## UniversidadeVigo

Citation for published version:<br>de Arriba-Pérez, F., Caeiro-Rodríguez, M. \& Santos-Gago, J.M. How do you sleep? Using off the shelf wrist wearables to estimate sleep quality, sleepiness level, chronotype and sleep regularity indicators. J Ambient Intell Human Comput 9, 897-917 (2018).<br>https://doi.org/10.1007/s12652-017-0477-5

## Peer reviewed version

This version of the article has been accepted for publication, after peer review and is subject to Springer Nature's AM terms of use but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at https://doi.org/10.1007/s12652-017-0477-5

## General rights:

Copyright © 2017, Springer-Verlag Berlin Heidelberg

# How do you sleep? Using off the shelf wrist wearables to estimate sleep quality, sleepiness level, chronotype and sleep regularity indicators 

Francisco de Arriba-Pérez*, Manuel Caeiro-Rodríguez and Juan M. Santos-Gago<br>Department of Telematics Engineering, University of Vigo, Campus Lagoas-Marcosende, 36310 Vigo, Galicia | Spain farriba@uvigo.es, Manuel.Caeiro@det.uvigo.es, Juan.Santos@det.uvigo.es<br>*Correspondence: farriba@uvigo.es; Tel.: +34 986814073


#### Abstract

This piece of research is situated in the domain of multi-modal analytics. New commercial off the shelf wearables, such as smartwatches or wristbands, are becoming popular and increasingly used for fitness and wellness in a new trend known as the quantified-self movement. The sensors included in these devices (e.g. accelerometer, Heart Rate) in conjunction with data analytics algorithms are used to provide information such as steps walked, calories consumed, etc. The main goal of this piece of research is to check if new wearable technologies could be used to estimate sleep indicators in an automatic way. The available medical literature proposes several sleep-related features and methods to calculate them involving direct user observation, interviews or specific medical instrumentation. Off the shelf wearable vendors also provide some sleep indicators, such as the sleep duration, the number of awakes or the time to fall asleep. Taking as a reference the results and methods described in the medical literature and the data available in commercial off the shelf wearables, we propose new sleep indicators offering a greater interpretative value: sleep quality, sleepiness level, chronotype. The results obtained after initial experiments demonstrate the feasibility of this approach to be applied in real contexts. Eventually, we plan to apply these solutions to support educational scenarios related to selfregulated learning and teaching support.


Keywords: sleep quality; sleepiness; chronotype; sleep regularity, wearables; wearables analytics.

## 1. Introduction

Wearable technology (Crabtree and Rhodes 1998; K Tehrani 2014) refers to electronic devices and systems incorporated in some part of our body or clothes. Nowadays, smartwatches, sneakers with built-in GPS or wristbands are examples of this technology that are becoming popular as a new kind of off the shelf electronics. These systems can be used for different purposes, particularly to monitor physiological and environmental data. In a previous study (de Arriba Pérez et al. 2016a) we identified more than 17 sensors in more than 140 commercial off the shelf wearables that could be used for these purposes, such as: accelerometer, heart rate, GPS, gyroscope and compass, microphone, ambient light, barometer, altimeter, camera, thermometer, etc.

The adoption of commercial off the shelf wearable devices is experiencing a great rising, mainly in relation to fitness and wellness activities (IDC 2016a). Indeed, a new kind of philosophy is being developed: the quantified-self movement (Ermes et al. 2008). This movement embraces people interested on monitoring their activities and healthrelated features in a continuous basis. These people can be motivated towards different goals, such as to investigate some health problem, to make progress towards a training goal, or just simply because they are curious. Taking advantage of the new wearables people can collect data about what they are doing and about how their body is behaving at all time in a transparent and autonomous way. Later, the data collected can be processed by data analytics techniques
and prepared for attractive graphic representations showing details such as the steps walked, calories consumed, exercise time, etc. There exist many criticisms and scepticism related to the validity and usefulness of data collected from such off the shelf devices. Several studies related with accuracy topic have been published (Richmond; Natale et al. 2012; Guo et al. 2013; Valenti and Westerterp 2013). Despite this, it is accepted that the new devices can be used for personal purposes, but not for medical ones. In any case, there exists some experiences that show these devices can be used as a previous stage before clinical tests (Paul Burton 2015; Wallis Snowdon 2016; Rudner et al. 2016).

The goal of this piece of research is to use affordable off the shelf wearables to estimate sleep indicators without requiring user interaction in educational scenarios. There are several reasons that motivated us to work towards this goal in this domain. First, the scientific literature demonstrates that sleep has influence in the ability to acquire new knowledge, skills and competences and consequently on the academic achievement of the students (Medeiros et al. 2003; Dewald et al. 2010). There are research projects that specifically study the relationships between sleep features and cognitive/learning processes, such as the relation between the student motivation and the sleepiness in classroom (Horzum et al. 2014), or the impact of sleep deprivation in the student performance (Belenky et al. 2003). Other researchers have identified a positive correlation between sleep quality and academic performance (Curcio et al. 2006), between alterations of sleep rhythms and levels of care when working with high levels of concentration, like the study of a subject (Lockley et al. 2004; Johnston 2005), etc. In addition, a greater daytime somnolence, coming from a bad sleep quality, can be detrimental to students' cognitive functions (Curcio et al. 2006). This disorder can be solved through the adoption of healthy sleep timetables (Curcio et al. 2006) (modifying the sleeping hours). Sleep deprivation can also affect both to the attention, divergent thinking and memorization capabilities (WIMMER et al. 1992; Harrison et al. 2000). Therefore, the analysis of students' sleep can be used as a key factor towards the improvement of learning support systems. For example, the level of fatigue, the pattern of rest or the chronotype can be used to support alert services and recommendation facilities (e.g. to propose a study plan). Second, despite existing off the shelf wearables provide some sleep indicators, such as the sleep duration, the number of awakes or the time to fall asleep, the specialized scientific sleep literature uses other indicators that feature sleep and sleep behaviour in a more precise way (Sano et al. 2015). Particularly, in the fields of medicine and psychology some broadly accepted sleep indicators can be found (Carskadon et al. 1986; Buysse et al. 1989; Quan et al. 1999): sleep quality, sleepiness level and chronotype. It would be good if these top indicators could be estimated from off the shelf wearables, using the provided sleep indicators and other physiological signals that can be collected from them. For example, Heart Rate (Bunde et al. 2000), respiration (Snyder et al. 1964) or temperature (Kräuchi 2002) are very well correlated to sleep and can be collected by off the shelf wearables. And finally, third, we hope in the new future our higher education students will carry on affordable off the shelf wearables including sensors that provide the needed data to calculate the proposed sleep indicators.

This paper analyses how a sleep indicators can be calculated from data collected in commercial electronics wrist wearables (Kikhia et al. 2015) and provides initial experimental results for their validation. We do not pretend to replace medical procedures, but to check if similar indicators can be offered using off the shelf wearables. This will enable to extend the student profile with sleep information. The study has been carried out using three different wearables: Fitbit, Microsoft Band and Jawbone. These are affordable examples of new wrist devices that are being increasingly used by people for fitness and wellness purposes. Indeed, they already provide some sleep related data, such as number of awakes, time to fall asleep, sleep duration and a kind of "sleep efficiency" indicator. Nevertheless, this data has a low interpretative value and it cannot be used to come to the conclusion about how well a person sleeps.

Typically, two persons sleeping the same number of hours can feel very different (Sano et al. 2015). To provide results with more interpretative value, our search for new indicators is focused on clinic sleep analysis procedures.

The rest of the paper is organized as follows. The following section reviews broadly accepted sleep indicators available in the scientific literature, focused mainly in the data used to calculate them. Next, section 3 reviews the existing commercial off the shelf wrist wearables, focusing on the sleep indicators already provided and on the data that can be collected from their sensors and be related to sleep, particularly data such as the used to calculate the scientific sleep indicators. After these two reviews and taking into account the findings, section 4 introduces proposals of sleep indicators that can be calculated from data collected in commercial electronics wrist wearables in an autonomous and transparent way. Next, an initial validation of these indicators is described in section 5. Section 6 proposes some educational scenarios in which the proposed solutions can be applied. Finally, the conclusions of the piece of research are discussed in section 7 .

## 2. Sleep indicators in the scientific literature

This review has a twofold purpose. First, to identify the more accepted sleep indicators adopted in the medical scientific literature to describe features and behaviours about how a person sleeps. Second, to identify the variables used to calculate such indicators. In the following section we check if these variables can be obtained from commercial off the shelf electronic wrist wearables.

In the scientific literature, there exist various proposals for sleep indicators and there also exist different procedures to calculate them, no one of them involves the use of commercial off the shelf devices. Broadly speaking, there are two types of techniques to obtain indicators: i) subjective tests, in which the individual under study answers a series of questions, a questionnaire, based on their perceptions and memories; and ii) objective tests: such as clinical trials where an expert performs a series of analytical experiments to the patient in a specialized centre. The main advantage of the subjective tests is they do not require many resources, particularly no expert is needed. Nevertheless, its main drawback is the lack of rigor in the answers.

Taking into account its internal consistency, widespread recognition and use in clinical studies the most relevant sleep indicators in the scientific literature are the following ones:

- Sleep Quality. Sleep is an essential vital function (the human being cannot live without sleep), restorative (sleep repairs the body every day) and fundamental to ensure wakefulness (sleeping to feel awake the next day). The sleep quality index represents, in most cases, the level of rest and repair that has occurred in the individual. Perhaps, the most outstanding method to estimate this indicator is the Pittsburgh Sleep Quality Index (PSQI) test (Buysse et al. 1989). This test is a subjective questionnaire that assesses sleep quality and disturbances over a 1-month time interval. It consists of 19 questions divided in 7 fields. It has a Cronbach alpha coefficient of 0.83 , a sensitivity of $89.6 \%$ and a specificity of $86.5 \%$. It makes a very good measurer and currently, this indicator is very used in the medical scientific literature (Medeiros et al. 2003; Alt et al. 2013). The variables used in this procedure are reflected in the table 1.

Table 1. Variables in PSQI

| Variables |  |  |  |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- | :---: | :---: | :---: |
| 1. | Bedtime | 7. | Bad breath | 13. | User perceived sleep quality |  |  |  |
| 2. | Fall as sleep | 8. | Snoring | 14. | Drug ingestion |  |  |  |
| 3. | Rise time | 9. | Feel cold | 15. | Somnolence developing a activity |  |  |  |
| 4. | Real sleep duration | 10. | Feel hot | 16. | Mood and energy developing a activity |  |  |  |
| 5. | Awakes | 11. | Nightmares | 17. | Number of co-sleepers |  |  |  |
| 6. | Awakes to the toilet | 12. | Pains in the night |  |  |  |  |  |

- Sleepiness or Somnolence. Sleepiness is a state of relatively strong desire for sleep. This state is usually accompanied by lethargy, weakness and, what is more meaningful for our work, lack of mental agility. Different methods have been proposed in the literature to estimate its value or level, both subjective and objective. Some of the most outstanding proposals are the following ones:
- Sleepiness scale of Epworth (Johns and others 1991). It is a subjective test used to measure average daytime sleepiness. Currently, this is also very used in the medical scientific literature (Kritikou et al. 2013; Duarte et al. 2014). It consists of 8 questions and it has a high Cronbach alpha coefficient in the range of 0.73 to 0.88 . For each one of the questions it is possible to identify a variable, cf. Table 2.
- Multiple Sleep Latency Test (MSLT), Maintenance of Wakefulness Test (MWT) and OSLER test (Bernal et al. 2012) are objective tests. In the MSLT, the patient is observed in a room while performing a number of micro-sleeps. The MWT and OSLER tests study the ability of the person to stay awake at a low stimulation environment. All these methods share the use of variables such as the start and end of the sleep period, and REM periods, cf. Table 2.

Table 2. Variables in Sleepiness test.

| Epworth |  | MSLT/TMV/OSLER |  |
| :--- | :--- | :--- | :--- |
| 1. | Fatigue sitting or reading | 9. | Start test |
| 2. | Fatigue watching TV | 10. | Sleep start |
| 3. | Fatigue in a public place | 11. | REM start |
| 4. | Fatigue as a passenger of a car in silence | 12. | End test |
| 5. | Fatigue in a comfortable place |  |  |
| 6. | Fatigue sitting, talking to another person |  |  |
| 7. | Fatigue sitting, after eating |  |  |
| 8. | Fatigue sitting in a car waiting for the traffic |  |  |

- Chronotype. The chronotype of a person identifies his/her propensity to sleep at a particular time during a 24-hour period. Eveningness and morningness are the two extremes, with most individuals having some flexibility in the timing of their sleep period. This indicator is also very used in recent works (Medeiros et al. 2003; Lucassen et al. 2013). It can be calculated using the Horne and Östberg questionnaire (Horne and Ostberg 1975). This test is a subjective questionnaire to determine morningness-eveningness in human circadian rhythms. It consists of 19 questions to estimate the time chosen by the user to sleep if she would enjoy complete freedom to choose it. In addition, the test also inquiries about the periods in which the subject believes her performance and welfare have improved. Each one of the 19 questions involve a particular variable, as shown in table 3.

Table 3. Variables in Chronotype.

| Variables |  |  |  |
| :--- | :--- | :--- | :--- |
| 1. | Rise time without obligations | 11. | Time with maximum performance |
| 2. | Bedtime without obligations | 12. | Level of fatigue at 11 pm. |
| 3. | Need alarm | 13. | Sleeping in an abnormal time, when you awake |
| 4. | Easy get up | 14. | Having guard between 4 and 6 am when sleep |
| 5. | Alert level after awaking | 15. | Time with high performance for physical work |
| 6. | Appetite level after awaking | 16. | Physical exercise between 10 and 11 pm |
| 7. | Fatigue level after awaking | 17. | Five hours with maximum performance |
| 8. | Bedtime without liabilities | 18. | Time of day with maximum comfort. |
| 9. | Physical exercise between 7 and 8 am | 19. | What perception has its own chronotype |
| 10. | Time when subject feel tired. |  |  |

In addition to the previous ones there exist other type of sleep indicators. It is especially relevant the sleep regularity (Soehner et al. 2011; Sano and Eng 2016) because it plays a main role for the detection of sleep patterns. The formula (1) used in (Sano and Eng 2016) is as follows:

$$
\text { Sleep regularity }=\frac{1+\frac{1}{T-\tau} \int_{0}^{T-\tau} s(t) s(t+\tau) d t}{2}
$$

$s(t)=1$ during wake and $s(t)=-1$ during sleep, and T. Suppose data are collected for $y=[0, T]$. Choose $\tau=24$.

Basically, the sleep regularity is calculated taking into account the similarity between sleep periods in consecutive days. A review of methods and factors used to measure the sleep regularity can be found at (Bei et al. 2016). Others indicators are directly focused on specific sickness or sleep-disorders. For example: the melatonin secretion test to calculate an insomnia level indicator; the measurement of hypocretin in LCR to calculate a narcolepsy indicator; indicators related to the "restless legs syndrome", etc.

Previous indicators are not based on any technology, but there are some cases in which specific devices and medical instrumentation has been used to measure sleep related features and disorders, such as the ones involved in polysomnography (Manber et al. 1998) and actigraphy (Ancoli-Israel 2005) techniques. The polysomnography involves the use of an Electroencephalography (EEG), using electrodes and a specific instrument to collect brain signals, an Electrooculography (EOG), to detect ocular movements, an Electromyography (EMG), to detect muscle movements, and other variables that are added depending on the user's sleep complaints (e.g., respiration, heart rate, oximetry). This technique is very complex because the different kind of devices needed. By the contrary, the actigraphy technique involves the use of a single device: the actigraph. It is not as precise as the polysomnography, but the results are acceptable for regular studies not focused on the study of specific sicknesses (Jean-Louis et al. 2001). This technique based on accelerometer sensors can detect several sleep indicators, such as wake versus sleep time, sleep latency, awakenings during the night (Ancoli-Israel 2005), REM phase (Herscovici et al. 2007), etc. In some cases, actigraphy has also been used to support sleep quality and sleepiness measures (Keller et al. 2008), and for other non-sleep purposes, such as comfort measurement (Telfer et al. 2009).Nevertheless, despite actigraphs are much more comfortable than electrodes and other devices for clinic tests they are not available for the general consumer. For this reason, off the shelf wrist wearables provide new chances to collect data related to sleep, because they allow to apply techniques similar to actigraphy but at a lower price.

In addition to the previous methods, there exist other physiological signals related to sleep. For example: the Heart Rate (HR), the Respiration rate or the Temperature. Deep sleep, light sleep, and rapid eye movement (REM) are well-known sleep stages that maintain a good correlation to HR (Bunde et al. 2000), as it has been demonstrated both in human and animal experiments (Baust and Bohnert 1969). Indeed, to study some heart conditions, doctors take into account the relation between the HR and sleep, being considered as a possible explanation of sudden death (Vanoli et al. 1995). Similarly, the respiration rate, which is well correlated to HR, also change during the different sleep stages and can be used to estimate them (Snyder et al. 1964). Moreover, changes in the skin temperature can disturb the sleeping time, both when it starts and to maintain it (Raymann et al. 2007).All these signals vary in accordance to a circadian rhythm (Kräuchi 2002), also during night-time.

## 3. Data and sleep-related features in commercial off the shelf wearables

This section reviews data and indicators provided by existing commercial off the shelf wrist wearables, more specifically sleep-related data and indicators. From all the wearables available, the wrist wearables (smartband and smartwatches) have the largest market share and the previsions for the near future are similar (IDC 2016a): more tan $80 \%$ of market share by 2020 and around 170 millions of shipments. Currently, these devices include a great variety of sensors that can be used to collect physiological and environmental data.

We performed a review of more than 145 wearables searched in different sources: review papers (Swan 2012; ur Rehman et al. 2015) and well known websites like (Inc 2016), (Apple 2016), (SAMSUNG 2016), (Jawbone 2016), (Xiaomi 2016), (LG 2016). From this review, 17 sensors were identified and classified in two categories, as it is shown in Table 4: environmental and biometric. Fig. 1 show the percentage of main sensors availability in current wearables. The accelerometer is the most popular one, it is available in $85 \%$ of wearables. The accelerometer sensor is included in the wearables of main vendors representing about the $59 \%$ of the market share (IDC 2016b): Apple, Xiaomi, Fitbit, Garmin and Samsung. Next sensor in availability rate is the Heart Rate sensor (HR), that can be found in around $33 \%$ of the wearables in the market. Nevertheless, this sensor is just available at mid-range and high-end devices. The GPS, gyroscope and compass are used in conjunction with the accelerometer to measure distances, but they do not offer any other type of quantification data by themselves. Finally, the microphone is available in less than $18 \%$ of the devices, and in most cases, it can only be used to communicate to a smartphone, but not as a recorder directly. The rest of sensors can be found, but just on very specific devices. For example, the skin temperature sensor can be found in only $5 \%$ of the 145 reviewed wearables. In any case we foresee in the following years most of these sensors (e.g. skin temperature sensor, GSR and oximeters) will be more common taking into account the recent evolution of wearables (IDC 2016a), mainly in fitness environments.

Despite the previous results, it is important to notice that the availability of sensors in wearables does not necessarily mean that the corresponding data can be collected from them, because quite frequently the wearable vendors constrain or limit the access to raw data. For example, in some of the Fitbit devices accelerometer and altimeter raw data is not available for developers, just for the vendor to calculate indicators of the start and end of the sleep periods. In (de Arriba Pérez et al. 2016a) we have identified other issues to collect data from off the shelf wearables: battery duration, differences on data models, duration of tokens to authorize the collection of data, etc.

Table 4. Sensors in wearables.

| Environmental sensors | Biometric sensors |
| :--- | :--- |
| GPS | Heart Rate sensor (HR) |
| Accelerometer / pedometer | Blood Pressure sensor |
| Gyroscope | Respiration rate sensor |
| Compass (magnetic sensor) | Oximeter |
| Light sensor | Galvanic Skin Response (GSR) |
| Altimeter | Skin Temperature sensor (ST) |
| Pressure sensor / barometer |  |
| Humidity sensor |  |
| UV sensor |  |
| Microphone |  |
| Camera |  |



Fig. 1 Percentage of sensors availability in wrist wearables

In addition to raw sensor data, wearable vendors also provide more elaborated information. Table 5 shows some common sleep features offered by well-known wearable vendors, such as Fitbit, Microsoft, Jawbone, etc. Of course, these companies also provide fitness-related indicators: calories, distance, steps, velocity, number of stairs, type of exercise (running, cycling) duration of a high exercise, etc.

Table 5. The most common sleep information available through rest APIs of manufacturers using wrist wearable devices.

| Sleep data from wearable manufacturers |  |
| :--- | :--- |
| Bedtime | Number of awakes |
| Sleep start | Awake duration |
| Sleep end | Phases of sleep (asleep, restless, awake) |
| Fall asleep | Sleep Efficiency |

Focusing on sleep, there are some indicators provided by vendors, currently. The more common ones are obtained from accelerometer raw data values using actigraphy techniques (Ancoli-Israel 2005) (Johnson et al. 2003). The wearable raw accelerometer data is collected and transferred to a smartphone that process it to estimate the time slept by a certain user (Fitbit 2016a; Microsoft 2016a; Xiaomi 2016). By the way, the use of actigraphy techniques and accelerometer data in smartphones is not new, apps such as "Sleep as Android" (Sleep as Android 2016) already provided this functionality using the smartphone accelerometer data directly. Instead, wearables are directly wear by the person and as a result they may register user movements in a better way. Main wearable vendors also use actigraphy techniques and algorithms to provide a sleep efficiency indicator. Fitbit explains in its public web site that this value is calculated using the following formula: 100 * time asleep / (time asleep + time restless + time awoken during sleep) (Fitbit 2016b). Microsoft follows a similar approach (Microsoft 2016b) as the ratio between the sleeping time and the total time. From our experiments, we consider these indicators are not valid, because they provide values over an $85 \%$ of sleep efficiency in almost all cases. Jawbone is another wearable vendor who also provides a sleep indicator that performs better than the two previous ones. The method used to calculate this indicator is not publicly available, but we think it uses the sleep duration, the number of times the user has awaken and the ratio between time in bed and real slept time. No vendor provides any type of indicator similar to sleepiness, chronotype or sleep regularity.

## 4. Proposed sleep indicators and procedures

In previous sections it has been shown how several procedures to calculate sleep indicators have been proposed, including the clinical and actigraphy techniques and other preliminary studies involving the use of video cameras (Liao and Kuo 2013). The clinic methods or video camera systems are uncomfortable for users, require a costly infrastructure and violate user privacy. We look for a comfortable and automatic solution, requiring no interaction from the final user, such as the one described in (Chent et al. 2013) for smartphones, but in our case using commercial off the shelf electronic wearables.

The more common indicators described in the scientific literature to characterize sleep are: sleep quality, sleepiness and chronotype. These indicators are calculated from data collected in interviews or through direct observations. We wonder if such data, or a high part of it, could be collected from commercial off the shelf wearables in an automatic way. To answer this question, firstly we composed table 6 gathering the variables used to calculate the Sleep Quality (SQ), Sleepiness Level (SL) and Chronotype (CT) indicators, as described in the scientific literature. This table includes the variables from tables 1, 2 and 3, that could be collected from wearables. We have discarded purely perceptual variables, such as "Perceived sleep quality user", "Appetite level", because they are subjective and cannot be measured by a wearable. Furthermore, the variables that correspond to the same measurements but in different contexts have been
merged. For example: "Fatigue level after awaking" or "fatigue watching TV", "fatigue sitting, after eating" and similar have been grouped into the fatigue variable, because wearables cannot be used to collect the user context. As it can be observed, most variables are not included in the three indicators, just bedtime and fatigue-concentration. In addition, as it is shown in table 7, fatigue-concentration needs a great number of sensors to be identified correctly, and such sensors are not common in wearables.

Secondly, we composed table 7 showing the correspondences between the variables and the wearable sensors and processed data. As conclusion, we can say that accelerometer and Heart Rate (HR) are key sensors because they are available in a high percentage of devices and involved in many of the sleep-related factors.

Next sub-sections introduce our procedures to calculate the proposed sleep indicators using the accelerometer and HR sensors as main data sources. In addition, the Skin Temperature (ST) and the Galvanic Skin Response (GSR) sensors are going to be considered as optional parameters, because despite both of them are rare in current wearables they provide very good correlations to sleep features, according to the literature (Sano and Eng 2016).

Table 6. Variables in tests used to calculate scientific literature sleep indicators.

| Variables | SQ | SL | CT | Variables | SQ | SL | CT |
| :--- | :---: | :---: | :---: | :--- | :---: | :---: | :---: |
| Bedtime | X | X | X | Ambient temperature | X |  |  |
| Rise time | X |  | X | Pain level | X |  |  |
| Fall as sleep | X |  |  | Nightmares detection | X |  |  |
| Awakes | X |  |  | Drug ingestion | X |  |  |
| Phases of sleep | X | X |  | Alert level |  |  |  |
| Breathing per minute | X |  |  | Physical exercise- |  |  | X |
|  |  |  |  | Activity |  |  |  |
| Blood Oxygenation | X |  |  | Fatigue-concentration | X | X | X |
| Snoring level | X |  |  | Comfort level |  |  | X |
| Noise level | X |  |  |  |  |  |  |
| Body temperature | X |  |  |  |  |  |  |

Table 7. Sensors/data used to calculate the variables of sleep.

| Variables | Raw sensor data | Processed sensor data |
| :--- | :--- | :--- |
| Bedtime/Rise time/Fall as | Accelerometer, HR | Bedtime, Sleep start, Sleep end, |
| sleep/Awakes/Phases of sleep/Hr |  | Fall asleep, Number of awakes, |
| minimum |  | Awake duration, Real sleep: |
| (Bedtime) - (Fall asleep + Awake |  |  |
|  |  | duration), Phases of sleep, Sleep |
|  |  | Efficiency, HR minimum |
| Breathing per minute/Blood | Respiration rate, SpO2 | Respiration levels |
| Oxygenation |  | Sounds |
| Noise level/Snoring level | Microphone | Temperature, humidity |
| Body temperature/Ambient humidity | Skin temperature, |  |
|  | Humidity |  |


| Physical exercise and Activity | Accelerometer, HR, GPS | Distance, calories, steps |
| :--- | :--- | :--- |
| Fatigue and concentration | Accelerometer, GPS, | Mood, alert level |
| level/Comfort level/Alert level | GSR, Respiration rate, |  |
|  | SpO2, HR |  |

### 4.1. Sleep Quality

The Sleep Quality (SQ) indicator is based on the PSQI to provide a measure of the level of rest and repair of sleep. Taking into account the information introduced in the previous sections it can be seen that some of the data used to calculate the original PSQI can be collected from wearables. Particularly, the following factors are used: sleep duration, fall as sleep, awake, HR min and average ST.

As a first approach we developed a direct calculation of this indicator taking into account the data that could be collected or calculated from wearables and establishing some fixed weights (de Arriba Pérez et al. 2016b). This first version provided good results, better than the vendors' ones. Nevertheless, we noticed the perceived sleep quality by users experiencing the same data (e.g. slept hours, number of awakes) embodied significant differences. For example, two persons that sleep the same number of hours can feel very different about how it was. To tackle with this issue, we decided to use an adaptive predictor solution using Machine Learning techniques and the subjective perception of the user. As a result, the solution adapts to each person. The predictive algorithm is trained during an initial stage, using data collected from the wearable device and the answer of the user to a query about his/her sleep quality feeling. Based on our test a minimum of five daily samples is needed to provide a first predicted value, which improves when more daily samples are included. After the initial training, a SQ indicator with a value between 0 and 100 is produced on a daily basis. This is a main difference in relation to the classic PSQI, because it only provides a value referred to a onemonth period.

During the training, the query about the sleep feeling to the user is performed always at the same time after noon, at 12:30. This time was chosen deliberately for two reasons. First, to not disturb to the user if he/she wakes up very late. Second, because if the query is performed closer to the wake-up time other factors different to the rest feeling could influence in the user answer. To perform the question, we use an Android app, cf. Fig. 2. The app shows some data about the sleep period: init. time, finish time, total sleep time, HR at rest and number of awakes. Then, the user has to choose his/her own rest feeling using a Likert scale (with 5-points): 0 -very tired face; 25 -tired face; 50 -average face; 75-happy face; 100 -very happy face).


Fig. 2 Quiz app to query the user about his/her subjective sleepiness quality

We tried different algorithms to predict the SQ indicator using the Weka library (Mark et al. 2011) for Java. No one of the more common nominal predictors, such as the C 4.5 , provided good results, because they don't take into account the order of the values. These were the initial algorithms tested to perform the estimations:

- C 4.5. This classifier generates a pruned or unpruned decision tree from a training data set. At each tree node it is established a classification decision until the top of the branch, where the result (predicted value) is located (Quinlan and Ross 1993).
- Bagging. It's used as a bagging a classifier to reduce variance. (Breiman 1996).
- Adaboost. It's used for boosting a nominal class classifier, often dramatically improves performance, but sometimes overfits (Freund et al. 1996).
- Knn. K-nearest neighbours classifier. This algorithm predicts an outcome based on the proximity/similarity of the attributes used to train it. It can select appropriate value of K based on cross-validation and it can also use distance weighting (Aha et al. 1991).
- Ramdom Forest. It's used for constructing a forest of random trees. As in the decision tree, each of its branches contains a condition to evaluate. Depending on the evaluation of the condition, the next branch is taken or not. This is repeated until the top of the branch. The main difference is that the tree is made up by a combination of prediction trees, depending each tree on a random vector (Breiman 2001).
- Naive Bayes. This is a probabilistic classifier based on simplifications of the Bayes Theorem, assuming that the predictor variables are independent. In this way, each variable offers a certain success percentage, independently of the other ones (John and Langley 1995).
- OneR. It uses the minimum-error attribute for prediction, discretizing numeric attributes. It selects the attribute that best explains the output variable and uses this attribute as a node in a decision tree (Holte 1993).
- ZeroR. It is used as a baseline. If the remaining classifiers have lower or similar values, it indicates that the sample has not been chosen properly or the classifiers fail to construct a predictive model. It always predicts the most frequent value of the sample (Weka).

Table 8 shows the accuracy of the classifier model obtained with these algorithms using a data set of 374 samples. As it can be observed, no one offers an accuracy over $50 \%$, the algorithm fails more than half of the decisions taken. Then, we searched for algorithms to predict a numeric variable taking into account the order and the proximity among values. To perform this experiment, the following algorithms were chosen:

- K*. This is an instance-based classifier. The class of a test instance, based on the class of those training instances similar to it, is determined with some similarity function. Its main difference is that it uses an entropy-based distance function (Cleary et al. 1995).
- Linear Regression. To generate the prediction model it performs an approximation by a straight line of the attributes and values to predict (Weka).
- SMOreg. It implements the support vector machine for regression (Shevade et al. 2000).
- Gaussian Processes classifier. This classifier implements statistic processes for regression (MacKay 1998).
- Multilayer perceptron. It is an artificial neural network model that uses back propagation to classify instances (Ruck et al. 1990).
- Locally weighted learning (LWL). In this classifier at each point of interest a local model is created based on neighbouring data (Frank et al. 2003).
Table 9 shows the standardized (values between 0 and 1) mean absolute and standard deviation errors for different algorithms. The $\mathrm{K}^{*}$ provides the better statistical results, with the lower mean absolute error, approximately better in a $5 \%$ than other algorithms.

Table 8. Accuracy classifiers using nominal variable value.

| C4.5 | Bagging(C4.5) | AdaBoost <br> (C4.5) | Knn | Random <br> Forest | Naive <br> Bayes | OneR | Zero R |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
|  |  | 45.99 | 34.70 | 49.76 | 43.66 | 43.49 | 36.36 |
| 46.24 | 47.78 |  |  |  |  |  |  |

Table 9. Mean absolute error + (standard deviation error) classifiers using numeric variable value.

| K* | Linear <br> Regression | SMOreg | Gaussian <br> Processes | Multilayer <br> Perceptron | Knn | LWL |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 0.155 | 0.19 | 0.2025 | 0.195 | 0.2025 | 0.2125 | 0.1825 |
| $(0.02)$ | $(0.023)$ | $(0.023)$ | $(0.02)$ | $(0.043)$ | $(0.028)$ | $(0.023)$ |

As a result, the final classifier works as follows. User sleep data and the subjective answer to the question about his/her sleep quality is collected daily. With a minimum of five days' data the $\mathrm{K}^{*}$ algorithm can be used to predict a new SQ.

### 4.2. Sleepiness Level

The Sleepiness Level (SL) indicator is proposed to identify relaxing states close to sleep during situations in which the user should be active. The idea of this indicator is to detect in real time if the user is sleepy from data related to the
active/inactive state of the user. For example, to detect somnolence in learners during a classroom or study session. Therefore, there exists a temporal requirement related to the availability of data.

The data used proposed to calculate this indicator is the following one:

- Accelerometer: to analyse the movements of the person.
- Heart Rate: it provides a signature of the user during rest situations.
- Skin Temperature: it provides an estimation of the user comfort level.

To estimate the SL indicator, a machine learning method is applied to a large number of tagged tracking records was used. The classifier needs to be trained on periods in which the user is both sleeping and active. Therefore, two different kinds of measures were used. The first one involves data from the HR, temperature and accelerometer sensors collected while the person is deeply sleeping at night. The second one involves the collection of data from the same sensors at moments in which she/he feels really active. In this case an app is used to enquire and get the state of the user, cf. Fig. 3. These measures represent the tagged tracking records used to train the algorithm. After an initial training period collecting this data, the algorithm is tuned by querying the person about her/his agreement with the estimated indicator.


Fig. 3 Sleepiness detection application

We decided to use the C 4.5 classifier because it is fast and simple. In any case, we contrasted it to other classifiers, some of them heavier in computational cost and other ones lighter. Results are shown in table 10. The classifiers used in this section are the same ones used in the first part of the experiment, described in section 4.1. The results show that the C4.5, Adaboost over C4.5 and Random Forest are the best options. All of them provide similar results and the differences are not statistically significant. As it can be observed, an average absolute error of just $4 \%$ and a precision over $97 \%$ is obtained. Therefore, taking into account the speed and simplicity of the different methods, C4.5 was
selected as the option to estimate in real time the sleepiness state. These results are interpreted as a very good approximation towards its implementation in a real system beyond this prototype.

Table 10 Comparing different classifiers.

|  | C4.5 | Adaboost (C4.5) | K-nn | Random Forest | Naive Bayes | OneR | Zero R |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| Mean absolute | 0.04 | 0.02 | 0.03 | 0.04 | 0.17 | 0.09 | 0.46 |
| error |  |  |  |  |  |  |  |
| Root Mean <br> squared error | 0.15 | 0.14 | 0.17 | 0.14 | 0.34 | 0.30 | 0.48 |
| Accuracy | 97.26 | 97.84 | 97.12 | 97.51 | 83.55 | 90.98 | 63.97 |

### 4.3. Basic Chronotype

This indicator seeks to identify the student chronotype by observing the hours slept/raised and taking into account differences between instructional days and holidays-weekends. The working routines or the academic activities, in case of students, usually stand in the way of natural sleeping periods. In addition, during weekends people can vary the routines, making up for the sleep hours missed during the working week, or by the contrary, missing more sleeping hours because social activities.

To calculate this indicator, we used an equivalence table relating the start and end sleeping times to a value in the morningness-eveningness scale (Lama et al. 2008), cf. Table 11. The time of the start and end of sleep are taken from the wearable data. Different values in the range -4 ( -2 start bedtime plus -2 end bedtime) for "Definitely evening" and 4 ( 2 start bedtime and 2 end bedtime) for "Extreme morning" are assigned to each person depending on the hour at which she/he starts and ends sleep. We take into account data from all the days, without distinguishing between working and non-working days.

Table 11. Score and rank equivalent to chronotype.

| Type | Start Bed Time | Values | End Bed Time | Values |
| :---: | :---: | :---: | :---: | :---: |
| Extreme morning | $-/ 21: 30$ | 2 | $-/ 05: 00$ | 2 |
| Moderately morning | $21: 30 / 22: 45$ | 1 | $05: 00 / 06: 30$ | 1 |
| Intermediate | $22: 45 / 00: 45$ | 0 | $06: 30 / 08: 30$ | 0 |
| Moderately evening | $00: 45 / 02: 00$ | -1 | $08: 30 / 10: 00$ | -1 |
| Definitely evening | $02: 00 /-$ | -2 | $10: 00 /-$ | -2 |

### 4.4. Regularity indicators

In addition to the following indicators and taking into account the history of values corresponding to the previous indicators new derived indexes are proposed based on the regularity:

Sleep Regularity (SR). This indicator is focused on measuring how regular is the sleep of a person in consecutive days. Generally, the development of human activities in a regular basis, in concordance to the regularity of physiological rhythms, has positive effects in the health and performance of people. Indeed, the sleep regularity is being studied in
recent pieces of research and the results show a positive relation between the sleep regularity and the academic performance (Bei et al. 2016).

There exist different ways in which SR can be calculated:

- In section 2 of this paper is described one method focused on sleep periods (Sano and Eng 2016), that we name as Time-based SR (TSR). This first method provides a direct measure of the regularity, but it does not take into account the sleep quality or the chronotype. Furthermore, it can be affected by differences among week days, depending on the schedules of the educational and working activities, and weekend days, where sleep periods are usually changed to take advantage of the free time and therefore usually people wakes up later.
- We propose a different SR indicator and method based on the sleep quality, named as Quality-based SR (QSR). This method focuses on the SQ indicator because embodies many features related to the sleep, such as the total sleep time, the level of rest during sleep, etc. Therefore, this indicator is calculated from the regularity of the SQ in accordance to the following algorithm (2):
- First, SQ measures for a specified period are collected and included in a vector. Each position of the vector contains the SQ measure of a certain day $(\overrightarrow{S Q})$.
- Second, the vector is normalized taking into account the maximum value and cero. As a result, each position contains a value between 0 and 1.1 represents the best $S Q$ obtained in the period ( $\overrightarrow{S Q}_{\text {norm }}$ ).
- Third, a medium SQ vector for the same period is generated, containing all positions the mean value of SQ normalized vector. ( $\overrightarrow{M e a n S Q}_{\text {norm }}$ )
- Fourth, the Euclidean distance $\left(d_{E}\right)$ among the two vectors is calculated and normalized, where $N$ is the size of the vector $\overrightarrow{S Q_{\text {norm }}}$.

$$
Q S R=\frac{d_{E}\left(\overrightarrow{S Q_{\text {norm }}}, \overrightarrow{M e a n S Q}\right)}{N} * 100
$$

Sleep-GPA Sensitivity (CS). Another indicator of interest in our piece of research is the influence of the sleep patterns by external agents. The difficulties to calculate this factor in a precise way are high, but we propose a rough estimation that can be useful for our purposes. The goal is to check if a good or bad performance in assessments is related to high or low SQ values. The relation between these values provides an estimation of the sensitivity or dependency between these two variables. To calculate it, we take into account the days before and after the assessment activity, obtaining a SQ mean value that is related to the assessment result. Finally, we calculate the correlation between these variables to get a CS estimation. The algorithm is as follows (3):

- The following variables are used: academic activities (for example an exam) Grade Point Average (GPA) and sleep quality.
- A vector with the GPA of the activities is created $\left(\overrightarrow{\text { Activitles }}_{\text {grades }}\right)$.
- A vector with the mean SQ values in the days before and after the activities is created $\left(\overrightarrow{S Q}_{\text {avg }}\right)$. We check the sleep regularity in the 2 days before and 2 days after the activity.
- The correlation between these two vectors is calculated.
- Finally, the absolute value of the result is calculated and multiplied by 100 to get the estimated value as a percentage.

$$
\begin{gather*}
S Q_{\text {avg }}=\frac{\sum_{t}^{t+4} S Q}{5} \\
C S_{\text {activity }}=A b s\left(\operatorname{corr}\left(\overrightarrow{\text { Actıvitıes }}_{\text {grades }}, \overrightarrow{S Q}_{\text {avg }}\right)\right) * 100 \tag{3}
\end{gather*}
$$

where $t$ is the time point two days previous to the activity.

## 5. Experimental Results

In order to verify the feasibility of the indicators proposed in the previous section a prototype software system involving different wearables and sensors was developed. The software uses the data collected from the wearables and performs the algorithm implementations. The system involves both the use of real time sensors data, required to estimate the sleepiness index, and stored data, needed to calculate the other indexes.

Three different off the shelf wearables have been used: Fitbit, Microsoft Band and Jawbone. Fitbit and Jawbone offer access to sleep data through public APIs for developers. Microsoft Band in addition to an API for developers, also provides a SDK to collect data directly from wearables in real time.

Our system involves several components to perform the proposed calculations (de Arriba Pérez et al. 2016a). First, some components are used to collect and homogenize the data from different wearables. Second, the data collected is stored in a database to get a quick access to: daily, weekly and monthly data. A task manager is included to perform data synchronization and calculations. These modules have been developed in Java and are executed in a Tomcat server. The Jersey library (Jersey 2016) is used to provide the access API. Finally, Machine Learning calculations are based on the Weka library (Mark et al. 2011) for Java.

The two ways in which data can be collected from wearables are:

- From the sensors available in the devices. This option provides to the developer full control to apply our algorithms to the raw data collected. In the cases where it is possible, an app was developed to be installed in the student smartphone to collect this data and to transfer it to our analysis server. This was done in the case of the Microsoft Band. To do it, the app uses a specific SDK that provides a connection between the collecting device (the wearable) and the device with the app (the smartphone). This way of collecting data is needed in case real-time data availability is a requirement.
- Using the API available at the vendor warehouse. Most of the wearable vendors provide a smartphone app to collect data from the wearable sensors and to transfer such data to their own warehouses.
Manufacturers also provide an API to enable users and third-party developers to get access to the data maintained in their warehouses. We have used this option of access to collect data from Fitbit and Jawbone wearables, because in this case the first option is not feasible. In this case, real-time data cannot be collected.

Independently of the access mode the data collected from the sensors needs to be processed before the application of the algorithms. The initial measurements of sensors are voltages or light intensities that are transformed into the expected measures (e.g. HR, temperature). In the case of the HR, off the shelf wearables usually implement a complex process known as Photoplethysmography (PPG) (Stahl et al. 2016), that is an alternative to the Electrocardiography (ECG). The ECG measures the potential generated by electrical signals that control the expansion and contraction of heart, while

PPG uses a light-based technology, measuring differences in blood volume through a reflected signal (infra-red or red) over the skin. This principle can be used to measure the HR, but not Heart Rate variability, because it does not provide enough accuracy (Healey and Jennifer Anne 2000). The intensity of the reflected light is drawn along each heartbeat and then it is approximated with a line that identifies the maximum and minimum light intensity points. These points are used to provide the final HR (Allen 2007; Nitzan et al. 2014)

In addition to the previous components, our prototype also includes a Dashboard to enable the researcher to check the collected data, to arrange the results, to visualize specific analysis and to facilitate the performance of certain tests (cf. Fig. 4). In the top side of the figure it is shown a summary of the current night sleep period. The graphic shows inactivity periods and awake instants. In the right side, it is included a summary of the sleep data for the current night. The central part of the figure includes three sector diagrams showing the sleep efficiency estimated by our system, estimated by the vendor and the amount of awake and sleeping (light and relaxing) time. The bottom part of the figure shows a histogram with historical data from previous days and a statistical summary.

Data of franfuco4444@gmail.com device: Microsot_Api for date 2016-03-16 $\ddagger$


Summary of the last days 30 :


Regularity: $67 \%$ Efficiency max-mean-min: $94 \%-84 \%$ $.23 \%$
Fall as sleep max-mean-min: 20 min -
6.6 min - 3 min

Duration max-mean-min: 9.28 hours -
6.92hours - 2.93 hours

Real duration max-mean-min:
8.12 hours -6.17 hours -1.88 hours Number of awake avg: 4.9 HR max-mean-min: 60-56-53

Fig. 4 Dashboard for researchers

Using the prototype, some experiments were performed oriented towards the evaluation of the accuracy of the proposed indicators, mainly of the sleep quality, chronotype and the sleepiness level (for regularity indicators we need more time
to obtain a relevant result with a final experiment). Five subjects participated in this evaluation, all of them were informed and accepted the use of the data collected for this study. The study was performed in the usual environment of the participants, without requiring any kind of special condition or instrumentation, just the wearables: Fitbit, Microsoft Band and Jawbone. All the 5 participants were involved in the SQ and chronotype experiments, while just the subject with the Microsoft Band were involved in the SL (sleepiness) experiment because this device is the only one that can be used to collect real time data.

Next, we describe some key features of the subjects involved in the experimentation:

- Sleep features. The 5 participants are in the $27-44$ age range with a median value $38 \pm 6,43$ age. They were administered the Spanish version of the Horne and Östberg chronotype test (Lama et al. 2008) , showing a mean difference of 18 points: 3 have an intermediate chronotype and 2 moderately evening. One of the participants shown a very irregular pattern during nights, because personal reasons, not for any medical condition.
- Minimum knowledge about the experiment. Just one of the subjects involved had a good knowledge about features of the wearables, issues related to the transfer of data and mathematical processes to estimate. The other subjects did not have any specific knowledge about the operations. They were asked to proceed with their lives in a regular way, without paying special attention to the wearable, just answering the daily question about the perceived level of rest.
- Use of devices and number of samples. The assignment of devices to subjects was performed randomly. All the subjects wear the devices without any incidence related to their use. There were some issues related to the drain of batteries, because the subject misses to recharge the device. Each subject provided a minimum of 30day samples.


### 5.1. Sleep quality tests

The calculation of the sleep quality required the subject to use a smartphone app that has to be attended every day at mid-morning. The app queries the subject about the level of rest felt the previous night. The user can provide a value between 0 and 4 , being 0 not-restful sleep and 4 fully-restful sleep. Participants in the experiment were fully concerned to provide true answers. Nevertheless, if someday a subject provided an erroneous value, such a sample would be considered as an outlier and its effect in the training model would be minimized. In case subjects answer randomly, the classifier would also behave randomly or it would not provide any value, because no relation among the data could be identified. This would be reflected as a high mean error.

As it was explained in the previous section the predictor is trained and after an initial period of 5 days it provides estimations of the sleep quality. The testing of this predictor was performed during a period of 180 days. Table 12 shows the results of the Mean Error, Standard Deviation and Number of Samples. The table shows the results obtained using this Machine Learning Predictor (MLP) in relation to the results provided by vendors. These values are shown taking into account each of the wearables used.

As it is described in section 4.1, we used the $\mathrm{K}^{*}$ classifier with variables: sleep duration, fall as sleep, awake, HRmin (the minimum value of all the HR samples) and average ST. The collection of these variables depends on the wearable used. In the Fitbit and Jawbone cases, they can just be collected from the proprietary warehouse using the provided

APIs, without any choice to access raw data from sensors. Meanwhile, in the Microsoft Band case the two options were available. Therefore, we manage different data for each wearable:

- From Jawbone: sleep duration, fall as sleep and awake. HRmin and average ST are not available. All this data elements are collected from the Jawbone warehouse.
- From Fitbit: sleep duration, fall as sleep, awake and HRmin. The average ST is not available. All this data elements are collected from the Fitbit warehouse.
- From Microsoft Band: sleep duration, fall as sleep, awake, HRmin and average ST. The first three data elements are collected from the Microsoft Health warehouse, while the HRmin and the average ST are collected directly from the wearable. We choose the minimum value of all the HR samples taken during the sleep night period. For the temperature, we calculate the mean of all the samples. In both cases, outliers such a HR under 10 or a temperature of 0 are discarded. These values can be obtained when the subject manipulates the wearable. The sampling rate is 1 Hz for both sensors. The HR can also provide abnormal peak values. To reduce the effect of these peaks, we use the mean of the last 4 samples.
$\square$
Table 12. Results of the sleep quality tests.

| Predictor | Mean Error | Standard <br> Deviation | Number of samples |
| :---: | :---: | :---: | :---: |
| Fitbit MLP | $13,05 \%$ | $15,16 \%$ | 187 |
| Microsoft Band MLP | $13 \%$ | $16,93 \%$ | 50 |
| Jawbone MLP | $22,5 \%$ | $16,54 \%$ | 30 |
| Total MLP | 14,93 | $15,79 \%$ | 267 |
| Fitbit | $35,73 \%$ | $16,09 \%$ | 187 |
| Microsoft Band | $18,72 \%$ | $11,78 \%$ | 50 |
| Jawbone | $26,83 \%$ | $18,29 \%$ | 30 |
| Total Vendors | $30,55 \%$ | $15,67 \%$ | 267 |

The results obtained from this tests show that the MLP provides better results than vendor solutions. The mean error is reduced by a factor of two in relation to the vendors' estimations in mean, and in all cases the MLP provides better results. Fig. 5 to 8 show graphically the estimations provided by the MLP in comparison to Microsoft and Fitbit for two different subjects: user1 and user2. As it can be shown in the table 12 there are significant differences among the results provided by the vendors, providing Microsoft the best estimations. An explanation for this difference can be found in the sleep regularity patterns of different subjects. The number of samples of Fitbit is much larger than in the Microsoft case. Fig. 5 to 8 show the differences in sleep patterns for some of the subjects involved in our experiments. In addition to the graphics provided by Microsoft and Fitbit, that are shown almost flat along all the days, the result provided by our algorithm shown more waves, and it is more correlated to the answers provided by the subjects about his/her subjective sleep feeling. In most cases the error is just in one level among the five possible ones in the scale $(0,25,50$, 75, 100).


Fig. 5 Comparison chart between Microsoft sleep efficiency and the user 1 response


Fig. 6 Comparison chart between the machine learning predictor (using Microsoft data) and the user 1 response


Fig. 7 Comparison chart between Fitbit sleep efficiency and the user2 response


Fig. 8 Comparison chart between the machine learning predictor (using Fitbit data) and the user2 response

### 5.2. Sleepiness test

To evaluate the sleepiness index, we provided a different app to the subject, as it was showed in Fig. 3. In this case, the subject has to indicate from time to time if she feels sleepier or alert (wide_awake, sleepiness, normal). Furthermore, having as a goal to improve the classifier, a time period from one of the nights where HR and skin temperature values are observed as constant is selected as a sleepiness tag. Then, the model is trained and the estimation is provided in real time.

In summary, the experiment involved the collection of the following data:

- Biometric data: HR, skin temperature and accelerometer.
- Samples used to evaluate and train the classifier: 25199 samples in total, 9079 wide-awake samples and 16120 sleepiness samples.

In the Microsoft Band case, the HR, ST and accelerometer samples were collected directly from the wearable. Outliers in HR and ST were discarded. The accelerometer also can provide outlier values. Regular values are in the range 1 to 0 and therefore, values over 20 are discarded. For the HR, it is stored the mean of each 4 samples in a row. The accelerometer is used as movement sensor. We don't take into account the discrepancies at each axis $x, y$ and $z$ separately, but the module of the values. This value is calculated every time a movement is detected. The accuracy of the accelerometer depends on the sampling rate and in our case, it is rather low: 8 Hz . This value was decided to reduce the energy consumption as much as possible and in this way, extend the battery duration. Each one of these samples is saved in an auxiliary variable and compared with the next one. If the new value is greater than the former one, the auxiliary variable is updated, in other case the value is maintained. In other words, we keep the maximum value in the period. This process is repeated until the variable is stored in the database when the $4{ }^{\text {th }} \mathrm{HR}$ sample is collected and all the values (average HR, the last value of temperature and maximum accelerometer value) are saved.

To check the real relevance of each attribute we applied a Weka algorithm to evaluate them, using the Ranker search method. The results obtained are shown in table 13. The HR and skin temperature are the best attributes to train the classifier. This is rather logic, because the HR shows very low values while we are sleeping. In case of the skin temperature show high values during sleep time because it is usually maintained with blankets and sleeping clothes. By contrast, during daily the skin temperature is more low and changes more frequently, all of this in the similar temperature conditions. Finally, the accelerometer is also representative because it is related with the movement of the user and therefore we maintain it in the formula.

The application of the cross validation technique to the classifier using the available samples produced an accuracy result over $96 \%$. This figure validates the classifier. This indicates that in a similar situation the classifier detects somnolence with a $96 \%$ success rate. All system is showed in Fig.9.

Fig. 10 shows an example of the result obtained after training the model and using data samples collected during a whole day. This classifier is trained, and then it can be used to predict the input data samples in real time. This means that the samples tagged by the user can be included automatically as training samples. Immediately, the new samples can be used to predict the current sleepiness level and to improve the results of former estimations.

Table 13 Results of Ranker

| Average merit | Rank | Variables |
| :--- | :--- | :--- |
| $0.572+-0.001$ | 1 | HR |
| $0.489+-0.005$ | 2 | Skin temperature |
| $0.42+-0.002$ | 3 | Accelerometer |



Fig. 9 Sleepiness recollection app


Fig. 10 Sleepiness graph sample of a subject during a day

### 5.3. Chronotype test

Based on the chronotype tables introduced in section 2, we produced a Spanish version of the questionnaire of momingness-eveningness of Horne and Östberg (Lama et al. 2008) and administered it to our testing subjects. We collected data from wearables for these subjects and applied our algorithm to estimate the chronotype indicator. The results obtained provide a full success of $100 \%$. As it was indicated at the beginning of this section, we successfully estimated the 3 intermediate and the 2 moderately evening. Table 14 shows the chronotype results with the number of samples used for each subject. It can be observer that the number of samples used in this case is greater than the ones used for the SQ estimation. The reason in this case we used all the samples available. In the SQ case, the samples corresponding to days without subject answer were discarded.
Table 14 Chronotype analytical results

| Subject | Device | Number of Samples | Detection |
| :---: | :---: | :---: | :---: |
| 1 | Fitbit | 94 | moderately evenning |
| 2 | Fitbit | 44 | intermediate |
| 3 | Fitbit | 114 | intermediate |
| 4 | Microsoft band | 55 | moderately evenning |
| 5 | Jawbone | 34 | intermediate |

## 6. Applications in educational contexts

The use of the indicators in educational context is the main goal of our piece of research. This section introduces a first view to the possible applications from two different points of view. The first one is focused on improving the academic performance and to promote the self-regulated learning. The second one is focused on the provision of new services to teachers.

### 6.1. Self-regulated learning

The relation between getting to know yourself and what is better for you seems clear. Self-regulated learning is a pedagogical approach that involves students taking responsibility of their own learning, involving different strategies related to goal setting, planning, monitoring and reflecting (Schunk and Zimmerman 1998; Pintrich 2004).

A main idea about self-regulation is self-awareness. Our proposal can be used to contribute to this approach in very precise ways: allowing a student to be aware about sleep quality, sleepiness level and chronotype. Therefore, an initial application of these indexes is to provide clear information to the student about herself. Being aware of these features the student can perform better decisions to develop learning strategies, studying habits and behaviour decisions. Using the proposed indexes, the student can get information about biometric features to evaluate the best behaviours and routines. The student can also check if changes produce positive or negative results, in this case related to sleep features. Some of the recommendations that could be provided from a supporting system could be:

- To recommend a certain timetable or agenda to facilitate a good sleep condition.
- To recommend some healthy exercises as a strategic to tackle with sleepiness periods and tiredness.

In addition to recommend good habits and to alert about potential problems, the main goal of this system from a SRL point of view is focused on the generation of a study timetable that take advantage of the periods in which the biometric features are best aligned. The sleepiness time periods could be used to perform some type of physical or relax activities. The adoption of the recommendations during the school time periods seems not feasible, but they can be followed in case of homework and extra-academic activities. In the case of school time the organization of the educational schools and the differences among learners make impossible to plan the timetables taking into account personal differences. Nevertheless, for the homework and extra-academic activities the student can arrange the timetables taking into account her personal features and preferences. In case a LMS is used, the system could check the pending tasks for the student and recommend some schedule to perform the them. Moreover, the system could alert the student if too much tasks are pending for the next days taking into account the sleep features of the student.

### 6.2. Services for teachers

The proposed indexes can also be provided to teachers to support their academic activities. For example, this information could be used to group students, to get knowledge about the students' engagement, to arrange the timetables, etc.

Collaborative and cooperative activities play a main role in current pedagogical approaches. Working with other ones, students develop main competences in the new curriculums, such as communication, collaboration, leadership, group work, etc. The arrangement of students into groups could be supported taking into account the proposed indexes:

- Groups are created taking into account the chronotype index. Students showing a similar chronotype could be grouped together because in this way they will be better "tuned" among them, at least in accordance to this biorhythm. This will increase the opportunities to be active at the same time periods and to be more productive. By the contrary, the grouping of students with opposite chronotype may provoke that students work mainly in an individual way, because they have different preferences, and few opportunities to work together.
- Groups are created taking into account the number of sleep hours in days previous to an exam. This feature can also be used to identify similarities among different students. In this way, students with similar habits can be grouped together.

These indexes can be used to estimate the student engagement in different activities. Observing the sleepiness level of the whole classroom, while taking into account the sleep quality indicator and the average chronotype, can provide valuable information about the best and worse activities from an engagement point of view. Sometimes teachers find difficulties to know if their students are really interested or paying attention. The availability of this information, probability in an aggregated way or as a classroom average, can contribute to get a better knowledge.
Finally, the teacher could also use this information to arrange the activities and change the timetable in certain ways. Beyond this, in the future some schools could even adapt their timetables to take into account the sleep features of their students.

## 7. Conclusions

The use of commercial off the shelf wrist wearables to provide sleep indicators is a feasible option already. Vendors usually include some types of sleep indicators, such as the sleep duration, the number of awakes or the time to fall asleep, but these are not rigorous enough and provide a low interpretative value. Based on the existing scientific literature and in the data that can be collected from commercial off the shelf wrist wearables, this paper proposes new sleep-related indicators (sleep quality, sleepiness level, chronotype and sleep regularity) and the corresponding algorithms to calculate them in an automatic and autonomous way. The proposed indicators are based on existing medical ones, but they are not intended to replace the medical procedures. By the contrary, the goal is to demonstrate that such indicators can be estimated from existing wearables with an acceptable accuracy. This opens an opportunity to offer data about sleep with a good interpretative value, particularly in the educational domain, where the relationship between sleep and academic performance has been well-established by several studies. In this way, the proposed indicators can be used to prevent restlessness of students and improve their performance and motivation. The paper shows an initial evaluation of the indicators and procedures following a self-reporting approach. First results are positive, better than the values provided by wearable manufacturers and with a high accuracy in accordance to the perception of subjects. In any case, further experiments in test environments are needed to confirm these studies, particularly the comparison with somnography procedures.

The focus on sleep indicators is new regarding existing methods, based mainly on the use of actigraphy techniques with accelerometer data. In our proposal, different types of sensors are used, such as the Heart Rate (HR) and the Skin Temperature (ST), providing valuable data about the sleep condition. Some issues about the use of the commercial off the shelf wearables need to be mentioned. First, the availability of sensors and their precision on these wearables is far from desirable. Basically, vendors are more interested in providing accelerometer, GPS, gyroscope and in some cases HR sensors that are the most "valuable" ones for fitness purposes. Second, the chances to collect data in real time are also very reduced, and this limits the options to calculate some indicators, particularly the SL. In any case, taking into account the rapid evolution of this domain, we hope in the next year a good number of vendors will provide new affordable commercial off the shelf wearables, particularly wrist bands and smart watches, that can provide these sensors data in real-time without any issue. Thirdly, the duration of the batteries is a main issue, because when data is collected in continuous basics they drain rapidly. Fourthly, standardization is needed because there are many differences among vendors, e.g. data models, states vocabularies or token durations to authorize data collection, introducing many complexities in the programming for the data collection and processing.

## 8. Acknowledge

This work has been funded by the European Regional Development Fund (ERDF) and the Regional Government of Galicia in part under agreement for funding the Atlantic Research Center and in part under the GRC2013-006 program and the employment contract granted by the University of Vigo in July 2016 for the performance of PhD studies.

## 9. Conflicts of Interest

Conflicts of Interest: The authors declare no conflict of interest.

## 10. References

Aha DW, Kibler D, Albert MK (1991) Instance-based learning algorithms. Mach Learn 6:37-66. doi: 10.1007/BF00153759

Allen J (2007) Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 28:R1R39. doi: 10.1088/0967-3334/28/3/R01

Alt JA, Smith TL, Mace JC, Soler ZM (2013) Sleep quality and disease severity in patients with chronic rhinosinusitis. Laryngoscope 123:2364-70. doi: 10.1002/lary. 24040

Ancoli-Israel S (2005) Actigraphy. In: Principles and Practice of Sleep Medicine. pp 1459-1467
Apple (2016) Apple Watch. http://www.apple.com/es/shop/buy-watch/apple-watch-sport. Accessed 16 Nov 2016
Baust W, Bohnert B (1969) The regulation of heart rate during sleep. Exp Brain Res 7:169-180. doi:
10.1007/BF00235442

Bei B, Wiley JF, Trinder J, Manber R (2016) Beyond the mean: A systematic review on the correlates of daily intraindividual variability of sleep/wake patterns. Sleep Med Rev 28:108-124. doi: 10.1016/j.smrv.2015.06.003
Belenky G, Wesensten NJ, Thorne DR, et al (2003) Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: A sleep dose-response study. J Sleep Res 12:1-12.
Bernal CC, Armengol ÁS, Ramírez JDA, et al (2012) Artículo Especial. Rev Esp Patol Torac 24:214-254.
Breiman L (1996) Bagging predictors. Mach Learn 24:123-140. doi: 10.1007/BF00058655
Breiman L (2001) Random Forests. Mach Learn 45:5-32. doi: 10.1023/A:1010933404324
Bunde A, Havlin S, Kantelhardt JW, et al (2000) Correlated and Uncorrelated Regions in Heart-Rate Fluctuations during Sleep. Phys Rev Lett 85:3736-3739. doi: 10.1103/PhysRevLett.85.3736

Buysse DJ, Reynolds CF, Monk TH, et al (1989) The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 28:193-213.

Carskadon MA, Dement WC, Mitler MM, et al (1986) Guidelines for the multiple sleep latency test (MSLT): a standard measure of sleepiness. Sleep 9:519-524.

Chent Z, Lint M, Chent F, et al (2013) Unobtrusive Sleep Monitoring using Smartphones. doi:
10.4108/icst.pervasivehealth. 2013.252148

Cleary JG, Cleary JG, Trigg LE (1995) K*: An Instance-based Learner Using an Entropic Distance Measure. Proc 12TH Int Conf Mach Learn 108--114.

Crabtree IB, Rhodes B (1998) Wearable Computing and the Remembrance Agent. BT Technol J 16:118-124. doi: 10.1023/A:1009642301754

Curcio G, Ferrara M, Gennaro L De (2006) Sleep loss, learning capacity and academic performance. Sleep Med Rev 10:323-337.
de Arriba Pérez F, Caeiro Rodríguez M, Santos Gago J (2016a) Collection and Processing of Data from Wrist Wearable Devices in Heterogeneous and Multiple-User Scenarios. Sensors 16:1538. doi: 10.3390/s16091538
de Arriba Pérez F, Santos Gago JM, Caeiro Rodríguez M (2016b) Calculation of Sleep Indicators in Students Using Smartphones and Wearables. New Adv Inf Syst Technol 445:169-178.

Dewald JF, Meijer AM, Oort FJ, et al (2010) The influence of sleep quality, sleep duration and sleepiness on school performance in children and adolescents: A meta-analytic review. Sleep Med Rev 14:179-189. doi: 10.1016/j.smrv.2009.10.004

Duarte J, Nelas P, Chaves C, et al (2014) Sleep-wake patterns and their influence on school performance in Portuguese adolescents. Atención Primaria 46:160-164. doi: 10.1016/S0212-6567(14)70085-X

Ermes M, Parkka J, Mantyjarvi J, Korhonen I (2008) Detection of Daily Activities and Sports With Wearable Sensors in Controlled and Uncontrolled Conditions. IEEE Trans Inf Technol Biomed 12:20-26. doi: 10.1109/TITB.2007.899496

Fitbit (2016a) Fitbit. https://www.fitbit.com/. Accessed 16 Nov 2016
Fitbit (2016b) How do I track my sleep?
http://help.fitbit.com/articles/en_US/Help_article/1314/?q=efficiency\&l=en_US\&fs=Search\&pn=1\#howissleepeff . Accessed 16 Nov 2016

Frank E, Hall M, Pfahringer B (2003) Locally weighted naive bayes. Proc Ninet Conf Uncertain Artif Intell 249-256.
Freund Y, Freund Y, Schapire RE (1996) Experiments with a New Boosting Algorithm. Proc Thirteen Int Conf Mach Learn 148--156.

Guo F, Li Y, Kankanhalli M, Brown M (2013) An evaluation of wearable activity monitoring devices.
Harrison Y, Horne JA, Rothwell A (2000) Prefrontal neuropsychological effects of sleep deprivation in young adults--a model for healthy aging? Sleep 23:1067-73.
Healey, Jennifer Anne (2000) Wearable and automotive systems for affect recognition from physiology.
Herscovici S, Pe'er A, Papyan S, et al (2007) Detecting REM sleep from the finger: an automatic REM sleep algorithm based on peripheral arterial tone (PAT) and actigraphy. Physiol Meas 28:129-140. doi: 10.1088/09673334/28/2/002
Holte RC (1993) Very Simple Classification Rules Perform Well on Most Commonly Used Datasets. Mach Learn 11:63-90. doi: 10.1023/A:1022631118932
Horne JA, Ostberg O (1975) A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. Int J Chronobiol 4:97-110.
Horzum MB, Önder İ, Beşoluk Ş (2014) Chronotype and academic achievement among online learning students. Learn Individ Differ 30:106-111. doi: 10.1016/j.lindif.2013.10.017
IDC (2016a) IDC Forecasts Wearables Shipments to Reach 213.6 Million Units Worldwide in 2020 with Watches and Wristbands Driving Volume While Clothing and Eyewear Gain Traction. http://www.idc.com/getdoc.jsp?containerId=prUS41530816. Accessed 16 Nov 2016
IDC (2016b) Worldwide Wearables Market Increases 67.2\% Amid Seasonal Retrenchment, According to IDC. http://www.idc.com/getdoc.jsp?containerId=prUS41284516. Accessed 16 Nov 2016
Inc V (2016) Vandrico Inc. http://vandrico.com/wearables/. Accessed 16 Nov 2016
Jawbone (2016) Jawbone. https://jawbone.com/up/trackers. Accessed 16 Nov 2016
Jean-Louis G, Kripke DF, Cole RJ, et al (2001) Sleep detection with an accelerometer actigraph: comparisons with polysomnography. Physiol Behav 72:21-28. doi: 10.1016/S0031-9384(00)00355-3
Jersey (2016) Jersey. https://jersey.java.net/. Accessed 16 Nov 2016
John GH, Langley P (1995) Estimating continuous distributions in Bayesian classifiers. Proc Elev Conf Uncertain Artif Intell 338-345.
Johns MW, others (1991) A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 14:540-545.

Johnson ML, Kushida CA, Anderson WM, et al (2003) Practice parameters for the role of actigraphy in the study of sleep and circadian rhythms: an update for 2002. Sleep 26:337-341.
Johnston SL (2005) Societal and workplace consequences of insomnia, sleepiness, and fatigue.
K Tehrani AM (2014) Wearable technology and wearable devices: Everything you need to know. In: Wearable Devices Mag. http://www.wearabledevices.com/what-is-a-wearable-device/. Accessed 16 Nov 2016

Keller PS, El-Sheikh M, Buckhalt JA (2008) Children's Attachment to Parents and Their Academic Functioning: Sleep Disruptions as Moderators of Effects. J Dev Behav Pediatr 29:441-449. doi: 10.1097/DBP.0b013e318182a9b4

Kikhia B, Stavropoulos TG, Meditskos G, et al (2015) Utilizing ambient and wearable sensors to monitor sleep and stress for people with BPSD in nursing homes. J Ambient Intell Humaniz Comput 1-13. doi: 10.1007/s12652-015-0331-6

Kräuchi K (2002) How is the circadian rhythm of core body temperature regulated? Clin Auton Res 12:147-149. doi: 10.1007/s10286-002-0043-9

Kritikou I, Basta M, Vgontzas AN, et al (2013) Sleep apnoea, sleepiness, inflammation and insulin resistance in middleaged males and females.
Lama MAR de, Otálora BB, Teresa M, et al (2008) Versión castellana del cuestionario de matutinidad-vespertinidad de Horne y Östberg (revisado).

LG (2016) LG G Watch R. http://www.lg.com/es/wearables/lg-LGW110-g-watch-r. Accessed 16 Nov 2016
Liao W-H, Kuo J-H (2013) Sleep monitoring system in real bedroom environment using texture-based background modeling approaches. J Ambient Intell Humaniz Comput 4:57-66. doi: 10.1007/s12652-011-0067-x
Lockley SW, Cronin JW, Evans EE, et al (2004) Effect of reducing interns’ weekly work hours on sleep and attentional failures. N Engl J Med 351:1829-1837.

Lucassen EA, Zhao X, Rother KI, et al (2013) Evening Chronotype Is Associated with Changes in Eating Behavior, More Sleep Apnea, and Increased Stress Hormones in Short Sleeping Obese Individuals. PLoS One 8:e56519. doi: 10.1371/journal.pone. 0056519

MacKay D (1998) Introduction to Gaussian processes.
Manber R, Bootzin RR, Loewy D (1998) Sleep Disorders. In: Comprehensive Clinical Psychology. pp 505-527
Mark H, Ian W, Eibe F (2011) Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publishers

Medeiros ALD, Mendes DBF, Lima PF, Araujo JF (2003) The Relationships between Sleep-Wake Cycle and Academic Performance in Medical Students. Biol Rhythm Res 32:263-270. doi: 10.1076/brhm.32.2.263.1359

Microsoft (2016a) Microsoft Band. https://www.microsoft.com/microsoft-band/en-us/features. Accessed 16 Nov 2016
Microsoft (2016b) Track your sleep. https://support.microsoft.com/en-us/help/4000337/band-health-and-exercise-sleeptracking. Accessed 16 Nov 2016

Natale V, Drejak M, Erbacci A (2012) Monitoring sleep with a smartphone accelerometer.
Nitzan M, Romem A, Koppel R (2014) Pulse oximetry: fundamentals and technology update. Med Devices (Auckl) 7:231-9. doi: 10.2147/MDER.S47319

Paul Burton (2015) Apple Watch Saves Life Of Tabor Academy Football Player. http://boston.cbslocal.com/2015/09/20/apple-watch-tim-cook-heart-rate-tabor-academy-marion-massachusetts-paul-houle/. Accessed 16 Nov 2016

Pintrich P (2004) A conceptual framework for assessing motivation and self-regulated learning in college students.
Quan S, Gillin J, Littner M, Shepard J (1999) Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. editorials.

Quinlan JR (John R, Ross J (1993) C4.5 : programs for machine learning. Morgan Kaufmann Publishers
Raymann RJEM, Swaab DF, Van Someren EJW (2007) Skin temperature and sleep-onset latency: Changes with age and insomnia. Physiol Behav 90:257-266. doi: 10.1016/j.physbeh.2006.09.008

Richmond S The real world wrist-based heart rate monitor test: Are they accurate enough? http://www.wareable.com/fitness-trackers/heart-rate-monitor-accurate-comparison-wrist. Accessed 16 Nov 2016

Ruck DW, Rogers SK, Kabrisky M, et al (1990) The multilayer perceptron as an approximation to a Bayes optimal discriminant function. IEEE Trans Neural Networks 1:296-298. doi: 10.1109/72.80266
Rudner J, McDougall C, Sailam V, et al (2016) Interrogation of Patient Smartphone Activity Tracker to Assist Arrhythmia Management. Ann Emerg Med 68:292-294. doi: 10.1016/j.annemergmed.2016.02.039
SAMSUNG (2016) SAMSUNG. http://www.samsung.com/es/consumer/mobile-devices/wearables/filter/. Accessed 16 Nov 2016
Sano A, Eng B (2016) Measuring College Students' Sleep, Stress, Mental Health and Wellbeing with Wearable Sensors and Mobile Phones. Massachusetts Institute of Technology
Sano A, Phillips AJ, Yu AZ, et al (2015) Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. In: 2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN). IEEE, pp 1-6
Schunk D, Zimmerman B (1998) Self-regulated learning: From teaching to self-reflective practice.
Shevade SK, Keerthi SS, Bhattacharyya C, Murthy KRK (2000) Improvements to the SMO algorithm for SVM regression. IEEE Trans Neural Networks 11:1188-1193. doi: 10.1109/72.870050
Sleep as Android (2016) Sleep as Android. http://sleep.urbandroid.org/. Accessed 16 Nov 2016
Snyder F, Hobson JA, Morrison DF, Goldfrank F (1964) Changes in respiration, heart rate, and systolic blood pressure in human sleep.
Soehner AM, Kennedy KS, Monk TH (2011) Circadian Preference and Sleep-Wake Regularity: Associations With Self-Report Sleep Parameters in Daytime-Working Adults. Chronobiol Int 28:802-809. doi: 10.3109/07420528.2011.613137

Stahl SE, An H-S, Dinkel DM, et al (2016) How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? BMJ Open Sport Exerc Med 2: e000106. doi: 10.1136/bmjsem-2015-000106
Swan M (2012) Sensor Mania! The Internet of Things, Wearable Computing, Objective Metrics, and the Quantified Self 2.0. J Sens Actuator Networks 1:217-253.
Telfer S, Spence WD, Solomonidis SE (2009) The Potential for Actigraphy to Be Used as an Indicator of Sitting Discomfort. Hum Factors J Hum Factors Ergon Soc 51:694-704. doi: 10.1177/0018720809352789
ur Rehman MH, Liew CS, Wah TY, et al (2015) Mining Personal Data Using Smartphones and Wearable Devices: A Survey. Sensors 15:4430-4469.
Valenti G, Westerterp KR (2013) Optical heart rate monitoring module validation study. In: 2013 IEEE International Conference on Consumer Electronics (ICCE). IEEE, pp 195-196
Vanoli E, Adamson PB, Ba-Lin, et al (1995) Heart Rate Variability During Specific Sleep Stages.
Wallis Snowdon CN (2016) Apple watch saved Alberta man's life, makes international headlines - Edmonton - CBC News. http://www.cbc.ca/news/canada/edmonton/apple-watch-saved-alberta-man-s-life-makes-international-headlines-1.3495397. Accessed 16 Nov 2016
Weka ZeroR. http://weka.sourceforge.net/doc.dev/weka/classifiers/rules/ZeroR.html. Accessed 13 Feb 2017a
Weka LinearRegression. http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/LinearRegression.html. Accessed 13 Feb 2017b
WIMMER F, HOFFMANN RF, BONATO RA, MOFFITT AR (1992) The effects of sleep deprivation on divergent thinking and attention processes. J Sleep Res 1:223-230. doi: 10.1111/j.1365-2869.1992.tb00043.x
Xiaomi (2016) Mi Band. http://www.mi.com/en/miband/\#01. Accessed 16 Nov 2016

