

# Evaluation of Smartwatch and Wristband-based Sleep Analyses

## 1. Background and Objective

Our attention, memory and mood are strongly associated with our quality of sleep. Many studies have shown that various pathological conditions can be caused by poor sleep quality. Nowadays, the concept of sleep health management is becoming well accepted, and consumer sleep technologies are commonly utilized in mobile devices including high-end wristbands and smartwatches. However, the accuracy of sleep evaluation based on these wearable consumer devices has not been studied systematically. Most wearable devices on the market are entertainment-oriented, and cannot generate reliable sleep assessments. The objective of this study is to evaluate the sleep analysis report obtained by different models of wristband and smartwatch manufactured by Huawei. Our evaluation is based on the comparison with a well-accepted ECG-based sleep analysis, known as Cardiopulmonary Coupling (CPC). The CPC algorithm has been approved by FDA in the U.S. and China. The advantage of using the CPC analysis is that it has been extensively verified in sleep medicine, while at the same time wearing a compact one-lead ECG causes minimal disturbance to a test subject's sleep.

## 2. Method

In this study, various models of smartwatch and wristband made by Huawei were tested. These devices include a photoplethysmography (PPG) sensor for heart rate monitoring, they also implemented a sleep analysis algorithm based on heart rate variability developed by Nanjing Fengsheng Yongkang Software Technology Co., Ltd.

All subjects wear both a single-lead ECG monitor and a wristband (or smartwatch) to record their signals. All subjects were asked to recall their sleep for the test night, including 1) time for bed, 2) time to fall asleep (sleep latency) and 3) wake-up time in the morning. The bed time and wake-up time were used to define the start and end points for sleep time. The data from wristband (or smartwatch) and ECG recordings were extracted and analyzed by wristband-based sleep analysis algorithm and CPC analysis, respectively.

Healthy subjects from three Chinese cities (Dongguan, Suzhou and Nanjing) were recruited in this study. To exclude subjects with bad signals, we evaluate the PPG signal quality by the amount of misdetection of heartbeat (based on the PPG signal). Detailed statistics of test subjects will be described in the Results section.

In this study, sleep was classified into four states: stable sleep (SS, also denoted as “deep” sleep), unstable sleep (US, also denoted as “light” sleep), REM sleep (REM) and unknown state. Only the first three states were used as outcomes, while the unknown states, mostly due to bad signal quality, were eliminated from both outputs. The classification by the

CPC analysis is treated as the true events. To investigate the accuracy of the classification obtained by wristband/watch, seven measures were calculated: composite sleep duration; stable sleep duration; unstable sleep duration; REM sleep duration; detection of stable sleep epoch (each epoch is 1 minute); detection of unstable sleep epoch; detection of REM sleep epoch.

Accuracy of detection is defined as  $(TP+TN)/(TP+TN+FP+FN)*100\%$ .

TP: number of true positive;

TN: number of true negative

FP: number of false positive;

FN: number of false negative

Accuracy of duration is defined as  $(1-|T_{PPG} - T_{CPC}|/T_{CPC})*100\%$

where  $T_{PPG}$  indicates the time duration derived from wristwatch-based analysis, and  $T_{CPC}$  indicates the time duration derived from CPC analysis.

Composite accuracy of duration is defined as

$$1 - \frac{D_{stable\ sleep} + D_{unstable\ sleep} + D_{REM\ sleep}}{2 * T_{total}},$$

where  $D$  indicates the absolute value of difference of  $T_{PPG}$  and  $T_{CPC}$ , and  $T_{total}$  indicates the summation of time duration.

### 3. Results

Various wearable devices have been tested by our group. Their results are listed below.

#### 3.1 HUAWEI FIT Smartwatch

Healthy subjects from 258 subjects were collected. To exclude subjects with bad signals, we evaluate the PPG signal quality by the amount of misdetection of heartbeat (based on the PPG signal), and use the subset of 200 subjects with the best signal quality. Their gender distribution: 87 males (43.5%); and age range: 18 – 45 years (median age 27yr). The subjects reported total sleep time (TST) ranged from 135 to 550 minutes (median TST = 405min).

The median values of accuracy of six measures are the following:

- 93.84% for composite sleep duration
- 88.78% for stable sleep duration
- 90.71% for unstable sleep duration
- 85.62% for REM sleep duration
- 84.70% for stable sleep epoch detection
- 75.29% for unstable sleep epoch detection
- 81.12% for REM sleep epoch detection

Our results show that the classifications obtained by the HUAWEI FIT smartwatch and the ECG-based CPC analysis are consistent.

### **3.2 HUAWEI Band 2 Pro Wristband**

The Band 2 Pro model use the same PPG sensor as in the FIT model, and they have the same heart rate monitoring and sleep analysis functions. Thus, an additional 23 healthy subjects were recruited to verify that whether the result is consistent with that of the HUAWEI FIT. Their gender distribution: 15 males (65.2%); and age range: 23 – 33 years (median age 27yr).

The median values of accuracy of six measures are the following:

- 92.16% for composite sleep duration
- 84.09% for stable sleep duration
- 82.61% for unstable sleep duration
- 87.18% for REM sleep duration
- 82.57% for stable sleep epoch detection
- 73.68% for unstable sleep epoch detection
- 80.34% for REM sleep epoch detection

Our results show that the classifications obtained by the HUAWEI Band 2 Pro wristband are consistent with the ECG-based CPC analysis.

### **3.3 HUAWEI Watch 2 / Watch 2 Pro Smartwatch**

The Watch 2 and Watch 2 Pro models are identical in term of their heart rate monitoring and sleep analysis functions. Furthermore, they use the same PPG sensor as in the FIT model. Thus, an additional 20 healthy subjects were recruited to verify that whether the result is consistent with that of the HUAWEI FIT. Their gender distribution: 16 males (80%); and age range: 24 – 41 years (median age 26yr).

The median values of accuracy of six measures are the following:

- 93.7% for composite sleep duration
- 86.5% for stable sleep duration
- 92.4% for unstable sleep duration
- 79.4% for REM sleep duration
- 87.4% for stable sleep epoch detection
- 74.4% for unstable sleep epoch detection
- 81.5% for REM sleep epoch detection

Our results show that the classifications obtained by the HUAWEI Watch 2 and Watch 2 Pro smartwatch are consistent with the ECG-based CPC analysis.

### **3.4 HUAWEI Talkband B5 Wristband**

The Talkband B5 model is identical in term of their heart rate monitoring and sleep analysis functions. Furthermore, they use the same PPG sensor as in the FIT model. Thus,

an additional 20 healthy subjects were recruited to verify that whether the result is consistent with that of the HUAWEI FIT. Their gender distribution: 16 males (80%); and age range: 23 – 36 years (median age 28.4yr).

The median values of accuracy of six measures are the following:

- 95.1% for composite sleep duration
- 90.0% for stable sleep duration
- 92.1% for unstable sleep duration
- 91.5% for REM sleep duration
- 87.7% for stable sleep epoch detection
- 73.5% for unstable sleep epoch detection
- 77.0% for REM sleep epoch detection

Our results show that the classifications obtained by the HUAWEI Talkband B5 wristband is consistent with the ECG-based CPC analysis.

### **3.5 honor Band 4 Wristband**

The honor Band 4 model is identical in term of their heart rate monitoring and sleep analysis functions. Furthermore, they use the same PPG sensor as in the Huawei FIT model. Thus, an additional 24 healthy subjects were recruited to verify that whether the result is consistent with that of the HUAWEI FIT. Their gender distribution: 14 males (58%); and age range: 22 – 35 years (median age 26yr).

The median values of accuracy of six measures are the following:

- 92.1% for composite sleep duration
- 83.1% for stable sleep duration
- 84.9% for unstable sleep duration
- 82.0% for REM sleep duration
- 78.0% for stable sleep epoch detection
- 66.4% for unstable sleep epoch detection
- 78.1% for REM sleep epoch detection

Our results show that the classifications obtained by the honor Band 4 wristband is consistent with the ECG-based CPC analysis.