Sleep Insights from the Finger Tip: How Photoplethysmography Can Help Quantify Sleep

Shuli Eyal[1], Anda Baharav[1,2]

[1]HypnoCore LTD, Petach-Tikva, Israel
[2]WinMed, Wingate Institute, Netanya, Israel

Abstract

Sleep is essential for a healthy and productive life, yet its importance is largely overlooked, allowing populations to sleep less and to develop sleep disturbances. This trend results into an epidemic of poor quality and insufficient sleep that is turn jeopardizes health, performance, mood, memory, social relationships and productivity. A first step to overcome this epidemic relies on uncovering it at the individual and societal levels. The availability of different wearable devices that can track physiological signals represents a great opportunity to define and quantify the problem. Many such devices incorporate a photoplethysmography (PPG) sensor. This triggers interest in studies aiming to get insight into human sleep structure based on information obtained from PPG sensors.

This study aimed to validate a new automated sleep analysis which is simply based on the inter-beat-interval series obtained from PPG, and that uses features of heart rate variability. The candidate algorithm was tested against gold standard scoring of whole night state-of-the-art sleep studies.

The PPG-based sleep scoring performs very well in differentiating sleep stages, however, the sleep/wake separation is not sufficient and requires improvement. This last task is facilitated by the fact that the majority of devices with PPG capabilities, are equipped with accelerometers providing additional information for better separation. Combining accelerometers and PPG signals from wearable devices in a sleep analyser is likely to provide a reliable and accurate automated detection of sleep and wakefulness, including sleep macro- and micro-architecture.

1. Introduction

Human life has changed very rapidly during the last century. People move less, eat more, sleep less. Technology made this possible, yet the price is high as we live in a sleepless and heavy-bodied society. The solution to this crisis may be found in technology itself. Smartphones became an indispensable companion for the vast majority of people, wearable devices tend to follow this trend. Now many use their smart communication companion and their wearable device to monitor their level of physical activity, their nutrition balance, or their sleep. Knowing some metrics regarding those vital human functions may help improve the human sleep-life balance, thus improving health, mood, performance, memory and longevity. Same wearables that track activity can help monitoring sleep, based on accelerometer and/or instantaneous heart rate signals. Poor quality sleep and insufficient sleep jeopardize health, performance and wellbeing [1]. Medically provided sleep solutions are limited, and expensive, leaving most sufferers undetected, and thus untreated efficiently. Sleep self-assessment, followed by digital personalized expert guidance, may represent a suitable solution for individuals as well as for the health establishment.

Most fitness trackers and smart watches offer sleep analysis derived from their accelerometer sensors, offering information regarding sleep and wakefulness periods only [2,3], yet they lack the ability to discriminate between various sleep stages (Light Sleep (LS), Deep Sleep (DS), or Rapid Eyes Movement Sleep (REM)) sleep fragmentation, stress levels and Heart Rate (HR) during sleep. Many such devices incorporate a sensor that uses Photoplethysmography (PPG) to measure HR. The signals originating in this sensor allow detection of the instantaneous Inter Beat Interval (IBI), which enables the evaluation of the autonomic nervous fluctuations [4].

Our ECG-based validated sleep diagnostic software [5] based on HR variability (HRV) analysis, has been modified to rely solely on IBI series obtained from ECG signal (IBI_{ECG}). The wide availability of PPG based devices for HR detection during physical activity triggered our interest to evaluate whether the IBI obtained from PPG signals (IBI_{PPG}) during sleep, allows to evaluate sleep structure.

Several studies [4,6] indicate that HRV based on IBI_{PPG} may be used as an alternative to the HRV calculated from electrical signals of the heart, IBI_{ECG}. The two time series

 ISSN: 2325-887X DOI:10.22489/CinC.2017.274-197

(IBI_{ECG} and IBI_{PPG}) are not identical, as the pulse wave travels from the heart to the wrist/finger, and the time required for this, known as the pulse transient time (PTT), has its inherent beat to beat variability.

The goal of this study is to evaluate the applicability, accuracy, and reliability of our sleep analysis algorithm when adjusted to rely on IBI_{PPG}.

2. Methods

Data from 88 whole night gold standard sleep studies (PSG) that included both ECG (standard lead II), and finger PPG (Nonin transmissive PPG) were used to check the performance of a PPG based algorithm. 35 sleep studies (20% male, 51.6±8.3 years, BMI 30.3±5.1) were used as a training set, and the remaining 53 studies (17% male, 51.8±8.9 years, BMI 29.0±5.6) served as the test set.

All the sleep studies used represent a subset of a case-control study characterizing polysomnographic traits of chronic fatigue syndrome. The study was conducted at the Center for Disease Control in Wichita, Kansas, and is described elsewhere [7]. The PSG tests were performed using a N7000 system by Embla. ECG and PPG signals were recorded as part of the standard PSG protocol. The transmissive pulse oximeter probe was applied to either the right or left index finger. The PPG signal was sampled at 75Hz and 8bit resolution, and the ECG signal at 200Hz and 16bit.

Each PSG recording was scored manually according to the American Academy of Medicine Scoring criteria [8], by a single registered technologist, in 30sec epochs. Each epoch was scored as either wake, stage 1 or 2, slow-wave (SWS also referred as DS), or REM sleep.

The automated PPG analysis included: (1) A preprocessing procedure applied to the PPG signal that consisted of up-sampling the PPG signal to 200Hz, followed by low-pass filtering the result with a cutoff frequency of 10Hz. (2) The extraction of the IBI series from the PPG signal. Portions of the PPG signals were tainted by clipping, which precluded defining the beats location based on detecting of the apex points. Therefore, we chose to define the IBI as the time-interval between two consecutive inflection points in the upslope part of the filtered PPG signal. (3) An automated correction procedure in which outlier IBI points were located based on their surrounding values and were removed.

The automated sleep analysis applied to the PPG signal was a version of a validated ECG-based sleep diagnostic software [5]. The algorithm used a similar approach, and a similar set of HRV features, yet this time for IBI_{PPG} instead of IBI_{ECG}. The time-domain features included sample statistics of IBI duration and differences, nonlinear features such as characteristics features of Poincaré plot, and detrended fluctuation analysis [9]. The frequency domain features were based on time frequency decomposition, which allowed optimizing the resolution in time and frequency providing spectral powers in the very-low (0.008-0.04Hz), low (0.04-0.15Hz), and high (0.15-0.5Hz) frequency bands [10]. Finally, the stage definition was based on a Bayesian classifier. The outcome of the classifier was then combined with a correction procedure based on arousals and awakenings detection, obtained separately from the analysis of consecutive beat series.

The algorithm uncovered sleep architecture in 30 seconds epochs, including wakefulness, LS (corresponding to combined stages 1 and 2 in PSG), DS, REM sleep [11], and additional events such as autonomic arousals and awakenings.

The classification of the sleep stages was performed in three steps. The algorithm first differentiated between wakefulness and sleep, then sleep epochs were further divided into REM and Non-REM, and finally the Non-REM epochs were classified into LS and DS.

The training set was used to assess whether the performance of the IBI_{PPG} algorithm performed the same as when using the IBI_{ECG}, and to optimize the algorithm in case of underperformance when compared to the IBI_{ECG}. When optimization became satisfactory, the algorithm was applied to the test set, for validation.

The hypnogram of the automated IBI_{PPG} algorithm was compared to the gold standard hypnogram to test for epoch-by-epoch agreement. Comparison with PSG was performed for the period between lights off and lights on. The following basic statistical parameters were used: agreement, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). In addition, we used Cohen's kappa coefficient ('kappa'), which is considered a more robust measure of agreement in cases with imbalance in the occurrence of the different states, as in our case. Cohen's kappa values are usually interpreted as follows: below 0.20 is considered slight agreement, 0.21-0.4 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.8 substantial agreement, and above 0.81 almost perfect agreement. The results presented here relate to the output of the test set only.

3. Results

The test set included a total of 49,726 epochs. For 48,541 epochs (98.6%) PPG was present and adequate for analysis. All those epochs were scored using both our candidate algorithm and gold standard, independent manual evaluation.

An example of a hypnogram obtained using the IBI_{PPG} based algorithm as compared to that of the gold standard scored PSG can be seen in Figure 1. The high similarity between the two hypnograms is evident. One can observe that, at the beginning of the night, the IBI_{PPG} algorithm misses a region of about 12 minutes of sleep, mistakenly scoring it as wake.

The epoch-by-epoch sleep/wake comparison between the output of the IBI_{PPG} automated algorithm with PSG results

yields fair agreement according to the obtained kappa of 0.31. The results indicate a relatively low sensitivity to wake (38%), and very high specificity (92%), as shown in Table 1. The same validation technique, over the same sleep sessions, but when the tested algorithm was the one based on IBI_{ECG}, showed substantially higher agreement with a kappa of 0.46, sensitivity of 52%, and specificity of 93%.

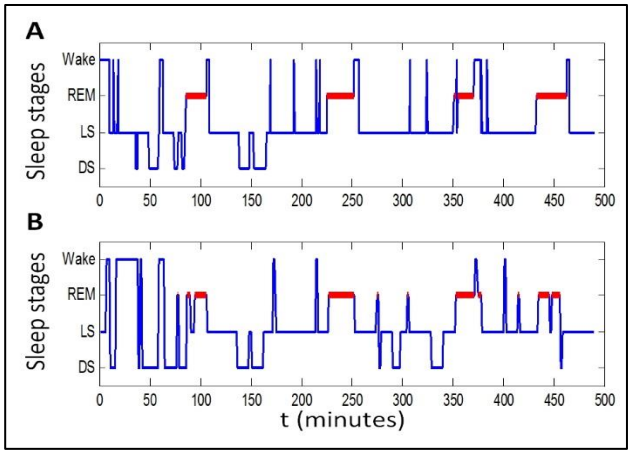

Figure 1. An example of single night hypnogram: A. obtained using gold standard manual scoring, B. Same study scored automatically using the IBI_{PPG} algorithm.

When considering sleep stages only, the ability of the IBI_{PPG} algorithm to differentiate between REM and non-REM showed a sensitivity to REM of 51% and the corresponding specificity of 92%, with kappa of 0.46. These results indicated a much better performance compared to the sleep /wake classification.

Subdividing the non-REM further into DS and LS resulted in a sensitivity and specificity of 76% and 75% respectively with kappa of 0.42.

Table 1. Epoch-by-epoch agreement between sleep stages scored using IBI_{PPG} and gold standard PSG scoring.

	Wake/ Sleep	REM/ NREM	DS/ LS
Agreement (%)	84.3	83.6	75.1
Sensitivity (%)	38.1	51.3	76.8
Specificity (%)	91.7	91.7	74.6
PPV (%)	42.2	61.1	46.9
NPV (%)	90.3	88.2	91.6
Kappa	0.31	0.46	0.42

4. Conclusions

Measurement results depend on the measuring methods used, and when presenting and interpreting results, we need to consider the kind of metric used. Sleep measures differ when we evaluate sleep based on electroencephalogram, electro-occulogram, and muscle activity as in the gold standard PSG, or when we evaluate based through the perspective of autonomic nervous fluctuations, as in our ECG based algorithm [5], or based on observing behaviors (body position, movement, performance of motor tasks etc.), or finally based on cognition and disconnection from the environment. Clearly, sleep detection based on a single signal cannot be expected to issue identical results to those obtained based on information provided by the multiple sensors used in gold standard PSG. Moreover, when manual scoring is used, the interscorer agreement for the same night standard PSG scoring is around 83% only [12]. Our results suggest that the IBI_{PPG} sleep scoring performs very well in differentiating between different sleep stages (LS, DS and REM). The sleep/wake separation, however is less satisfactory.

The sleep/wake classification based on IBI_{ECG} performed significantly better than the one based on IBI_{PPG}. This suggests that there is no inherent limitation when trying to detect sleep/wake based on HRV analysis. The difference in the performance of the two algorithms may be due to errors in the detection of the beat location in the PPG signal, and/or to the PTT variability which is known to affect mainly the high frequency band of the HRV [6]. Further studies should be performed to determine the exact source of the above difference, and to evaluate the contribution of the PTT to the results. The analysis of the PPG pulse wave can contribute additional information and improve the results.

The tested algorithm is a first attempt to establish a method of defining sleep architecture based on signals originating in PPG signals obtained with consumer wrist-worn devices. The PPG device that was used in this study was a medical transmissive sensor (using red and infrared light) positioned on the finger. Most consumer PPG based devices use reflective green light and are positioned on the wrist. In addition, to save battery life, their signal is sampled at lower rates. The limited sampling rate may increase the error in the detection of the beat location, whereas the different location of the PPG on the wrist provides information on the blood flow in different blood vessels (different dimensions, depth and regulation mechanism) than the standard finger PPG, and needs separate testing, optimization, and validation.

Many devices with PPG signals are also equipped with accelerometers that may help improving the separation between sleep and wake. Combining accelerometer and PPG signals in a new sleep algorithm will provide a rather accurate automated detection of the four stages of sleep and wakefulness. Indeed, a recent study [13] presented results of an algorithm for sleep stages detection based on PPG and accelerometer signal. The sleep analysis is based not on the IBI-HRV analysis only, but on cardio-respiratory coupling with the respiration signal obtained

from the PPG. The additional information supplied by the accelerometer and respiratory signals explains the better results presented in the new study [13].

The presented results could be further improved by adding accelerometer data. Further studies of specific wrist-worn PPG devices may help in defining the factors that influence the detection of the IBI series, and when those are better defined the sleep analyser can gain reliability.

No doubt that when using wearable devices to define human physiology and behavior, we need basic testing and understanding of what we measure, and on how those measures relate to those used in the clinical-medical laboratories. Unless standardization of signal measurement, sampling and transmission is defined, each wearable device requires appropriate testing and validation of analysis algorithms.

References

[1] Banks D, Dinges F. Behavioural and physiological consequences of sleep restriction. J Clin Sleep Med 2007;3:519-528.

[2] de Zambotti M, Baker FC, Colrain IM. Validation of sleep-tracking technology compared with polysomnography in adolescents. Sleep 2015;38(9):1461–8.

[3] Poyares D, Hirotsu C, Tufik S. Fitness tracker to assess sleep: beyond the market. Sleep 2015;38(9):1351–2.

[4] Lu S, Zhao H, Ju K, Shin K, Lee M, Shelley K and Chon K. Can photoplethysmography variability serve as an alternative approach to obtain heart rate variability information? J. Clin. Monitoring Comput. 2008;22:23–9.

[5] Decker MJ, Eyal S, Shinar Z, Fuxman Y, et al. Sleep architecture and ventilation derived through analysis of the electrocardiogram. Sleep Breath. 2010;14:233-9.

[6] Gil E, Orini M, Bail´on R, Vergara JM, Mainardi L, Laguna P. Photoplethysmography pulse rate variability as a surrogate measurement of heart rate variability during non-stationary conditions. Physiol. Meas. 2010;31:1271–1290.

[7] Reeves WC, Heim C, et al. Sleep characteristics of persons with chronic fatigue syndrome and non-fatigued controls: results from a population-based study. BMC Neurol 2006, 6:41.

[8] Rechtschaffen A, Kales A, eds. A manual of standardized terminology, techniques and scoring system for sleep staging in human subjects. US Government Printing Office, NIH Publication 204, 1968.

[9] Penzel T, Kantelhardt JW, Grote L, Peter JH, Bunde A. Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea. IEEE Trans Biomed Eng. 2003; 50(10): 1143–1151.

[10] Shinar Z, Baharav A, Dagan Y, Akselrod S. Automatic detection of slow-wave-sleep using heart rate variability. Computers in Cardiology 2001;28:593-6.

[11] Baharav A, Kotagal S, Gibbons V, Rubin BK, Pratt G, Karin J, Akselrod S. Fluctuations in autonomic nervous activity during sleep displayed by power spectrum analysis of heart rate variability. Neurology 1995;45:1183-7.

[12] Rosenberg RS, Van Hout S. The American Academy of Sleep Medicine inter-scorer reliability program: sleep stage scoring. J Clin Sleep Med 2013;9(1):81-7.

[13] Fonseca P, Weysen T, Goelema MS, Møst EIS, et al. Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults. Sleep 2017;40(7):1-10.

Address for correspondence.

Shuli Eyal.
20 Hamagshimim St,
Petach Tikva, 4934829, Israel.
shuli@sleeprate.com