex mapping functions [53]. Although we acknowledge that such methods may achieve marginally better accuracy, we recognise that our study is based on a limited number of participants and it is easy to overfit the free living data. Moreover, our focus has been on deriving interpretable representation of the data and gait clustering. Therefore, we effectively compared linear thresholding of different properties of the data, estimated by a window-based approach versus estimated from the inferred model.

### C. Results of comparison

In Table I we report the different performance measures of the tested methods when applied to the accelerometer data from smartphones. In Table II we also report the balanced accuracy of all algorithms on accelerometer data from different body locations (see Section III).

First of all, the results in Table II show that it is feasible to identify gait using our modelling approach, with at least as good average performance compared to existing algorithms. In addition, the results underline the importance of appropriate pre-processing and threshold adjustment, in particular when applying algorithms to patients with PD, as indicated by the significant change in performance.

In most methods, we observed a difference in accuracy between PD patients and controls, and between before and after medication intake for PD patients. The latter was most notable in patients with a strong response to medication. This difference in accuracy between before and after medication intake was less prominent for the AR-iHMM, which also demonstrated less variability in the performance across PD patients. Moreover, the performance of the AR-iHMM was relatively robust to different body locations of the sensor in comparison to STD-thresholding, NASC, and CWT-thresholding (Table II).

It is worth noting that the prevalences of the gait and non-gait classes are not balanced in the free-living assessments from the Parkinson@Home validation study. Across PD patients, the mean walking time is 16% with a standard deviation of 6%. Most patients had longer walking episodes between 7 and 20 minutes before and one after medication intake, combined with many shorter walking bouts. The mean walking time is slightly higher for non-PD controls at 21% with the same standard deviation. Because we cannot assume that this prevalence is representative of truly free-living situations, we choose to evaluate the methods with measures that are independent of the prevalence of gait (i.e. sensitivity and specificity). However, the thresholds were set to optimize the balanced accuracy (mean of sensitivity and specificity), which implicitly optimizes for the situation where class prevalences are equal, and misclassification costs are equal as well. Different applications may require different settings of the thresholds to achieve a different trade-off between false-positives (with the risk of using non-gait data to predict medication induced fluctuations, which has no basis) and false-negatives (with the risk of identifying too few gait segments which limits our ability to track gait fluctuations throughout the day).

TABLE II

PERFORMANCE OF DIFFERENT METHODS FOR GAIT DETECTION ACROSS DIFFERENT SENSOR LOCATIONS (AFTER PRE-PROCESSING AND OPTIMIZING THRESHOLDS). THE PERFORMANCE IS EXPRESSED AS BALANCED ACCURACY, EVALUATED ON THE COMPLETE FREE-LIVING ASSESSMENTS OF PD PATIENTS AGAINST VIDEO ANNOTATIONS.

| Method | Left Shin | Right Shin | Left Wrist | Right Wrist | Lower Back | Smartphone |
|---|---|---|---|---|---|---|
| STD-thresholding | 94% | 94% | 77% | 76% | 91% | 90% |
| STFT-thresholding | 93% | 92% | 83% | 81% | 90% | 90% |
| NASC | 91% | 91% | 78% | 77% | 92% | 90% |
| CWT-thresholding | 92% | 90% | 75% | 72% | 90% | 89% |
| **Probabilistic modelling** | **94%** | **94%** | **83%** | **83%** | **93%** | **91%** |

## VII. RESULTS: MODELLING GAIT PATTERN CHANGES

An important potential application of free-living gait analysis in PD patients is monitoring real-life variations in the response to medication. The non-parametric nature of our approach allows us to identify significant statistical changes in the gait distribution which can be used to locate potential clinical changes in the gait pattern. In this section, we report the effectiveness of our probabilistic model for the problem of classifying gait episodes into "before medication" and "after medication" classes. We have compared the binary classification accuracy using different gait features and different device locations (described in Section III).

### A. Model based discrimination of gait before/after medication intake

For this classification problem, we considered the segments that were identified as gait by our model (i.e. states with sufficient power, see Section VI). The AR-iHMM segments the data into intervals with the state variables $z_t$ denoting the AR state representing the signal at time indexed $t$. If we then assume $K^+$ unique values for $z_t$ as $t = 1, \ldots, T$, we estimate $K^+$ sets of AR coefficients: $\{A_1^k, \ldots, A_r^k\}_{k=1}^{K^+}$. For each state $k$ we estimate the spectrum based on AR coefficients $A_1^k, \ldots, A_r^k$.

Common PSD features used to monitor PD related changes in the gait pattern include: position of the dominant peak (i.e. fundamental frequency); height of the dominant peak; width of the dominant peak; ratio of the first and second peak; energy in a specific frequency range, and others [54]. In our evaluation, we consider the height and position of the dominant peak, and the total energy in the range 0.5-10Hz (gait related information is expected in this frequency range). Because we expect that the relative rather than the absolute within-person changes are relevant to distinguish between before and after medication intake, we normalized all features per patient using z-scores.

Of the 25 PD patients taking part in the study, 18 had sufficient walking periods (at least 5 segments of 25 seconds) both before and after medication. For these patients, we trained a logistic classifier using each of the features described above, to predict whether their gait segments occurred before or after medication intake. For each patient, we computed the out-of-sample accuracy based on leave-one-subject-out cross validation[3].

[3]The accuracy of the leave-one-subject-out cross validation is affected both by the flexibility of the trained classifier, but also the intrinsic variation across features from different subjects.

### B. Prediction accuracy

As displayed in Table III, we could predict with reasonable accuracy whether a gait segment occurred before or after medication intake; note that not all patients had visible changes in their gait pattern after medication intake, so we expect that achieving perfect prediction accuracy will not be realistic. To examine how our approach for gait detection affects the ability to discriminate between before and after medication intake, we compared accuracy between using the gait segments identified by our model and using the gait segments as annotated on the video recordings. For the latter, we learned the AR-iHMM on all annotated gait data, and used all identified states to obtain the AR-based PSD. As shown in Table III, the accuracy to predict before/after medication intake using gait segments identified by the model was comparable to using annotated gait segments.

### C. Exploratory gait analysis

Because our probabilistic model is unsupervised, we can use the model not only as a tool to make predictions, but also as an exploratory tool to study the gait data. For example, in Figure 5 on the left we show the gait segmentation of a PD patient with a notable clinical improvement in gait pattern after medication intake (based on the video annotations), where different colors indicate different hidden states $z_t$. What we observe is that the probabilistic model not only allows us to identify non-gait segments (pink and green), but also discriminates between different variations in gait quality. In this patient, the model separates before medication gait (red) or after medication gait (yellow). By contrast, the right plots on Figure 5 show segmented gait of a PD patient whose gait does not notably improve after medication intake (based on the video recordings). Interestingly, we can still identify different gait segments both before and after medication intake, but their pattern of occurrence is similar in both conditions. Furthermore, inspection of the AR-based PSD estimates associated with the states in both figures indicates that the gait states on the right in Figure 5 are more similar to each other than the gait states associated with before and after medication periods on the left side in Figure 5. It should be noted that this contrast was not present in all patients, and we show two illustrative cases here. Further research is needed to identify all the reasons which can affect the reported change.

## VIII. DISCUSSION

In this report we study the problem of passively monitoring movement disorders such as Parkinson's disease (PD) in

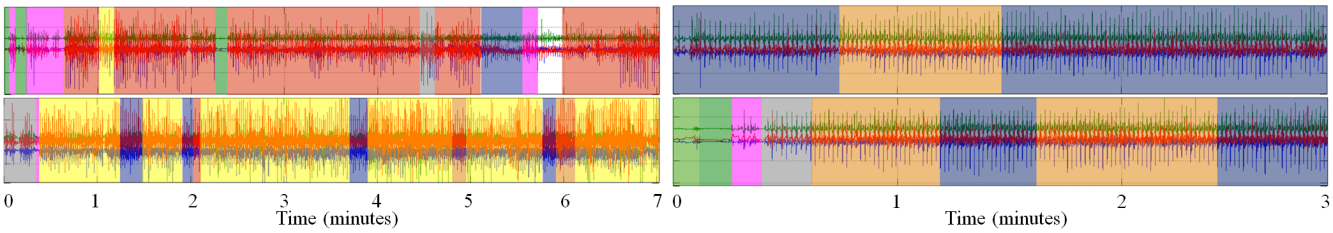


Fig. 5. Segmentation of smartphone data during the free-living assessment obtained from the AR-iHMM, from two PD patients with (left) and without (right) clinically observable changes and in the gait pattern after medication intake. Top: before medication intake. Bottom: after medication intake. On the left the red, yellow, blue and grey segments are all associated with gait data; on the right the yellow, blue and grey segments are all associated with gait data, with similar occurrence before and after medication intake; the remaining segments are not associated with gait.

daily living using wearable sensors. This is a challenging problem because of the large complexity and variation in daily living sensor signals, in combination with the scarcity of representative free-living datasets with reliable labels. This may explain why highly flexible methods such as deep learning have not been successful in the context of monitoring symptom fluctuations in PD [55]. This has stimulated the search for signal models that are based on principled assumptions which reduce the model's flexibility while still allowing it to capture subtle disease-related changes. In this work we propose a simple, structured probabilistic modelling approach for the analysis of free-living gait. Our approach is designed to simultaneously locate stationary gait segments and characterize the gait pattern, based on pre-processed accelerometer data. We achieve this by adopting a non-parametric switching autoregressive model, circumventing the need to use window-based analysis and the need to pre-define the number of gait and non-gait classes that can be observed in daily life.

We demonstrate our approach on a new reference dataset including 25 PD patients and 25 controls. The dataset is unique because it combines unscripted daily living activities in and around the house with detailed video annotations, which allows us to test our model on a much more realistic setting. First, we show that the identified classes can be used to accurately detect gait. Second, we show that states that represent gait can be used to predict medication induced fluctuations in PD patients.

### A. Benefits of AR-iHMM model based gait analysis

In addition to the evaluated benefits of the proposed AR-iHMM, there are some other properties which are not necessarily reflected in the reported comparisons. Our approach has two key advantages when it comes to estimating the spectrum of the free-living accelerometer data (or similar sensors): (1) the time boundaries of each segment of stationary data are adaptively selected by the model, which avoids the need for window-based analysis and problems associated with this; (2) in a fully probabilistic fashion, we can leverage multiple repeating patterns to get a more robust estimate of the spectrum. Additionally, our algorithm does not treat the problem as binary classification, but it is designed to learn multiple gait and non-gait states. This means that it can deal with changes in gait pattern during gait episodes. This avoids grouping different gait patterns together, which can introduce problems in further gait pattern analyses (see Figure 4). Moreover, because it is unsupervised, the model can be used as an exploratory tool to locate gait or non-gait segments that share the same (spectral) characteristics. The user can explicitly control the prior parameters of the model to determine the temporal granularity of the segmentation and focus on more or less detailed changes in the gait signals. This extra control allows us to focus on sufficiently stationary ("steady state") gait segments that are useful to make inferences about the gait pattern. Lastly, using this fully Bayesian model to describe the acceleration signals allows us to estimate the uncertainty involved with both the segmentation and the estimation of the spectrum.

### B. Limitations and future directions

In order to develop an intuitive, robust and easy to interpret probabilistic signal model for gait data in free-living, we have made some restrictive assumptions about the distribution and occurrence of such data in daily life. Despite the flexibility of inferring an unknown number of different spectral AR representations, we have focused on the states that have sufficient spectral power in the gait range to monitor changes after medication intake. We have shown that this approach is appropriate for monitoring the highly prevalent continuous impairments in PD patients (*bradykinetic gait*). However, by focusing on short-term interruptions of the gait, the model could potentially also be very suitable to monitor the more rare episodic hesitations (*freezing of gait*) [56], [57]. Because of the limited number of patients that presented with this symptom during the free-living assessments of the Parkinson@Home validation study, this remains to be evaluated using other data sets. In addition, the proposed framework does not use the axis meanings in the sensor outputs. This was done because in smartphones, the default orientation of the device can be different depending on the user. Our approach can also be applied to the three-dimensional dynamic component of the acceleration vector, or to any specific axis.

Due to overlap in typical gait frequencies and PD resting tremor, we can observe lower gait detection specificity mainly in the wrist-worn devices, compared to scenarios in which we have detected and separated PD tremor separately. Simultaneous PD tremor and gait detection in free living will be studied in future work.

Before the framework can used in medical contexts, further validation is necessary. For example, in the current study

TABLE III

BALANCED ACCURACY TO PREDICT WHETHER GAIT SEGMENTS OCCURRED BEFORE OR AFTER MEDICATION INTAKE, USING A LOGISTIC CLASSIFIER BASED ON PSD FEATURES OBTAINED FROM THE AR-IHMM, NORMALIZED PER SUBJECT. WE USE LEAVE-ONE-SUBJECT-OUT CROSS VALIDATION, AND PRESENT THE MEAN AND STANDARD ERROR ACROSS SUBJECTS. RESULTS ARE COMPARED BETWEEN USING GAIT SEGMENTS AS ANNOTATED ON THE VIDEO RECORDINGS ("ANNOTATED GAIT"), AND GAIT SEGMENTS IDENTIFIED BY OUR MODEL ("PREDICTED GAIT").

| Feature | Device locations | | | | | |
|---|---|---|---|---|---|---|
| | Smartphone | Left Wrist | Right Wrist | Left Shin | Right Shin | Lower Back |
| *Annotated gait* | | | | | | |
| Peak height | 68% (9%) | 66% (9%) | 68% (9%) | 70% (10%) | 72% (10%) | 70% (10%) |
| Peak position | 57% (8%) | 56% (6%) | 56% (6%) | 56% (8%) | 57% (9%) | 55% (8%) |
| Total energy | 72% (8%) | 67% (9%) | 69% (9%) | 69% (8%) | 74% (9%) | 72% (7%) |
| *Predicted gait* | | | | | | |
| Peak height | 67% (10%) | 63% (12%) | 66% (11%) | 64% (11%) | 66% (11%) | 68% (10%) |
| Peak position | 55% (12%) | 53% (14%) | 52% (15%) | 58% (12%) | 61 (12%) | 59% (11%) |
| Total energy | 75% (8%) | 65% (10%) | 68% (11%) | 70% (9%) | 74% (10%) | 76% (9%) |

protocol the gait before medication intake was measured after overnight withdrawal of dopaminergic medication. While this allowed for a detailed assessment of the changes after medication intake, in some patients the effects in daily life might be more subtle. Future work will aim to evaluate how well response fluctuations can be captured for naturally occurring shorter withdrawal periods in truly unsupervised conditions.

Finally, we emphasize that the developed framework aims to merely segment varying gait patterns in a principled and largely unsupervised manner. In order to assign meaning to detected changes in the gait pattern in fully unsupervised conditions, e.g. estimate causal effects of medication, further analysis using a carefully designed causal map is required. For example, real-life factors such as environment (e.g. crowded city versus park , indoors electromagnetic fields [58]) and voluntary behavior (e.g. making gestures while walking, specifically for the trouser pocket: type of clothing) might also influence the participants' gait and our ability to measure it. While the current data set includes much more environmental and behavioral variation than lab-based studies, the activities performed before and after medication intake were similar, which may not be the case in fully free-living conditions. Therefore, future work will include adjusting for potential real-life confounding, using additional contextual sensors such as GPS.

## REFERENCES


[1] F. Lipsmeier, I. Fernandez Garcia, D. Wolf, T. Kilchenmann, A. Scotland, J. Schjodt-Eriksen, W.-Y. Cheng, J. Siebourg-Polster, L. Jin, J. Soto, L. Verselis, M. Martin Facklam, F. Boess, M. Koller, M. Grundman, M. A. Little, A. Monsch, R. Postuma, A. Gosh, T. Kremer, K. Taylor, C. Czech, C. Gossens, and M. Lindemann, "Successful passive monitoring of early-stage Parkinson's disease patient mobility in Phase I RG7935/PRX002 clinical trial with smartphone sensors," *Movement Disorders*, vol. 32, no. 2, 2017.

[2] E. Warmerdam, J. M. Hausdorff, A. Atrsaei, Y. Zhou, A. Mirelman, K. Aminian, A. J. Espay, C. Hansen, L. J. Evers, A. Keller *et al.*, "Long-term unsupervised mobility assessment in movement disorders," *The Lancet Neurology*, 2020.

[3] A. L. S. de Lima, T. Hahn, N. M. de Vries, E. Cohen, L. Bataille, M. A. Little, H. Baldus, B. R. Bloem, and M. J. Faber, "Large-scale wearable sensor deployment in Parkinson's patients: the Parkinson@Home study protocol," *JMIR research protocols*, vol. 5, no. 3, 2016.

[4] S. Arora, V. Venkataraman, S. Donohue, K. M. Biglan, E. R. Dorsey, and M. A. Little, "High accuracy discrimination of Parkinson's disease participants from healthy controls using smartphones," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 3641–3644.

[5] B. M. Bot, C. Suver, E. C. Neto, M. Kellen, A. Klein, C. Bare, M. Doerr, A. Pratap, J. Wilbanks, E. R. Dorsey *et al.*, "The mpower study, Parkinson disease mobile data collected using researchkit," *Scientific data*, vol. 3, p. 160011, 2016.

[6] S. Del Din, M. Elshehabi, B. Galna, M. A. Hobert, E. Warmerdam, U. Suenkel, K. Brockmann, F. Metzger, C. Hansen, D. Berg *et al.*, "Gait analysis with wearables predicts conversion to Parkinson's disease," *Annals of neurology*, vol. 86, no. 3, pp. 357–367, 2019.

[7] C. Curtze, J. G. Nutt, P. Carlson-Kuhta, M. Mancini, and F. B. Horak, "Levodopa is a double-edged sword for balance and gait in people with Parkinson's disease," *Movement Disorders*, vol. 30, no. 10, pp. 1361–1370, 2015.

[8] D. Hodgins, "The importance of measuring human gait." *Medical Device Technology*, vol. 19, no. 5, pp. 42–44, 2008.

[9] E. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "Nonparametric bayesian learning of switching linear dynamical systems," in *Advances in Neural Information Processing Systems*, 2009, pp. 457–464.

[10] W. Tao, T. Liu, R. Zheng, and H. Feng, "Gait analysis using wearable sensors," *Sensors*, vol. 12, no. 2, pp. 2255–2283, 2012.

[11] L. Brognara, P. Palumbo, B. Grimm, and L. Palmerini, "Assessing gait in Parkinson's disease using wearable motion sensors: a systematic review," *Diseases*, vol. 7, no. 1, p. 18, 2019.

[12] A. Brajdic and R. Harle, "Walk detection and step counting on unconstrained smartphones," in *Proceedings of the 2013 ACM International Joint conference on Pervasive and Ubiquitous computing*. ACM, 2013, pp. 225–234.

[13] C. Randell, C. Djiallis, and H. Muller, "Personal position measurement using dead reckoning," in *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.* IEEE, 2003, pp. 166–173.

[14] E. K. Antonsson and R. W. Mann, "The frequency content of gait," *Journal of Biomechanics*, vol. 18, no. 1, pp. 39–47, 1985.

[15] A. Samà, C. Pérez-López, D. Rodríguez-Martín, A. Català, J. M. Moreno-Aróstegui, J. Cabestany, E. de Mingo, and A. Rodríguez-Molinero, "Estimating bradykinesia severity in Parkinson's disease by analysing gait through a waist-worn sensor," *Computers in biology and medicine*, vol. 84, pp. 114–123, 2017.

[16] D. M. Karantonis, M. R. Narayanan, M. Mathie, N. H. Lovell, and B. G. Celler, "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring," *IEEE Transactions on Information technology in Biomedicine*, vol. 10, no. 1, pp. 156–167, 2006.

[17] A. El-Attar, A. S. Ashour, N. Dey, H. Abdelkader, M. M. Abd El-Naby, and R. S. Sherratt, "Discrete wavelet transform-based freezing of gait detection in Parkinson's disease," *Journal of Experimental & Theoretical Artificial Intelligence*, pp. 1–17, 2018.

[18] M. A. Little, *Machine Learning for Signal Processing: Data Science, Algorithms, and Computational Statistics*. Oxford University Press, USA, 2019.

[19] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 293–304.

[20] Y. Makihara, N. T. Trung, H. Nagahara, R. Sagawa, Y. Mukaigawa,

and Y. Yagi, "Phase registration of a single quasi-periodic signal using self dynamic time warping," in *Asian Conference on Computer Vision*. Springer, 2010, pp. 667–678.

[21] H. Ying, C. Silex, A. Schnitzer, S. Leonhardt, and M. Schiek, "Automatic step detection in the accelerometer signal," in *4th International Workshop on Wearable and Implantable Body Sensor Networks (BSN 2007)*. Springer, 2007, pp. 80–85.

[22] L. Rong, D. Zhiguo, Z. Jianzhong, and L. Ming, "Identification of individual walking patterns using gait acceleration," in *2007 1st international Conference on Bioinformatics and Biomedical Engineering*. IEEE, 2007, pp. 543–546.

[23] M. Nyan, F. Tay, K. Seah, and Y. Sitoh, "Classification of gait patterns in the time–frequency domain," *Journal of Biomechanics*, vol. 39, no. 14, pp. 2647–2656, 2006.

[24] M. Mathie, B. G. Celler, N. H. Lovell, and A. Coster, "Classification of basic daily movements using a triaxial accelerometer," *Medical and Biological Engineering and Computing*, vol. 42, no. 5, pp. 679–687, 2004.

[25] N. Haji Ghassemi, J. Hannink, C. Martindale, H. Gaßner, M. Müller, J. Klucken, and B. Eskofier, "Segmentation of gait sequences in sensor-based movement analysis: a comparison of methods in Parkinson's disease," *Sensors*, vol. 18, no. 1, p. 145, 2018.

[26] J. Camps, A. Sama, M. Martin, D. Rodriguez-Martin, C. Perez-Lopez, J. M. M. Arostegui, J. Cabestany, A. Catala, S. Alcaine, B. Mestre *et al.*, "Deep learning for freezing of gait detection in Parkinson's disease patients in their homes using a waist-worn inertial measurement unit," *Knowledge-Based Systems*, vol. 139, pp. 119–131, 2018.

[27] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," *arXiv preprint arXiv:1604.08880*, 2016.

[28] J. McCamley, M. Donati, E. Grimpampi, and C. Mazza, "An enhanced estimate of initial contact and final contact instants of time using lower trunk inertial sensor data," *Gait & posture*, vol. 36, no. 2, pp. 316–318, 2012.

[29] W. Zijlstra and A. L. Hof, "Assessment of spatio-temporal gait parameters from trunk accelerations during human walking," *Gait & posture*, vol. 18, no. 2, pp. 1–10, 2003.

[30] S. Del Din, A. Godfrey, and L. Rochester, "Validation of an accelerometer to quantify a comprehensive battery of gait characteristics in healthy older adults and Parkinson's disease: toward clinical and at home use," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 3, pp. 838–847, 2015.

[31] S. T. Moore, V. Dilda, B. Hakim, and H. G. MacDougall, "Validation of 24-hour ambulatory gait assessment in Parkinson's disease with simultaneous video observation," *Biomedical engineering online*, vol. 10, no. 1, p. 82, 2011.

[32] A. Weiss, T. Herman, N. Giladi, and J. M. Hausdorff, "Objective assessment of fall risk in Parkinson's disease using a body-fixed sensor worn for 3 days," *PloS one*, vol. 9, no. 5, 2014.

[33] S. M. Rispens, K. S. van Schooten, M. Pijnappels, A. Daffertshofer, P. J. Beek, and J. H. van Dieen, "Identification of fall risk predictors in daily life measurements: gait characteristics' reliability and association with self-reported fall history," *Neurorehabilitation and neural repair*, vol. 29, no. 1, pp. 54–61, 2015.

[34] C. Pérez-López, A. Samà, D. Rodríguez-Martín, A. Català, J. Cabestany, J. M. Moreno-Arostegui, E. De Mingo, and A. Rodríguez-Molinero, "Assessing motor fluctuations in Parkinson's disease patients based on a single inertial sensor," *Sensors*, vol. 16, no. 12, p. 2132, 2016.

[35] J. Bellanca, K. Lowry, J. VanSwearingen, J. Brach, and M. Redfern, "Harmonic ratios: a quantification of step to step symmetry," *Journal of Biomechanics*, vol. 46, no. 4, pp. 828–831, 2013.

[36] R. Moe-Nilssen and J. L. Helbostad, "Estimation of gait cycle characteristics by trunk accelerometry," *Journal of Biomechanics*, vol. 37, no. 1, pp. 121–126, 2004.

[37] L. J. Evers, Y. P. Raykov, J. H. Krijthe, A. Silva de Lima, R. Badawy, K. Claes, T. M. Heskes, M. A. Little, M. J. Meinders, and B. R. Bloem, "Real-life gait performance as a digital biomarker for motor fluctuations: the Parkinson@Home validation study." *Journal of Medical Internet Research 21/08/2020:19068 (forthcoming/in press)*, 2020.

[38] C. G. Goetz, B. C. Tilley, S. R. Shaftman, G. T. Stebbins, S. Fahn, P. Martinez-Martin, W. Poewe, C. Sampaio, M. B. Stern, R. Dodel *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.

[39] M. R. Munetz and S. Benjamin, "How to examine patients using the abnormal involuntary movement scale," *Psychiatric Services*, vol. 39, no. 11, pp. 1172–1177, 1988.

[40] A. Zhan, M. A. Little, D. A. Harris, S. O. Abiola, E. R. Dorsey, S. Saria, and A. Terzis, "High frequency remote monitoring of Parkinson's disease via smartphone: Platform overview and medication response detection," *arXiv preprint arXiv:1601.00960*, vol. abs/1601.00960, 2016.

[41] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, "Estimation of imu and marg orientation using a gradient descent algorithm," in *2011 IEEE International conference on Rehabilitation Robotics*. IEEE, 2011, pp. 1–7.

[42] V. T. Van Hees, L. Gorzelniak, E. C. D. Leon, M. Eder, M. Pias, S. Taherian, U. Ekelund, F. Renström, P. W. Franks, A. Horsch *et al.*, "Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity," *PloS one*, vol. 8, no. 4, p. e61691, 2013.

[43] E. Bergamini, G. Ligorio, A. Summa, G. Vannozzi, A. Cappozzo, and A. M. Sabatini, "Estimating orientation using magnetic and inertial sensors and different sensor fusion approaches: Accuracy assessment in manual and locomotion tasks," *Sensors*, vol. 14, no. 10, pp. 18 625–18 649, 2014.

[44] R. Badawy, Y. P. Raykov, L. J. Evers, B. R. Bloem, M. J. Faber, A. Zhan, K. Claes, and M. A. Little, "Automated quality control for sensor based symptom measurement performed outside the lab," *Sensors*, vol. 18, no. 4, p. 1215, 2018.

[45] S.-J. Kim, K. Koh, S. Boyd, and D. Gorinevsky, "$l\_1$ trend filtering," *SIAM review*, vol. 51, no. 2, pp. 339–360, 2009.

[46] P. Welch, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, 1967.

[47] D. J. MacKay, "Probable networks and plausible predictions—a review of practical bayesian methods for supervised neural networks," *Network: computation in neural systems*, vol. 6, no. 3, pp. 469–505, 1995.

[48] E. B. Fox, "Bayesian nonparametric learning of complex dynamical phenomena," Ph.D. dissertation, Massachusetts Institute of Technology, 2009.

[49] M. J. Beal, Z. Ghahramani, and C. E. Rasmussen, "The infinite hidden Markov model," in *Advances in Neural Information Processing Systems 14*, 2002, pp. 577–584.

[50] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Sharing clusters among related groups: Hierarchical Dirichlet processes," in *Advances in Neural Information Processing Systems 17*, 2005, pp. 1385–1392.

[51] Y. P. Raykov, E. Ozer, G. Dasika, A. Boukouvalas, and M. A. Little, "Predicting room occupancy with a single passive infrared (PIR) sensor through behavior extraction," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 1016–1027.

[52] P. Barralon, N. Vuillerme, and N. Noury, "Walk detection with a kinematic sensor: Frequency and wavelet comparison," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2006, pp. 1711–1714.

[53] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning requires rethinking generalization," *arXiv preprint arXiv:1611.03530*, 2016.

[54] S. Del Din, A. Godfrey, B. Galna, S. Lord, and L. Rochester, "Free-living gait characteristics in ageing and Parkinson's disease: impact of environment and ambulatory bout length," *Journal of Neuroengineering and Rehabilitation*, vol. 13, no. 1, p. 46, 2016.

[55] N. Y. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker, and T. Plötz, "PD disease state assessment in naturalistic environments using deep learning," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015, pp. 1742–1748.

[56] S. T. Moore, H. G. MacDougall, and W. G. Ondo, "Ambulatory monitoring of freezing of gait in Parkinson's disease," *Journal of neuroscience methods*, vol. 167, no. 2, pp. 340–348, 2008.

[57] C. Barthel, E. Mallia, B. Debû, B. R. Bloem, and M. U. Ferraye, "The practicalities of assessing freezing of gait," *Journal of Parkinson's disease*, vol. 6, no. 4, pp. 667–674, 2016.

[58] J. A. Hughes, J. A. Brown, A. M. Khan, A. M. Khattak, and M. Daley, "Analysis of symbolic models of biometrie data and their use for action and user identification," in *2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, 2018, pp. 1–8.