Non-Linear Heart Rate Variability Measures in Sleep Stage Analysis with Photoplethysmography

Matti Molkkari[1], Mirja Tenhunen[2], Adrian Tarniceriu[3], Antti Vehkaoja[1,4], Sari-Leena Himanen[1,2], Esa Räsänen[1]

[1] Tampere University, Tampere, Finland
[2] Tampere University Hospital, Tampere, Finland
[3] PulseOn SA, Neuchatel, Switzerland
[4] PulseOn Oy, Espoo, Finland

Abstract

We assess the feasibility of heart rate variability (HRV) estimated from interbeat interval (IBI) data measured with wrist-worn photoplethysmography device for sleep stage classification. In particular, we examine fractal correlations in the IBIs as the function of both time and scale.

Optical heart rate sensor by PulseOn Ltd was utilized for monitoring IBIs from 18 healthy young adult subjects. Reference ambulatory polysomnography recordings were scored by a sleep physician. The HRV was studied by detrended fluctuation analysis by computing scale-dependent spectra of scaling exponents $\alpha(s)$. Dynamic changes were tracked by calculating the spectra $\alpha(s,t)$ in moving temporal windows whose length varied with the scale.

The dynamic landscapes of the alpha spectra show distinctive fractal correlations according to the underlying sleep stages. Respiratory effects, blood pressure variations, and thermoregulatory influence appear to be discernible as well. Classification of the alpha spectra yields up to $73\,\%$, $60\,\%$ and $54\,\%$ average accuracies for 3-class (wake, REM, NREM), 4-class (wake, REM, N1+2, N3) and 5-class (wake, REM, N1, N2, N3) cases, respectively.

1. Introduction

General interest for intelligent health and wellness products has become increasingly prevalent in the recent times. Sleep quality has a profound effect on everyday well-being. Sleep consists of distinct recurrent phases serving diverse physiological purposes. Knowledge of this sleep structure is essential for assessing sleep quality or diagnosing sleep disorders. In clinical practice the sleep stages are determined by polysomnography (PSG) predominantly from the brain activity, requiring evaluation by experts. Wearable sensor technology has permitted cheap and unobtrusive monitoring of physiological signals that previously required expensive examinations in clinical conditions. However, the precision of sleep stage monitoring by these devices is not yet medically sufficient.

We assess the feasibility of heart rate variability (HRV) measures for sleep stage classification utilizing interbeat intervals (IBIs) obtained from wrist-worn photoplethysmography (PPG). We study HRV by detrended fluctuation analysis (DFA), which is a popular tool for determining fractal correlations in non-stationary time series. DFA assesses power law scaling of the fluctuations $F(s) \propto s^{\alpha}$ from a local trend as a function of the scale s. The time series is interpreted as the increments of a random walk, and the detrending permits meaningful analysis in the presence of non-stationarities [1]. We propose a novel expansion to the method in order to accurately study dynamic changes in this scaling as the function of both scale and time.

2. Theory

In the standard formulation of DFA the fluctuation function $F(s)$ is computed as the mean fluctuations over non-overlapping windows of length s [1, 2], but its statistical properties may be improved by utilizing overlapping windows [3]. As power laws are transformed into linear relationships on logarithmic scale, the scaling exponents are conventionally determined by linear regression on the logarithmic fluctuation function [1, 2]. However, in practice many phenomena deviate from exact power law scaling, which is often observed only over limited ranges of scales. This could be taken into account by computing a whole spectrum of scaling exponents $\alpha(s)$ as a function of the scale [4–6]. The scaling may also exhibit temporal variations due to changes in external conditions, or the process itself may be comprised of distinct intrinsic modes. Straightforward segmentation of the time series suffers from the following limitations: The largest scale is dictated by the segment length to ascertain sufficient sta-

 ISSN: 2325-887X DOI: 10.22489/CinC.2019.287

tistical accuracy, and the temporal resolution of the shorter scales is compromised by the longer segments required for the larger scales. Therefore we adopt the following procedure to dynamically adjust the segment length as a function of the scale.

1. For each scale s, divide the time series into scale-dependent segments of length $l(s)$ by, e.g., $l(s) = 5s$.
2. In each segment $\mathcal{S}_{s,t}$, identified by the temporal indices t, compute the fluctuation function at the scales $\{s-1, s, s+1\}$. These computations should be carried out by utilizing maximally overlapping DFA windows.
3. Estimate the local alpha spectra $\alpha(s,t)$ by, e.g., finite difference scheme from the logarithmic fluctuation function, as the smoothness of the overlapping DFA procedure permits direct numerical differentiation.

3. Methods

Optical wrist heart rate monitor manufactured by PulseOn Ltd was used in the collection of beat-to-beat interval data. The device uses proprietary algorithms for heartbeat detection and signal quality assessment so that it provides only reliable beat-to-beat intervals to the user [7]. Data was collected from 18 students (average age 28 years, range 21–42 years) with no history of diagnosed sleep disorders. Nox A1 ambulatory PSG system from Nox Medical was used to record the reference data. The measurement devices were worn on the subject in the afternoon preceding the measurement night and the recording was done at subject's home. Subjects turned on the recording in the PSG and the wrist device before going to bed. Sleep physician visually confirmed and corrected the results of the initial automatic sleep staging performed by Noxturnal sleep study software in standard $30\,\mathrm{s}$ epochs. An ethical assessment for the study was provided by the ethical committee of Tampere University Hospital (R17171).

Based on the PPG sensor signal quality assessment the data from one subject had to be discarded. In addition to this intrinsic filtering, possible additional artifacts, e.g. missed beats or non-sinus originated extra beats, were removed if the IBIs differed more than $50\,\%$ from the local median within 51 beat window.

We utilize temporal slices of the alpha spectra $\alpha(s,t)$ as features for classification of the sleep phases. First, we limit our analysis to scales < 500 beats and aggregate the features by computing the mean spectra and its standard deviation from all the slices within each epoch. Then we normalize the spectra to the unit interval based on the quantiles of observed scaling exponents and their standard deviations for each scale and subject separately. These HRV features are supplemented by similarly aggregated heart rate data within the dynamic DFA segments.

We apply principal component analysis (PCA) to reduce the dimensionality of the feature vectors. The classification is then performed by support vector machines (SVMs) with radial basis functions. We cross-validate our results by leave-one-subject-out strategy with balanced accuracy as our performance metric, which is defined as accuracy with each sample weighted proportional to the inverses of class prevalences. Hyperparameters (PCA components and SVM penalty parameter) are optimized by a simple grid search.

The raw predictions from the classifier are regularized by Bayesian filtering to reduce noise in the hypnograms by taking into account common sleep structure. We perform bootstrapped sequential importance resampling (SIR) with the importance distributions based on epoch-by-epoch sleep state transition probabilities. The particles are weighted based on the observation probabilities derived from the confusion matrices of the classifier predictions. The leave-one-subject-out scheme is enforced when computing these probabilities.

4. Results

The IBIs are found to exhibit rich patterns in their fractal correlations during different sleep phases. An example of these correlations as a function of the scale and time is shown in Fig. 1. While a general trend is noticeable during different sleep phases, particularly at the shortest scales, considerable variance exists within the same sleep phase. Regardless, distinct fractal correlations are found during different sleep phases when averaging over the subjects and temporal slices of the alpha spectra, which is illustrated in Fig. 2(a). The differences in the mean correlations are statistically significant over extended ranges of scales, except for separating the N1 and REM states. However, there is considerable inter- and intra-subject variance, which complicates straightforward analysis. The inter-subject variance is reduced by the quantile-based normalization, as can be seen in Fig. 2(b). The intra-epoch variance exhibits dependency on the sleep state, as shown in Fig. 2(c), which is utilized for the classification.

The prevailing trend is that the deeper the sleep phase, the less correlated the IBIs become. The qualitative behavior as a function of the scale is similar for each sleep phase. Elevated correlations at the very shortest scales are well-known artifacts intrinsic to DFA. The first minimum in the alpha spectra is likely related to cyclic respiratory effects, resulting in superimposed anticorrelations lowering the scaling exponent. Increased correlations at the scale of roughly 10 beats is observed across all the sleep phases, and this may be related to blood pressure variations [8]. The increase is larger in deeper stages of sleep, which is consistent with previously studied short-term variabilities of blood pressure and IBIs [9].

The correlations decrease again towards the scales in the region of 20 beats, which could be associated with temper-

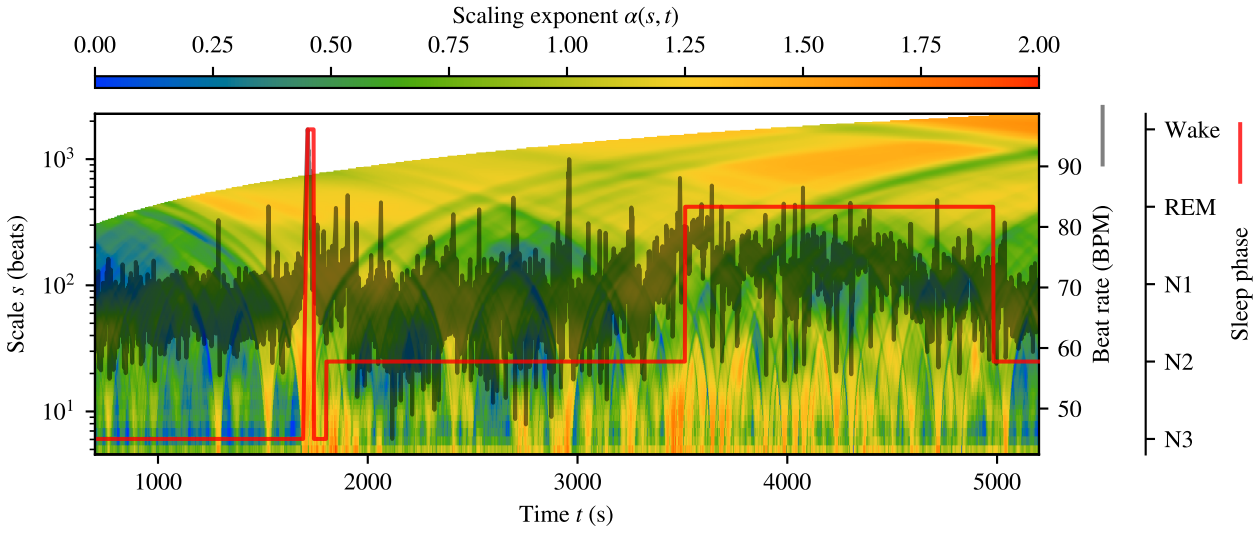

Figure 1. Dynamic DFA example. The scaling exponents and the beat rate are derived from the PPG measurements and the sleep phases are from the reference PSG annotations.

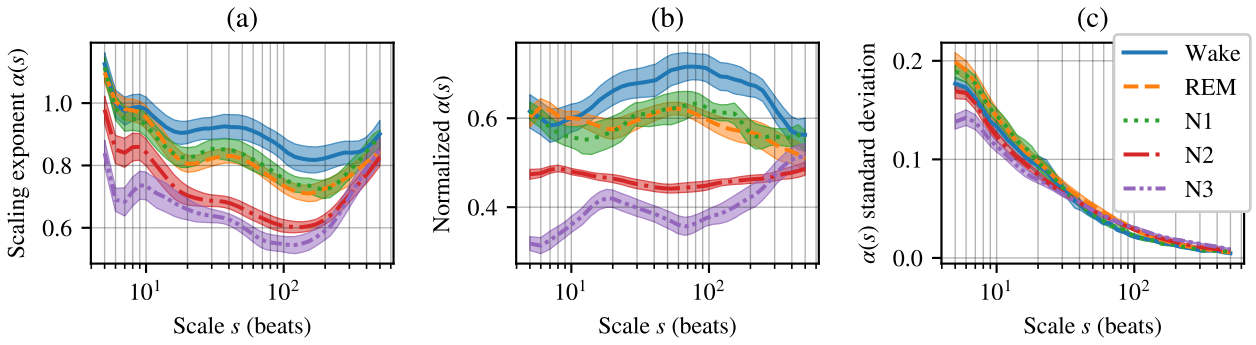

Figure 2. Mean alpha spectra (a) with their standard errors during different sleep phases when averaged over the subjects and time. In (b) the data has been normalized into the unit interval based on quantiles of observed scaling exponents separately for each scale and subject. The standard deviation of the spectra within the epochs is shown in (c).

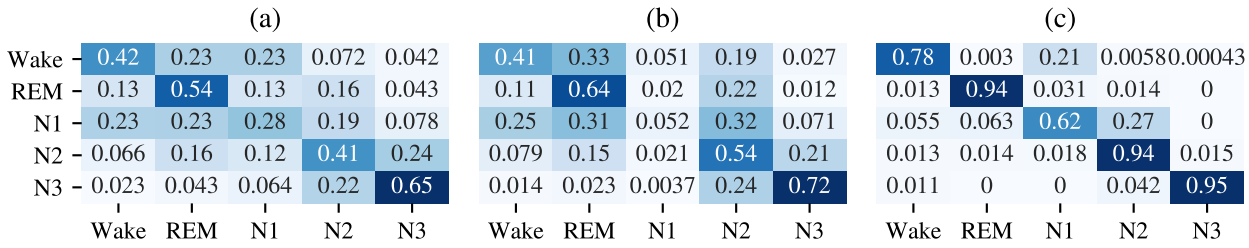

Figure 3. Confusion (a,b) and state transition matrices (c). The results in (a) are the raw predictions by the SVM classifier, and the results of the corresponding filtered predictions are shown in (b). Balanced accuracy as defined here is equivalent to the mean of the confusion matrix diagonals. Epoch-to-epoch sleep state transition probabilities are illustrated in (c). The matrices are computed by combining the data for all the included subjects.

ature regulation [10]. The authors propose no physiological explanation for the subsequent longer scale fluctuations in the spectra, but point out that they are qualitatively similar during all the sleep stages with the magnitude of the fluctuations decreasing in deeper sleep.

We note that after approximately 200 beats the correlations start to converge towards common long-range behavior. This is expected, as the sleep phases are determined from the reference annotations for the epochs indicated by the mean time within the dynamic DFA segments, resulting into overlap with different sleep phases at the longest scales. We justify this from the viewpoint of predicting the sleep phases from the correlations, in which case the data cannot be separated *a priori*. This also provides an explanation for the indistinctiveness of the N1 state, as these phases are very short, resulting in their alpha spectra being largely influenced by the surrounding Wake and N2 states.

Even with relatively large variance in the correlations, limited classification of the sleep phases is possible based on the fractal correlations alone. Leave-one-subject-out cross-validation strategy yields $43\,\%$ class-balanced accuracy for the 5-class case (Wake, REM, N1, N2, N3). Incorporating heart rate into the feature set and introducing the filtering of the raw predictions increases this accuracy to $47\,\%$. Different metrics for the classification results are listed in Table 1. The results for 4-class (Wake, REM, N1+2, N3) and 3-class (Wake, REM, NREM) cases are comparable to Fonseca et al., except our results are based on the fractal correlations instead of ensembles of HRV metrics [11]. The filtering is observed to improve the metrics, but as is evident from the full confusion matrices of Fig. 3, the N1 state is essentially precluded. The overall state transition probabilities are shown in Fig. 3(c), but for each subject their data is excluded prior to the filtering.

Table 1. Classification results for different number (#) of classes. For each score the left column contains the raw results from the classifier and the right column depicts the results after filtering the predictions.

#	Accuracy		Bal. accuracy		Cohen's κ	
5	0.465	0.542	0.460	0.471	0.288	0.352
4	0.535	0.601	0.538	0.570	0.313	0.384
3	0.668	0.725	0.618	0.622	0.373	0.434

5. Conclusion

Wrist-wearable PPG sensors are adequate for advanced HRV analysis. Respiratory effects, blood pressure variations, and thermoregulatory influence appear to be visible in the alpha spectra. Other longer scale fluctuations are also present in the spectra, but their interpretation requires further research. The dynamic DFA reveals distinct fractal correlations during different sleep phases. The method allows increased temporal resolution at the shortest scales while simultaneously incorporating larger scale behavior into the analysis. The temporal resolution is gained at the expense of increased variance. However, the variance itself also contains information, as its magnitude varies according to the sleep phases. Limited sleep phase classification is possible from the fractal correlations alone, but supplementing other information from wearable sensors with the dynamic DFA features is expected to enhance the accuracy of low-cost ambulatory sleep stage monitoring. The filtering of the raw classifier predictions is also a promising direction for further research.

References

[1] Peng CK, et al. Mosaic organization of DNA nucleotides. Physical Review E 1994;49(2):1685–1689.

[2] Kantelhardt JW, et al. Detecting long-range correlations with detrended fluctuation analysis. Physica A 2001; 295(3):441–454.

[3] Kiyono K, Tsujimoto Y. Nonlinear filtering properties of detrended fluctuation analysis. Physica A 2016;462:807–815.

[4] Viswanathan GM, et al. Deviations from uniform power law scaling in nonstationary time series. Physical Review E 1997;55:845–849.

[5] Echeverría JC, et al. Interpretation of heart rate variability via detrended fluctuation analysis and $\alpha\beta$ filter. Chaos 2003;13(2):467–475.

[6] Molkkari M, Räsänen E. Robust estimation of the scaling exponent in detrended fluctuation analysis of beat rate variability. In Computing in Cardiology. 2018; 219.

[7] Tarniceriu A, et al. Detection of beat-to-beat intervals from wrist photoplethysmography in patients with sinus rhythm and atrial fibrillation after surgery. In IEEE International Conference on Biomedical and Health Informatics. 2018; 133–136.

[8] Julien C. The enigma of Mayer waves: Facts and models. Cardiovascular Research 2006;70(1):12–21.

[9] Van de Borne P, et al. Effects of wake and sleep stages on the 24-h autonomic control of blood pressure and heart rate in recumbent men. American Journal of Physiology Heart and Circulatory Physiology 1994;266(2):H548–H554.

[10] Fleisher LA, et al. Thermoregulation and heart rate variability. Clinical Science 1996;90(2):97–103.

[11] Fonseca P, et al. Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults. Sleep 2017;40(7):zsx097.

Address for correspondence:

Matti Molkkari
Tampere University, computational physics laboratory
Korkeakoulunkatu 7, 33720 Tampere, Finland
matti.molkkari@tuni.fi