UNIVERSITY OF BOLOGNA
SCUOLA di Architettura ed Ingegneria

**Two-year master in**

**Biomedical Engineering**

Marcello Sicbaldi
927417

Project for Laboratory of Wearable Sensors and Mobile Health
A.A 2020/21

# Evaluation of PPG features during sleep and wake

# CONTENTS

# INTRODUCTION

Sleep/wake classification has been used both in clinical fields and personal health/wellness fields. The gold standard for sleep measurement, polysomnography (PSG), can be too expensive and unfeasible for use in large scale population research. Therefore, wearable accelerometers have been explored since the mid-1990s as a possible alternative for multi-day real life (out of the lab) sleep detection, as they are both low in price and provide reasonably accurate estimates of movement. Moreover, developing an automatic algorithm for sleep detection could be thought as the first step of a pipeline devoted to compare different physiological processes during sleep and wake. It is in this scenario that a wearable device such as Empatica E4 comes in handy: since it records multiple physiological signals simultaneously, it allows to compare different features due to different physiological phenomena. In the specific case of this project, the goal is to first design a sleep detection algorithm based on accelerometer data, and then compare PPG feature values during sleep and wake. In particular, the focus will be on comparing heart rate variability (HRV) indexes rather than morphological characteristics of PPG signals. HRV consists of changes in the time intervals between consecutive heartbeats called inter-beat intervals (IBIs). HRV indexes neurocardiac function and is generated by heart-brain interactions and autonomic nervous system (ANS) processes. HRV can be estimated using the PPG signal, since the location of a peak represents the instant of time at which a heartbeat occurs. Thus, the computation of HRV consists in the identification of the location of peaks in the PPG signal and then in the computation of inter-beat intervals.

# DATA

The data consists in accelerometers and PPG recordings of 10 subjects, 6 males and 4 females, between the age of 23 and 34.

| # Subject | Gender | Age | Recording Length |
|---|---|---|---|
| 1 | M | 33 | 26:39:27 |
| 2 | F | 34 | 24:28:53 |
| 3 | M | 24 | 24:04:00 |
| 4 | M | 23 | 24:00:24 |
| 5 | F | 27 | 24:33:20 |
| 6 | M | 26 | 43:34:40 |
| 7 | F | 29 | 25:07:16 |
| 8 | F | 25 | 25:03:54 |
| 9 | M | 23 | 23:58:03 |
| 10 | M | 25 | 24:55:21 |

All recordings don't show any particular anomalies with the exception of subject 6, for whom both the PPG and accelerometer data shows a very low level of activity after roughly 00:00 (figure 1). A possible hypothesis is that the subject stopped wearing the device, leaving it still somewhere. This

could create issues in the detection of the sleep window, since inactivity due to non-wear time could be attributed erroneously to sleep. For this reason, the signal was clipped from that time till the end of the recording. This does not create any problem in the following HRV analysis, since for this subject the whole recording lasted more than 43 hours (comprehending 2 nights), so a sleep window was detected regardless of the clipping.

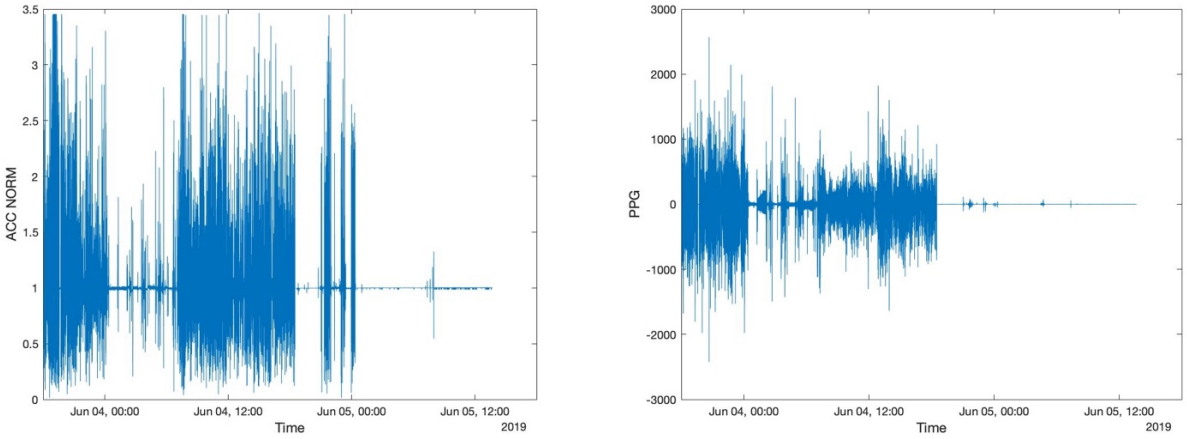


*Figure 1: Accelerometer and PPG data of subject 6. The signal was clipped in such a way to assure a correct detection of the first sleep window (≈ Jun 04, 00:00 – Jun 04, 7:00) and to discard inactivity levels due to possible non-wear time.*

## SLEEP DETECTION

The sleep detection algorithm is inspired by the heuristic algorithm developed by van Hees et al. [1], which is based on a thresholding of the arm angle. In particular, wrist-worn accelerometer data allow estimation of the arm angle relative to the horizontal plane (Z-angle), which reflects posture changes of the subject and can be estimated as follows:

$$angle_z = \frac{180}{\pi} \cdot \tan^{-1} \frac{a_z}{a_x^2 + a_y^2}$$

where $a_x$, $a_y$ and $a_z$, are the median values of the three orthogonal raw accelerations in gravitational ($g$) units (1 $g$ = 1000 m$g$), derived based on a rolling five second time window. For the Empatica E4 sensor, the z-axis corresponds to the axis positioned perpendicular to the skin surface (dorsal-ventral direction when the wrist is in the anatomical position), exactly as for the accelerometers used in [1]. Sleep is characterized as a period marked by a low frequency of changes in arm angle (figure 2): this suggests that thresholding it could allow to detect sleep periods.

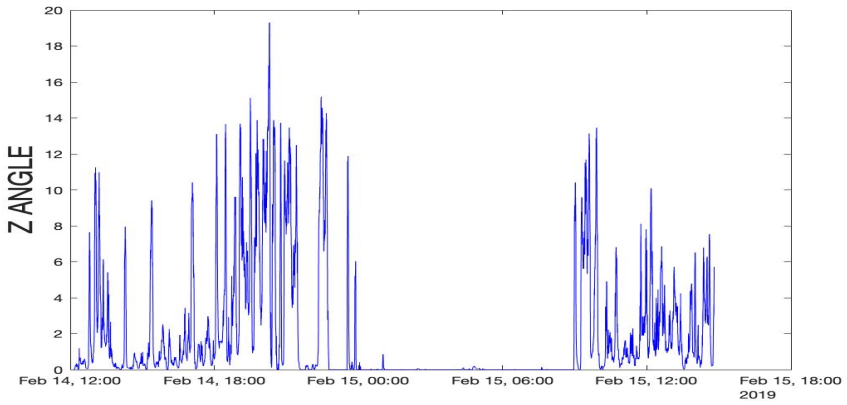


*Figure 2: Z-angle derived from accelerometer data of subject 1. It is very clear that from roughly midnight to 7:00 the level of the Z-angle is very low, as the subject was sleeping.*

The sleep detection algorithm consists in the following steps:

1. Compute a 5-second rolling median of the raw accelerations $a_x$, $a_y$ and $a_z$.
2. Compute the Z-angle using the accelerations obtained at step 1.
3. Compute averages of the Z-angle every 5 consecutive seconds.
4. Compute absolute differences between the successive 5-second averages.
5. Compute a 5-minute rolling median of the absolute differences.

These first five steps make the algorithm invariant to the potentially unstandardized orientation of the accelerometer relative to the wrist.

6. Compute the sleep/wake threshold as the $10^{th}$ percentile from the output of step 5 multiplied by 15.

Using a percentile as part of the threshold calculation allows the threshold to account for between-individual differences in z-angle distribution.

7. Detect the observation blocks for which the output from step 5 was below the critical threshold and keep the ones lasting longer than 2 hours.
8. Evaluate the length of the time gaps between the observation blocks identified by step 7, if the duration is less than 10 minutes then count these gaps towards the identified blocks.

The last three steps reflect assumptions about the nature of sleep. In particular, van Hees et al. used a multiplicative factor of 15 in step 6, stating that it was derived iteratively using visual inspection of the classification: that value worked well also with the data of this project. For step 7 they used a 30-minute time period, according to the assumption that people are typically not in bed for less than 30 minutes when they sleep. I extended this assumption to 2 hours, which is more appropriate for the goal of the project, that is to compare HRV during night and day: I want to discard naps or inactive periods during the day, which are not likely to last more than 2 hours. I care only about

identifying nocturnal sleep, which reasonably is constituted of several blocks of 2 straight hours of sleep which may interrupted by short awakenings. At this point only one thing is missing: the need to discard short wake periods during the night and consider them a part of the sleep period. The crucial aspect here is the choice of the maximum acceptable duration of this nocturnal wake:

- I do not want too large of a window (e.g., 60 minutes like in van Hees approach), because this long nocturnal wake period could interfere with HRV values during sleep.
- I do not want too small of a window, since this could cause to lose too much sleep data (if the subject wakes up for only 10 minutes three times during the night and I choose a window of 5, then I would get several sleep windows of much shorter duration with respect to the actual one).

I then need to find a tradeoff between these two considerations: I want to preserve as much sleep data as possible by including nocturnal awakenings, but I do not want them to be too long so that they can impact sleep HRV. A window of 10 minutes was chosen, and it proved to work well. The only subjects that experienced nocturnal activity were subjects 5 and 7, both lasting approximately 7 minutes.
A visual representation of what just described is given in figure 3 and 4.

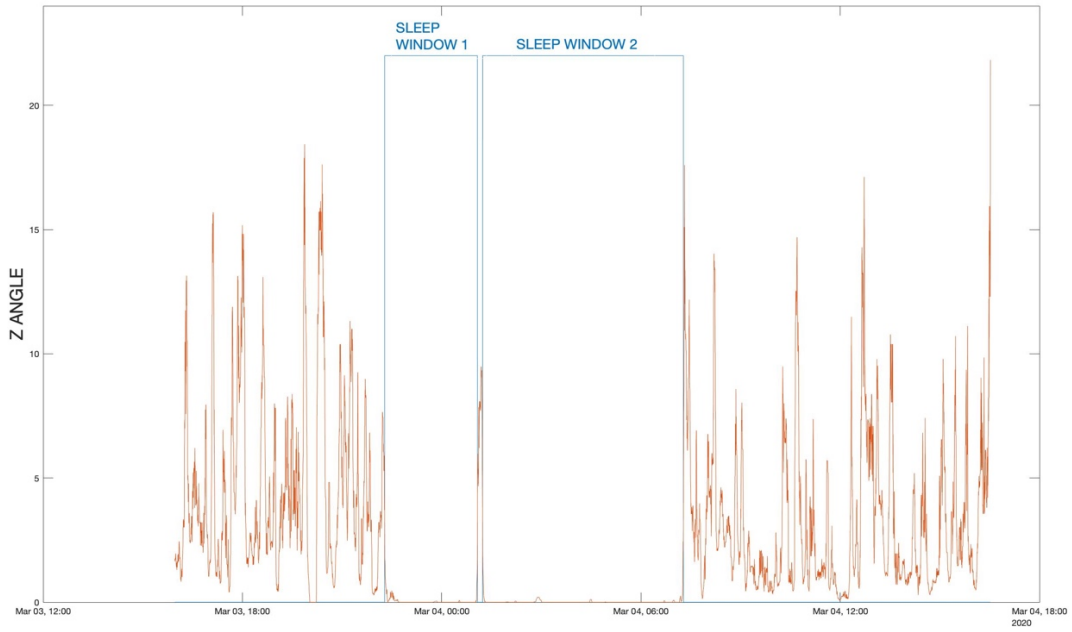


*Figure 3: output of step 7 of the algorithm, subject 5. We can see that there are two time periods in which the Z-angle is less than the threshold for more than 2 hours. Between these two windows there is a short interval in which the arm angle is far higher than the threshold, which most likely corresponds to a brief awakening during the night.*

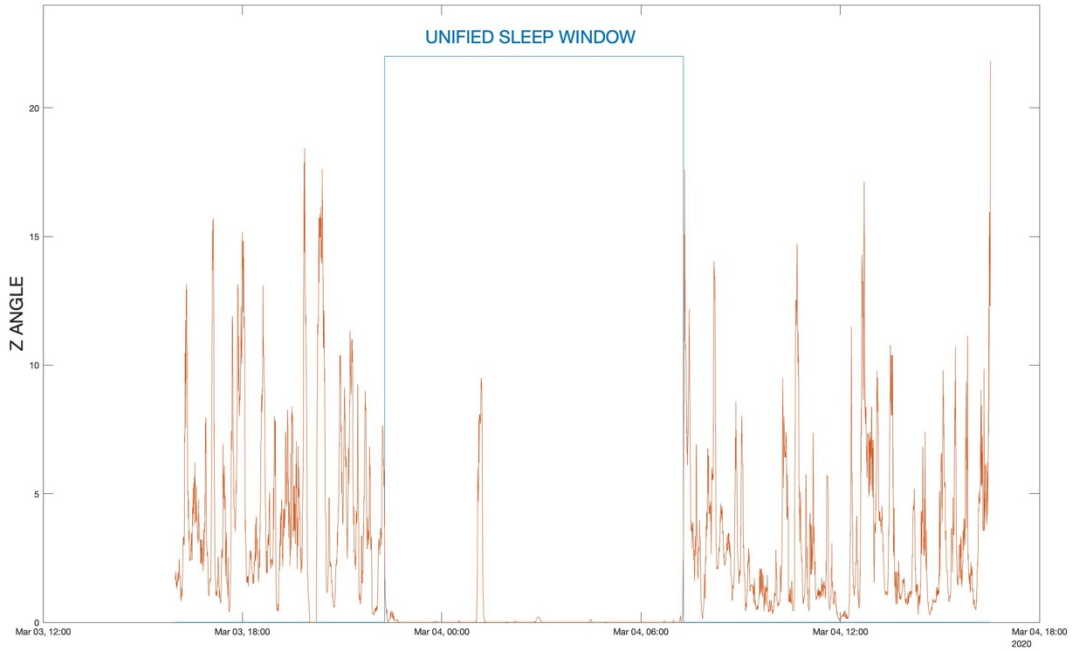


*Figure 4: output of the final step (8) of the algorithm, subject 5. Since the nocturnal wake detected in step 7 lasts for less than 10 minutes, the two sleep windows are unified into one.*

## DIVISION IN PPG WINDOWS OF SLEEP-WAKE

The sleep interval detected using accelerometer data was then used to split the PPG signal in windows of sleep and wake. In order to correctly compare inter-subject HRV values, I selected a window of sleep and wake of the same duration for every subject. To do this I identified the shorter sleep window between all subject (subject 8, 6 hours 24 minutes and 40 seconds) and applied it at the center of all other sleep windows. For what concern the wake, a window of that duration was selected depending on the subject, by qualitatively identifying the more active time during the day observing accelerometer data.

A visual example of what just said can be found in figure 5.

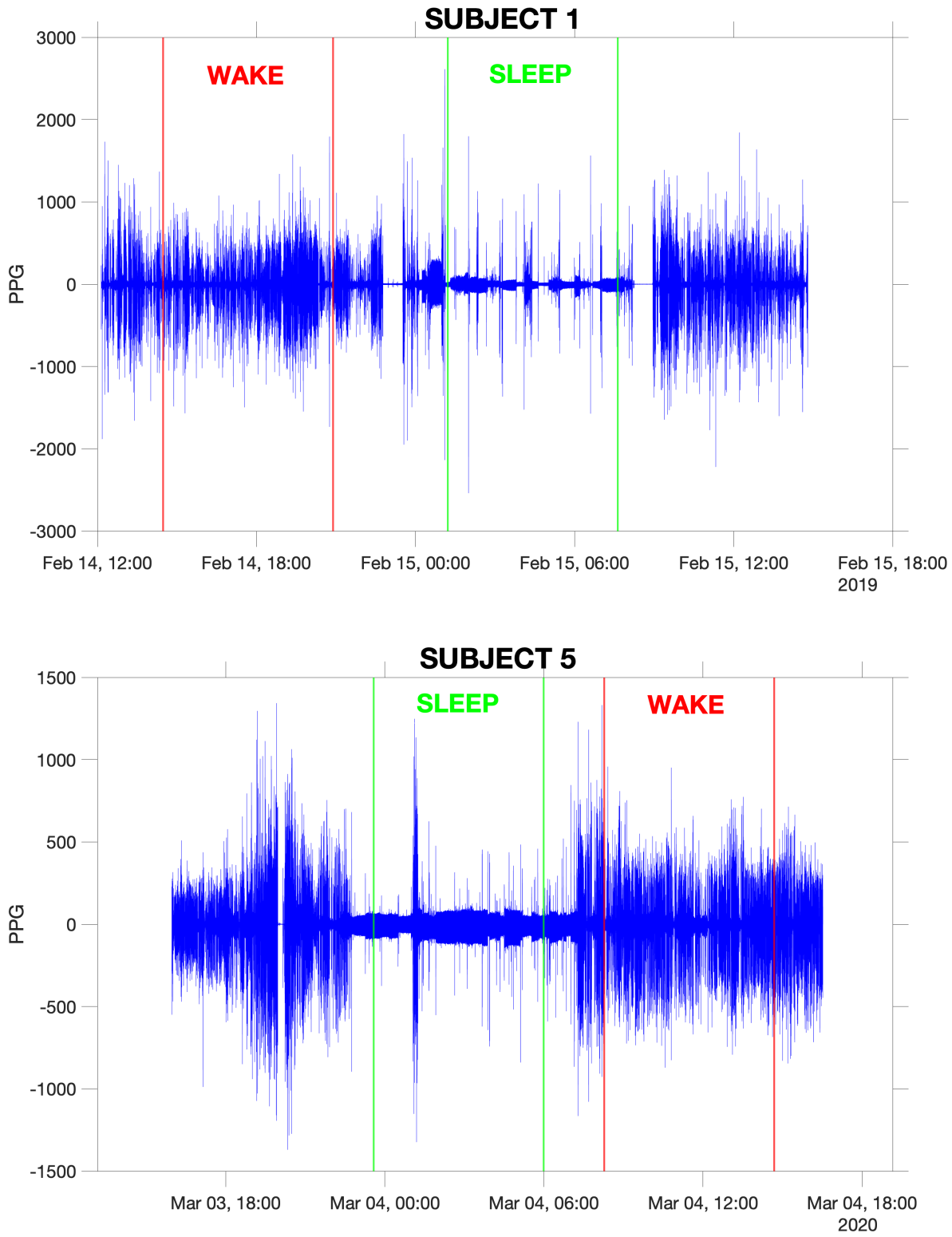


*Figure 5: PPG intervals of sleep and wake of two subjects. All windows have the same duration of 6h 24min and 40sec. The wake window is chosen depending on the subject and his more active time (qualitatively) during the day.*

## EVALUATION OF HRV FEATURES

The first step needed for the evaluation of HRV features is the detection of inter-beats intervals from the PPG signal. Unfortunately, said signal is highly affected by motion artifact, which hinders the extraction of reliable IBIs especially during wake, when the subject is moving a lot. The Empatica E4 device already provide an IBIs sequence obtained from the processing of the PPG signal, with an algorithm that removes incorrect peaks due to noise in the signal. However, for data where the subject is moving for more than 30% of the time (highly likely during wake), the given IBIs are not reliable enough to compute heart rate variability [2]. For this reason, another algorithm is used (provided by Serena Moscato), which evaluates PPG signal quality and again provides the indices of the pulses that can be used for subsequent analysis.

By visually observing the IBIs extracted by the algorithm through a boxplot (figure 6), we find some outliers values that are statistically impossible (e.g., 3 seconds). Therefore, to further clean the sequence, I set a threshold to eliminate extreme outliers, defined as those values x such that

$$x > Q3 + 3 \cdot IQR \ \ or \ \ x < Q1 - 3 \cdot IQR,$$

where Q3 = third quartile, Q1 = first quartile, IQR = interquartile range.

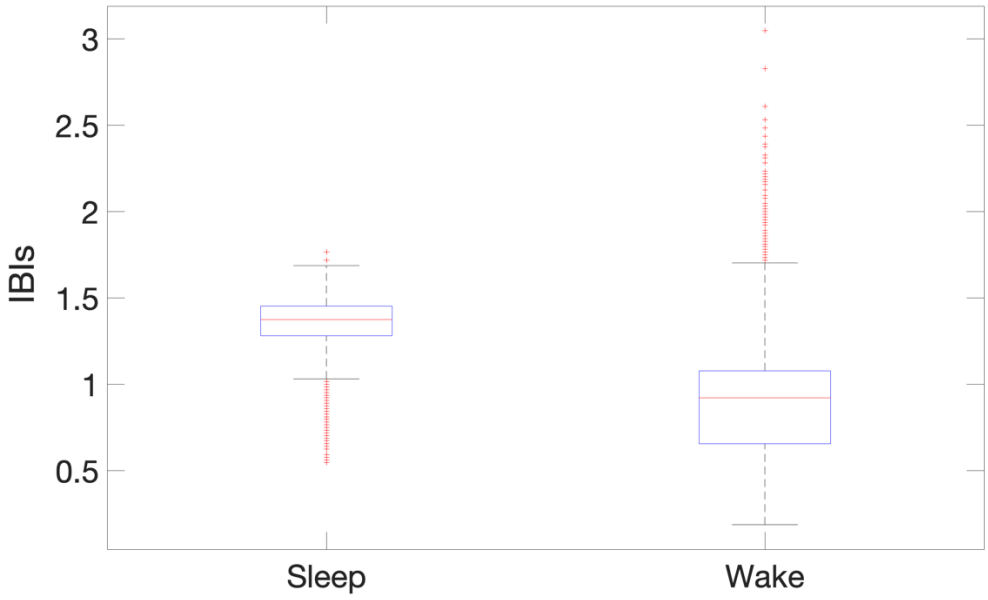


*Figure 6: IBIs distribution, subject 4. We can see unrealistic values especially during wake.*

The IBIs sequence is now ready to be analyzed, providing information on HRV.

The first measurement I derived from the sequence is the heart rate (HR), defined as $HR\ [bpm] = mean(\ \frac{60}{IBIs}\ )$, where IBIs is the sequence.

### *Time-domain features*

Time-domain indices of HRV quantify the amount of variability in measurements of the inter-beat intervals. To compare HRV during sleep and wake the following indices were chosen [3]:

- *SDNN:* standard deviation of normal beats, reflects the total variability of the IBI series (how dispersed is the data in relation to the mean value).

$$SDNN = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(IBI_i - \overline{IBI})}$$

- *RMSSD:* root mean square of successive differences between heartbeats, reflects beat-to-beat variability.

$$RMSSD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(IBI_{i+1} - IBI_i)^2}$$

### *Frequency-domain features*

Frequency-domain measurements estimate the distribution of power of the IBIs sequence into four frequency bands. These measurements are important in establishing a possible role of the sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) in modulating the HR.
The Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (1996) divided heart rate oscillations into ultra-low-frequency (ULF), very-low-frequency (VLF), low-frequency (LF), and high-frequency (HF) bands.

For this project the following metrics were chosen:

- *LF power*: power in the low-frequency band (0.04 – 0.15 Hz). It reflects both PNS and SNS activity.
- *HF power*: power in the high-frequency band (0.15 – 0.40 Hz). Reflects PNS activity and is influenced by breathing.
- *LF/HF ratio*: ratio between LF and HF power. Low LF/HF ratio reflects parasympathetic dominance while high LF/HF ratio indicates sympathetic dominance.

### *NON-LINEAR FEATURES*

Non-linear measurements reflect the unpredictability of the time series, which results from the complexity of the mechanisms that regulate HRV. Non-linear indices can be computed starting

from the Poincarè plot, which is graphed by plotting every inter-beat interval against the prior interval, creating a scatter plot. A Poincarè plot can be analyzed by fitting an ellipse to the plotted points. After fitting the ellipse, we can derive the following non-linear measurements:

- *SD*1: Standard deviation of the distance of each point of the Poincarè plot from the *y = x* axis, specifies the ellipse's width. Reflects short-term HRV and correlates with HF power.
- *SD2*: standard deviation of each point from the *y = x + average IBI,* specifies the ellipse's length. Reflects short and long-term HRV and correlates with LF power.

## COMPARISON OF FEATURES VALUES DURING SLEEP AND WAKE

Before diving into the results of the experiment, it is necessary to make some considerations about the IBIs data. As stated before, it is impossible to measure a continuous IBIs sequence using a wearable device because of the nature of photoplethysmography, which is susceptible to motion noise. Therefore the IBIs sequence, obtained by keeping only the reliable pulses, will present a lot of missing data, the more the higher is the movement of the subject.
Several studies ([4], [5], [6]) have investigated the effect of missing IBIs data on HRV analysis. For what concerns the time-domain, the most robust feature to missing data is the heart rate, while SDNN is the most affected, but still more robust than all frequency domain features, which are highly unreliable in the presence of missing IBIs. Moreover, also non-linear features are affected by this: hence the result obtained using features of all domains, with the exception of the heart rate, are difficult to interpret and may be meaningless. In the following, an attempt is made to extract as more meaningful information as possible from the results.

For what concerns heart rate, the results were satisfying, confirming its robustness to missing IBIs data. We can observe a lower heart rate during sleep, with a mean value across subjects of 59 bpm and a higher HR during wake, with a mean value of 80 bpm. In figure 7 it is shown a more complete representation of HR pattern between sleep and wake for all subjects.

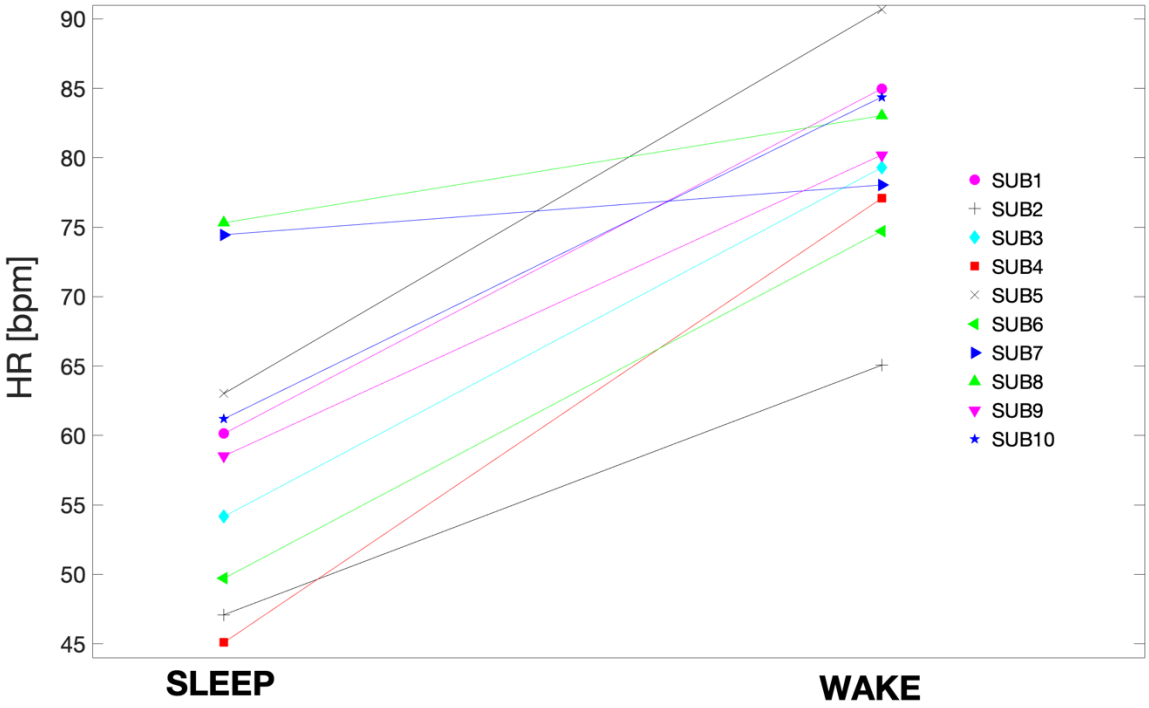


*Figure 7: HR of all subjects during sleep and wake. We can appreciate how HR varies between subjects during sleep and wake separately, but also how much HR changes for every single subject between sleep and wake. The lines connect the sleep HR with the wake HR for each subject, and the fact that they all have a positive slope shows that HR increased from sleep to wake for every subject. The higher the slope the higher the change in HR.*

Regarding HR, we can also observe that women had a higher heart rate during sleep compared to men, while the two groups had a similar heart rate during wake.

| | **MEAN HR, SLEEP** | **MEAN HR, WAKE** |
|---|---|---|
| **MEN** | 54.8 bpm | 79.2 bpm |
| **WOMAN** | 65.0 bpm | 80.1 bpm |

In figure 8 plots representing the remaining time-domain features are shown, although their interpretation remains doubtful due to the effect of missing data.

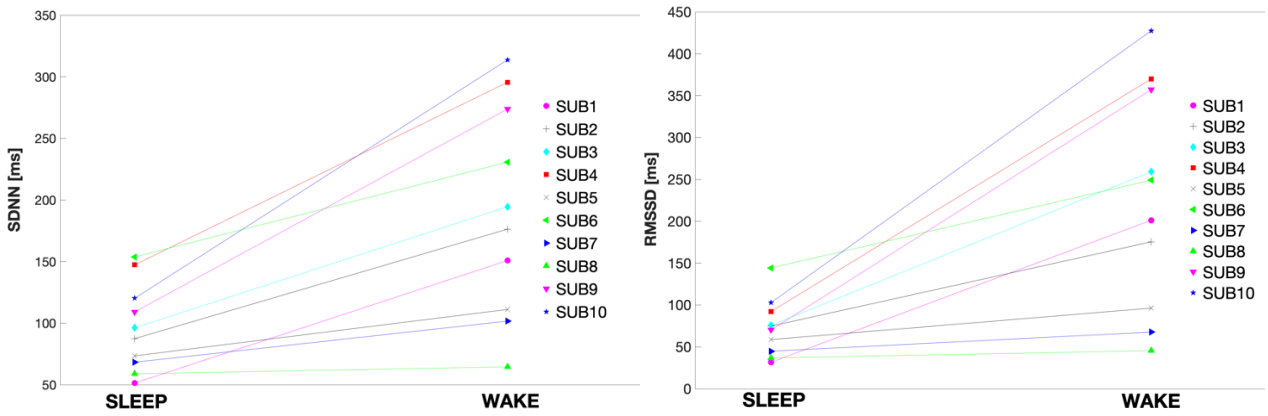

*Figure 8: SDNN and RMSSD during sleep and wake across subjects. We can observe that both increase for every subject from sleep to wake.*

As stated before, frequency domain features are the most affected by missing data. This explains the results obtained: I expected to observe an increase in LF/HF ratio during wake (sympathetic dominance) and a decrease during sleep (parasympathetic dominance). Instead the results show the contrary, with a mean LF/HF ratio of 2.1796 during sleep and 1.1198 during wake. Some studies shows that sympathetic activity can actually increase during REM sleep, causing a higher LF/HF ratio with respect to the wake, but this is not the case for NREM sleep, where PNS activity is dominant [7]. Since NREM sleep constitutes about 75 to 80 percent of total time spent in sleep, one should still expect to have an overall lower LF/HF ratio during sleep, which is not what we observe.

One thing I thought to do at this point was to see if the results were *consistently* wrong among all domains, i.e. if missing data had the same effect on features from different domains, exploiting knowledge about their correlations. To better explain with an example: we know that for long recordings, RMSSD is highly correlated with HF band power: in healthy people they both decrease during wake [8], while in this project they both increase (figure 8). The erroneous results are likely due to missing IBI data, but at least we proved that missing data alters HRV features from different domains in the same way, preserving their correlation. The same can be proved for others features, including non-linear ones. In the following table, correlations coefficient between different features are shown, to verify that missing data does not alter these known correlations.

| | **SLEEP** | **WAKE** |
|---|---|---|
| **RMSSD and HF POWER** | r = 0.9392 | r = 0.9632 |
| **SDNN and LF POWER** | r = 0.8812 | r = 0.9594 |
| **LF POWER and SD2** | r = 0.8456 | r = 0.9375 |

This table verifies known correlations between features from different domains.

Moreover, RMSSD conveys the same information as SD1 [3], and in fact their correlation coefficient is 1.

# CONCLUSIONS

This project shows the limitations of measuring HRV in everyday life using wrist wearable PPG sensors, as motion artifacts introduce lots of missing data in the IBIs sequence, which in turn affects most HRV measurements. To get more insights about differences in HRV between sleep and wake it is necessary to use more sophisticated instrumentation, for example a 24-h Holter ECG monitoring. The best-case scenario would be to have access to a polysomnography (to overcome limitations due to the automatic sleep detection algorithm and distinguish between different sleep phases) and an ECG system, but this of course would come at the cost of losing the advantages of

wearable sensors, such as wearability, ease of use and ability to collect data in free-living environments, offering insight on habitual behaviors.

# Bibliografia

[1] van Hees et al. *Estimating sleep parameters using an accelerometer without sleep diary,* 2018.

[2] https://support.empatica.com/hc/en-us/articles/360030058011-E4-data-IBI-expected-signal

[3] Shaffer et. al. *An Overview of Heart Rate Variability Metrics and Norms, 2017.*

[4] Baek, Shin. *Effect of Missing Inter-Beat Interval Data on Heart Rate Variability Analysis Using Wrist-Worn Wearables*, 2017.

[5] Kim et. al. *Effect of missing RR-interval data on heart rate variability analysis in the time domain*, 2007.

[6] Kim et. al. *The effect of missing RR-interval data on heart rate variability analysis in the frequency domain*, 2009.

[7] Elsenbruch et. al. *Heart Rate Variability During Waking and Sleep in Healthy Males and Females*, 1999

[8] Vanoli et. al. *Heart Rate Variability During Specific Sleep Stages: A Comparison of Healthy Subjects with Patients After Myocardial Infarction*, 1995