

Sleep Hunter: Towards Fine Grained Sleep Stage Tracking with Smartphones

Weixi Gu, *Student Member, IEEE*, Zheng Yang, *Member, IEEE*, Longfei Shangguan, *Student Member, IEEE*, and Yunhao Liu, *Member, IEEE*

Abstract—Sleep quality plays a significant role in personal health. A great deal of effort has been paid to design sleep quality monitoring systems, providing services ranging from bedtime monitoring to sleep activity detection. However, as sleep quality is closely related to the distribution of sleep duration over different sleep stages, neither the bedtime nor the intensity of sleep activities is able to reflect sleep quality precisely. To this end, we present Sleep Hunter, a mobile service that provides a fine-grained detection of sleep stage transition for sleep quality monitoring and intelligent wake-up call. The rationale is that each sleep stage is accompanied by specific yet distinguishable body movements and acoustic signals. Leveraging the built-in sensors on smartphones, Sleep Hunter integrates these physical activities with sleep environment, inherent temporal relation and personal factors by a statistical model for a fine-grained sleep stage detection. Based on the duration of each sleep stage, Sleep Hunter further provides sleep quality report and smart call service for users. Experimental results from over 30 sets of nocturnal sleep data show that our system is superior to existing actigraphy-based sleep quality monitoring systems, and achieves satisfying detection accuracy compared with dedicated polysomnography-based devices.

Index Terms—Smartphones; sleep stage; sensors.

1 INTRODUCTION

SLEEP, occupying nearly one-third of human lifetime, is a necessary and vital biological function. Physiological communities often regard sleep as a cyclical process composed of three stages: rapid eye movement (REM) stage, light sleep stage and deep sleep stage [1]. The biological characteristics of different sleep stages exhibit distinguishingly. REM is an active period of sleep marked by intense brain activities and dream occurrence. Light sleep stage is a period of relaxation, when the heartbeat, breathing rate and muscle activity slow down. Deep sleep stage triggers hormones to promote body growth, as well as the repair and restoration of energy. Sleep quality is actually determined by the distribution of different sleep stages rather than the time of sleep [2]. Even though people experience a full night's sleep, they may still feel fatigued after being woken up. A longer deep sleep period contributes to a better sleep. Moreover, a proper wake-up time is also helpful for mental and physical health [3]. Comparing with other sleep stages, people woken up in light sleep stage feel more refreshed generally.

Sleep quality monitoring requires a careful observation of individuals' sleep stages. The approaches to recording sleep stages are divided into two categories. The first category is based on polysomnography [4]. The methods in this category leverage electroencephalograph (EEG) to observe brain waves and then

recognize sleep stages accurately. EEG systems, *e.g.*, Zeo [5], however, are usually limited to medical and physiological studies. High cost and complicated operations make them hard to be accepted by the public. The approaches in the second category are based on actigraphy [6] utilizing some certain physical activities such as body movement or snore to predict sleep stage. The detection performance based on these approaches, *e.g.*, Jawbone up [7] and Sleep As Android [8], has not yet been evaluated against medically accepted methods. In this paper, we show their performance to be substantially lesser than the methods used in physiological studies.

In this paper, we present Sleep Hunter, a sleep stage detection system based on actigraphy that predicts sleep stage transitions by smartphone. The information collected by Sleep Hunter can be used to evaluate human sleep quality and provide smart call service, which wakes up users in light sleep stage intelligently. The principle behind our system is that apart from implicit brain wave changes, individuals usually exhibit distinguishable physical activities during different sleep stages. For example, in REM, breathing rate is commonly unstable and people tend to exhibit large body movements. Whereas in the deep sleep stage, breathing rate becomes slower and more regular, accompanied with slight body movements such as arm trembling and leg jerking [9]. Moreover, sleep environment, sleep duration and some certain personal factors also impact the transition of sleep stages. Sleep Hunter leverages built-in sensors of the smartphone to detect such sleep-related events and then predicts the transition of sleep stages overnight. Based on the detected sleep stages, it generates a corresponding score for sleep quality evaluation and provides smart call service. Compared with the methods in the first category, Sleep Hunter is a service that runs on a commercial off-the-shelf smartphone which requires no additional device. The simple operation makes it more convenient than polysomnography-based systems. Furthermore, Sleep Hunter

- Weixi Gu is with the School of Software, Tsinghua University, Beijing, China, 100083.
E-mail: guweixigavin@gmail.com
- Zheng Yang is with the School of Software, Tsinghua University, Beijing, China, 100083.
E-mail: yang@greenorbs.com
- Longfei Shangguan is with CSE, Hong Kong University of Science and Technology, Hong Kong.
E-mail: longfei@greenorbs.com
- Yunhao Liu is with the School of Software, Tsinghua University, Beijing, China, 100083.
E-mail: yunhao@greenorbs.com

TABLE 1
The physical activities happened in sleep process

Activity	Explanation
Tachypneic breath	A condition of rapid breathing, commonly between 12-20 breaths per minute.
Apneustic breath	A series of slow, deep breathing, lasting about 6-10 seconds, after which the air is suddenly expelled by the elastic recoil of the lung.
Cough	A sudden and often repetitively occurring reflex which helps to clear the large breathing passages from secretions, irritants.
Snore	A vibration of respiratory structures and a resulting sound, due to obstructed air movement during breathing while sleeping.
Somniloquy	A parasomnia that refers to talking aloud while asleep. It can be quite loud, ranging from simple sounds to long speeches, and can occur many times during sleep.
Macro body movement	A serials of significant activities happened in sleep cycles, such as turning body over, driving or raising legs and so on.
Micro body movement	A serials of tiny activities appeared in sleep cycles, including short convulsion and hand trembling, head moving.

integrates sleep-related events comprehensively and leverages a statistical model to predict sleep stage. The fine-grained detection results and promising performance make Sleep Hunter more suitable than those actigraphy-based products in the second category.

We face two challenges when codifying such idea into a practical system. The first challenge is how to identify discriminative activities from a variety of primitive data, given the condition that sensory data are sparse and full of noise. For example, audio signals recorded by the microphone contain not only sleep-related primitives, but also ambient noise. The second challenge is how to leverage sleep-related events to figure out features for capturing sleep stage transitions. Many events such as snore, body movement, sleep duration and even people's age have close relationships with sleep. The influence of these factors should be reasonably utilized.

To address the above challenges, we design a unique feature-extraction mechanism for each of sleep-related events based on their physical characteristics. Moreover, we exploit conditional random field (CRF), a statistical model to parse the relations behind these events according to our over 90 sets of nocturnal sleep data, and evaluate them on our testing dataset.

The contributions of our paper are listed as follows. 1) We put forward a set of efficient algorithms to detect sleep-related events and adopt a CRF to depict the relationship between such events and sleep stages. 2) We implement Sleep Hunter on Android platform and conduct evaluation experiments on 15 participants from various age groups. The result of the testing data over one month demonstrates that the detection accuracy of Sleep Hunter attains 64.55%, which is superior to the existing actigraphy-based applications to our best knowledge. 3) We conduct extensive case studies and show that Sleep Hunter is able to provide sleep quality reports and smart call services for users.

The rest of this paper is organized as follows. We firstly detail the preliminary of sleep in Section 2. We then introduce the key insight and design targets in Section 3. In Section 4, we briefly overview the architecture of Sleep Hunter. Next, the system design is described in Section 5 and the principles behind the sleep stage detection model are shown in Section 6. Then, we present the experimental evaluation in Section 7 and make a discussion of Sleep Hunter in Section 8. Finally, we provide a review of related work in Section 9 and make the conclusion in Section 10.

2 PRELIMINARY

2.1 Sleep cycle

Although it's common to think of sleep as a time of dormancy, sleep is actually an active physiological process. As shown in Fig. 1, physiological communities commonly regard sleep as a cycling process composed of three main sleep stages, namely Rapid Eye Movement (REM) stage (also called dream stage), light sleep stage and deep sleep stage. The light sleep and deep sleep are also called as Non-Rapid Eye Movement (NREM), which is characterized by a reduction in physiological activities. As sleep gets deeper, brain waves become slower and have greater amplitude. At the same time, breathing and heart rate slow down, and blood pressure drops. By analyzing the associated neurological features during sleep, medical studies further partitioned the three main sleep stages into five substages.

Stage 1 (S_1) is a time of drowsiness or transition from being awake to falling asleep. In this stage, the brain produces high-amplitude θ -wave, and eye movements and muscle activities begin to slow down. People in S_1 may experience sudden muscle jerks, and is sensitive to environmental disturbance.

Stage 2 (S_2) is a period of light sleep during which eye movements stop. The brain begins to produce bursts of rapid, rhythmic brain wave activities known as sleep spindles. Besides, body temperature starts to fall and heart rate begins to slow down. This stage commonly lasts for approximately 20 minutes and the body prepares to enter deep sleep after this stage.

Stage 3 (S_3) is often characterized by the presence of slow brain waves called δ -wave interspersed with smaller, faster waves. Blood pressure falls, breathing slows, and body temperature drops even lower. People commonly enter deep sleep in S_3 . So if a

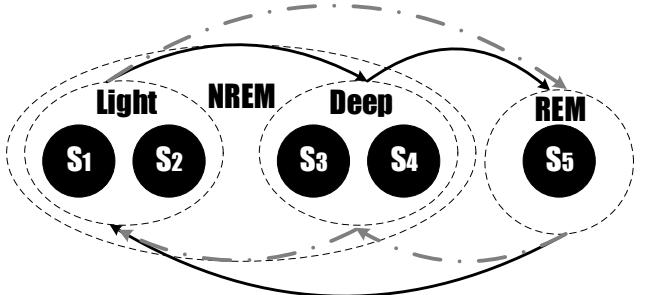


Fig. 1. An illustration of sleep stage transition

phone call is coming during this period, people are groggy and find it difficult to engage in conversations.

Stage 4 (S_4) is a period that brain produces more δ -waves, and eye movements and muscle activities weaken or stop. Some may experience sleepwalking or nightmares after this stage. S_4 is the deepest of the four sleep stages before the REM sleep, lasting for about 30 minutes on average.

REM (S_5) phase is an active period of sleep marked by intense brain activities. Brain waves in this phase are fast and unsynchronized, similar to those in the awake state. Besides, breathing becomes even rapid, irregular and shallow; eyes move rapidly in various directions and limb muscles become temporarily paralyzed in this stage.

Generally, people commonly experience a transition from light sleep (S_1 and S_2) to deep sleep (S_3 and S_4), lasting about 70 minutes before entering the REM phase. Fig. 1 illustrates the sleep cycle during the night, where sleep follows a predictable pattern, moving cyclically among light sleep stage, deep sleep stage and REM stage. Each cycle typically lasts for about 90 minutes and repeats four to six times over a night.

3 KEY INSIGHT AND DESIGN TARGETS

In this section, we first present the key insight for sleep cycle detection scheme, and then specify the design targets of our system.

3.1 Key Insight

Our key insight here is related to the following aspects. Firstly, apart from the implicit physiological activities (*e.g.*, body temperature changes and brain activity variations), sleepers usually exhibit distinguishable physical activities in different sleep stages [10]. For example, short breaths and large body movements such as body rollovers usually happen in light sleep, resulting from the fast heartbeat. On the contrary, slight body movements such as arm trembling and leg jerking [11] mostly occur in deep sleep due to the slow and regular breathing rate. Moreover, somniloquy and body trembles caused by frequent dreams generally appear in REM. Such physical activities can be detected via off-the-shelf smartphones, serving as the basis for the sleep stage analysis. Table 1 summarizes the physical activities that Sleep Hunter mainly monitors during the sleep process.

Secondly, sleep usually follows a predictable pattern, moving cyclically among light sleep stage, deep sleep stage and REM. Each sleep cycle typically lasts for about 90 minutes and repeats four to six times over a night. In each sleep cycle, sleepers firstly experience a transition from light sleep to deep sleep and then enter REM. This sequence is shown by solid black lines in Fig. 1. Nevertheless, sleep cycle is not an absolute case. The phenomenon of skipping some certain sleep stages usually occurs during sleep. For example, as shown by the grey dashed lines in Fig. 1, sleep stage could jump to REM from light sleep or return to deep sleep from REM directly. The dependence between two successive sleep stages, however, still exists. This inherent temporal sequence, intuitively, could also be utilized for analyzing sleep stages. Moreover, sleep environment, *e.g.*, ambient illumination, and certain personal factors, *e.g.*, age, also impact the sleep phases [11], which help us to predict the transition of sleep stages.

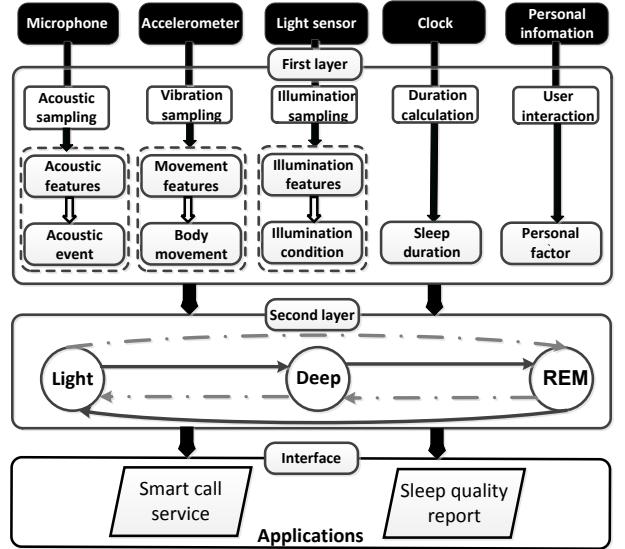


Fig. 2. The architecture of Sleep Hunter

3.2 Design Targets

The principal functions of Sleep Hunter are to monitor the sleep quality and wake people up at proper time instance. Specifically, Sleep Hunter is designed to meet the following requirements. (1) As Sleep Hunter operates over sleeping period, it should be non-intrusive. In other words, users should be unable to tell whether the system is running over night. This is a basic requirement and is widely accepted by sleeping-related products or applications [8], [12]. (2) Sleep Hunter needs to precisely detect the current sleep stage of the user, as well as recording the fine-grained duration of each stage across different users. Since sleep quality can also reflect the mental and physical health state, such fine-grained sleep information can further help to analyze the healthy condition of users. (3) As a long-term running application, Sleep Hunter should be energy-saving and efficient. It should be affordable for users to execute during the whole sleep process, and also be able to detect the state transition promptly.

4 SYSTEM OVERVIEW

4.1 System Architecture

Sleep Hunter is a two-layer system and provides an interface for its applications as shown in Fig. 2. The first layer is composed of five submodules: body movement detection module, acoustic event detection module, illumination condition detection module, sleep duration tracking module and personal factor collection module. Each module is responsible for collecting its related primitive data and extracting associated features. The second layer leverages CRF to integrate features from the upper layer. Based on these collected features, this layer makes sleep stage prediction for the corresponding period, which is called *detection phase*. The duration of detection phase is set to be 5 minutes. In other words, Sleep Hunter detects sleep stage every 5 minutes during sleep. Moreover, Sleep Hunter provides sleep quality report and smart call service for users in the interface layer based on the monitoring results of sleep stages.

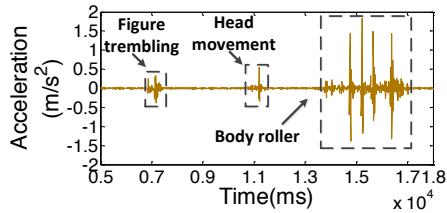


Fig. 3. Acceleration trace of body movement

5 SYSTEM DESIGN

In this section, we specify the design and implementation details for each component in Sleep Hunter.

5.1 Body Movement Detection (BM)

Sleepers usually exhibit various physical activities during different sleep stages. As reported by medical views [10], large body movements like body rollovers usually occur when people are in light sleep, which result from the fast heartbeat during this stage. In contrast, some tiny body movements such as body trembling and leg jerking usually occur in deep sleep stage. Moreover, some unconscious body movements such as the leg stretching and the arm rising would happen during REM, which are caused by frequent dreams. Accordingly, we could leverage the distinguishable movements to detect various sleep stages.

5.1.1 Body Movement Experiments

In order to better understand the phenomenon of body movements during sleep, 100 groups of sleep-related body movement experiments are conducted by 10 volunteers across different ages when they are in bed. Every volunteer contributes 10 groups of experiments. In each experiment, the volunteer puts a smartphone beside his/her head and enables the accelerometer to calculate the corresponding acceleration variance trace of the body movement. The sample rate of accelerometer is set to be 100Hz, which is same as the configuration of Sleep Hunter. Acceleration variance is calculated as $V(i) = a(i) - a(i - 1)$, where $a(i) = \sqrt{a_x(i)^2 + a_y(i)^2 + a_z(i)^2}$ and $a_x(i)$, $a_y(i)$ and $a_z(i)$ represent the accelerometer sample value of X-axis, Y-axis and Z-axis at time stamp i respectively. The sleep-related body movements include body rollover, leg stretching, arm raising, figure trembling, leg jerking and head movement. Fig. 3 plots a real acceleration variance trace of one volunteer's body movements.

5.1.2 Body Movement Extraction

Considering the distribution of inherent accelerometer's noise, we denote the threshold by ξ to classify the accelerations of body movement and noise. If $|V(i)| \geq \xi$, we regard it as an

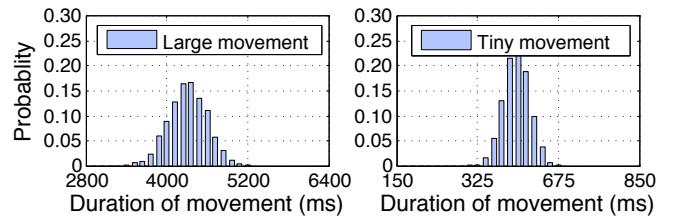


Fig. 4. Durations of body movements

occurrence of body movement. Otherwise, we take it as noise. However, if ξ is set too large, some tiny body movements are likely to be ignored. Conversely, if ξ is too small, some acceleration noises may be mistaken as body movements. Either case would make negative influence on the detection performance of body movement. In order to find an effective threshold. We vary ξ from 0.01 to 0.07 and analyze the precision and recall based on the data trace contains over 4000 body movements. The result is shown in Table. 2. According to the results, we set ξ as 0.05 which locally optimizes the identification scheme.

However, one issue of this mechanism is that body movements such as body rollover and leg stretching are not continuous. The temporal pauses in body movements would cause our mechanism mistakenly to split a single body movement into multiple movements. To solve this problem, we realize the longest temporal pause in our body movement experiments belongs to the body rollover, which lasts less than 1.5s. Sleep Hunter empirically merges two successive movements into a single one if they occur within 1.5s.

5.1.3 Body Movement Classification

For a better analysis of the relationship between sleep-related body movements and sleep stages, we calculate the durations of those large, long-lasting actions such as body rollover, leg stretching and arm raising, and tiny, short-lasting activities including arm trembling and leg jerking from our body movement experiments. Fig. 4 shows the distribution of body movement's durations in two sets. We found that all of the large, long-lasting body movements last at least 2.8s while those tiny, short-lasting activities last at most 0.85s. This obvious temporal gap helps us to distinguish these movements into two categories. In Sleep Hunter, we define the movements lasting less than 1s as micro body movements and those lasting longer as macro body movements, and leverage these two kinds of body movements as sleep-related features.

5.2 Acoustic Event Detection (AE)

Apart from body movements, people usually display some acoustic events during sleep. We concentrate on five common sleep-related acoustic events: somniloquy, tachypneic breath, apneustic breath, snore and cough. Based on the physiological research [10], somniloquy occurs frequently during REM because of continual dreams. Tachypnea, which is easily incurred by the rapid heartbeat, usually happens during light sleep. Moreover, apneusis often appears in deep sleep due to the slow heartbeat, and snore usually emerges in deep or light sleep. Furthermore, cough is not likely to come about during deep sleep, since it can easily disturb the sleep process.

5.2.1 Acoustic Event Experiments

We utilize the built-in microphone, whose sample rate is set to be 16KHz, to record acoustic data. A total of 30 participants join

TABLE 2
The precision, recall, F1 under different ξ settings

ξ	precision	Recall	F1
0.01	0.700	1.000	0.824
0.02	0.700	1.000	0.824
0.03	0.875	1.000	0.933
0.04	0.875	1.000	0.933
0.05	1.000	1.000	1.000
0.06	1.000	0.571	0.727
0.07	1.000	0.286	0.444

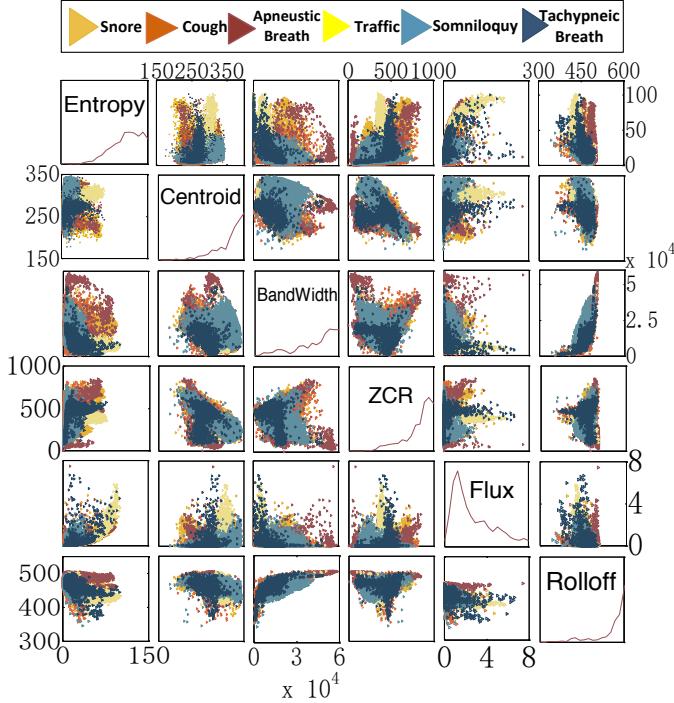


Fig. 5. Correlation matrix between features and acoustic events

our acoustic experiments and their smartphones are placed beside their heads during sleep. According to these recording data, we label acoustic events manually and thus collect 180 sound clips for each type of acoustic event from the 30 people, with each sound clip lasting around 5 seconds. The collected data, therefore, are sufficient for us to study the inner properties of these sleep-related acoustic events.

The acoustic analysis begins with dividing the audio stream from the microphone into frames of equal duration. Each frame is composed of 1024 acoustic samples, and thus its duration is 64ms, which is able to capture the acoustic characteristics of sounds [13].

5.2.2 Noise elimination

Since different kinds of noise exist in the sleep environment, the negligible effect of noise should be eliminated firstly. According to our observation, there are mainly three types of noises during sleep: ambient noise, noise made by body movement and traffic noise. Sleep Hunter leverages the scheme in [13] to differentiate the ambient noise from other types of noise and sleep-related acoustic events. The scheme recognizes ambient noise by its low root-mean-square (RMS) energy and high spectral entropy [14]. If RMS of the current frame is less than a predefined threshold Th_{rms} and its entropy is higher than $Th_{entropy}$, it is ambient noise. Here, we empirically set $Th_{rms} = 0.006$ and $Th_{entropy} = 25$, which locally optimizes the detection accuracy on our acoustic event experiments. Moreover, body movements such as body rollovers during sleep would make additional noise. Since they generally induce variances of acceleration and sound simultaneously, we preset threshold ϵ to measure this simultaneity and leverage it to detect the noise made by body movement. In other words, if the temporal deviation between the beginnings of an acoustic event and a body movement is less than ϵ , and the temporal deviation between their endings is also less than ϵ , we regard the acoustic event as noise made by body movement and

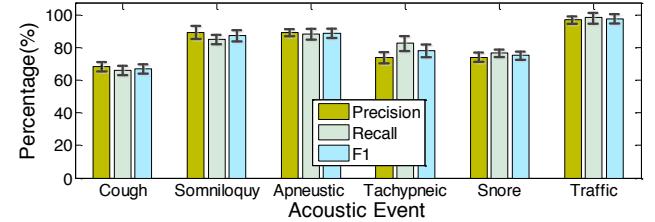


Fig. 6. The performance of acoustic event classification

then filter it out. In Sleep Hunter, we set $\epsilon = 0.5s$ based on our experiments. As the acoustic characteristics of traffic noise are complicated, we classify it with sleep-related acoustic events together.

5.2.3 Acoustic Feature Selection

In this section, we extract features of sleep-related acoustic events and traffic noise. Several most representative features for acoustic event classification are chosen. These features are also widely adopted by various works [12], [13], [15].

The features are divided into two categories as follows.

1) Time-domain feature:

1. **Zero Crossing Rate (ZCR):** $\frac{1}{2} \sum_{j=1}^m |sign(s_j) - sign(s_{j-1})|$, where s_j stands for the j th sample in a frame who has m samples. $sign()$ is the sign function. Zero Crossing Rate [16] is an indicator to represent the rate where the sample value changes from positive to negative or back in time domain. When an acoustic event possesses a mount of voiced and unvoiced sounds, its ZCR is high. However, if the sound of an acoustic event is stable.

2) Frequency-domain feature :

To better extract frequency-domain features, we firstly calculate the frequency spectrum of each frame by N -point Fast Fourier Transform(FFT), where N equals 1024. We then normalize the spectrum of each frame and let $f_t(j)$ be the normalized magnitude of the j th frequency bin in the spectrum of frame f_t . We compute the following spectral features for acoustic event classification.

1. **Spectral Entropy :** $-\sum_{j=1}^N f_t(j) \log f_t(j)$. Spectral Entropy [17] has been widely used to evaluate the flatness of the acoustic spectrum shape. A flat spectrum usually embraces high entropy value, whereas the entropy value of those embrace irregular spectrum shaped or contain notable patterns tends to be low.

2. **Spectral Centroid :** $\sum_{j=1}^N j \cdot |f_t(j)|^2 / \sum_{j=1}^N |f_t(j)|^2$. Spectral Centroid [14] aims to calculate the balancing point of the power spectral distribution. Intuitively, Spectral Centroid value of high frequency sounds such as snoring is likely to be higher than that of low frequency sounds such as apneustic or short breathing. As each acoustic event has unique power spectral distribution, we can categorize them based on this metric.

3. **Spectral Flux :** $-\sum_{j=1}^N (f_t(j) - f_{t-1}(j))^2$. Spectral Flux [14] reflects the stability of acoustic events. $f_t(j)$ and $f_{t-1}(j)$ stand for the j th frequency bin in the current frame and the last frame respectively. Ideally, one stable acoustic event usually possesses low spectral flux value for a slight change, i.e., breathing.

reflects the stability of acoustic events. $f_t(j)$ and $f_{t-1}(j)$ stand for the j th frequency bin in the current frame and the last frame respectively. Ideally, one stable acoustic event usually possesses low spectral flux value for a slight change, i.e., breathing.

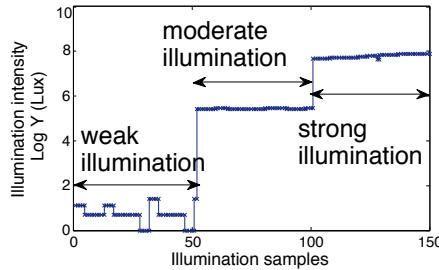


Fig. 7. Illumination intensity under different conditions

Whereas, unsteady audio events with significant variation embrace high value of flux, for example, somniloquy and snoring.

$$4. \text{ Bandwidth: } \sum_{j=1}^N (j - Cen)^2 \cdot |f_t(j)|^2 / \sum_{j=1}^N |f_t(j)|^2, \text{ where}$$

Cen indicates the spectral centroid of f_t . Bandwidth [14] describes the continuous frequency range beyond zero from lower to upper ends. It quantifies the concentration of a spectrum around its centroid. Intuitively, an even and smooth spectrum always own a high bandwidth. For instance, surrounding noise whose distribution of frequency spans extensively usually possess a broad bandwidth.

$$5. \text{ Spectral Rolloff: } \max(h) \sum_{j=1}^h f_t(j) < \text{threshold}.$$

Spectral Rolloff [14] indicates the percentage frequency bin below a predefined threshold, which is usually set to be 93%. It reflects the skewness of the spectral distribution. It reflects the skewness of the spectral distribution. The high-frequency acoustic events, who have right-skewed distributions, possess higher value of spectral rolloff than those whose frequency are low.

Fig. 5 illustrates the correlation matrix between the aforementioned acoustic features and the frames of different acoustic events collected by our experiments. Each row/column represents a feature. The curve in the boxes on the diagonal line stands for the distribution of the corresponding feature. We can learn that the flux of somniloquy lies in a relatively low level, which demonstrates that its frequency spectrum is stable. On the contrary, the flux of traffic is generally higher than those of others, which is coherent with our intuition that the frequency spectrum of traffic is abrupt. Moreover, the apneustic breath, whose rolloff is large, possesses a high frequency. Furthermore, we can see that most tachypneic breaths locate where entropy and bandwidth are both low. It shows a fact that the sound of this type of breath possesses an obvious pattern and its frequency spectrum is narrow. Since different acoustic events perform distinguishing distributions under various features, we can leverage such features to classify them.

5.2.4 Acoustic Event Classification

As acoustic features depicted in Fig. 5 are not likely to be classified by a simple linear function, we expect to use a nonlinear classifier to analyze these features and make further classification. Here, we adopt support vector machine (SVM) [18] to express this kind of nonlinear regression. The 10-fold cross-validation [19] has been done on the collected frames of acoustic event across the 30 participants and the classification performance is illustrated in Fig. 6. We can observe that the best classification performance is achieved by traffic. Its average precision and recall reach 98.3% and 96.9% respectively. The average F1 [20] is up to 97.6%. Even though the worst result belongs to

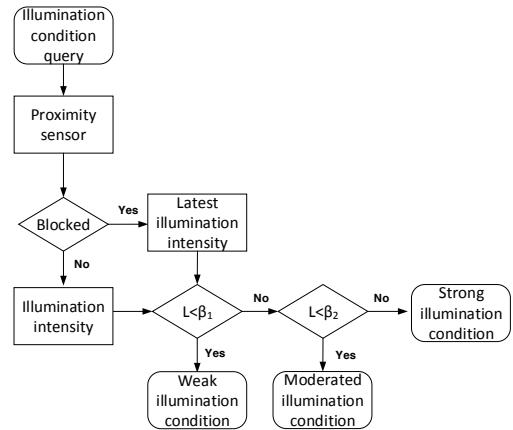


Fig. 8. The flow chart of illumination condition detection

cough, its average precision and recall are still over 66%, and average F1 runs up to 67.13%. Moreover, for other four types of acoustic events (*i.e.*, somniloquy, apneustic breath, tachypneic breath and snore), average F1 values of them are from 70% to 88%, which demonstrates the classification performance is acceptable. Furthermore, the standard deviations of all the indices illustrated by the error bars are less than 6%, which shows that the classification performance of SVM is stable and robust. Therefore, we could filter out the traffic noise and recognize the sleep-related acoustic events, and then leverage them as features to predict sleep stages.

5.3 Illumination Condition Detection (IC)

Generally, sleep quality is also affected by the ambient illumination conditions [11]. People may enjoy a good sleep in dim environment while feel difficult to fall asleep under strong illuminative conditions. To characterize the relationship between sleep stages and illumination intensity around, we categorize the illumination intensities into different conditions, and then explore the transition of sleep stages under these conditions. Fig. 7 illustrates the illumination intensity under different sleep environments, which is broadly categorized into three conditions: bedroom without light (Weak illumination condition, $\leq 10\text{Lux}$ (β_1)); bedroom with dim light (Moderate illumination condition, $10 \sim 2000\text{Lux}$ (β_2))); bedroom with strong lights (Strong illumination condition, $\geq 2000\text{Lux}$). Therefore, we could measure the sleep environment according to the three illumination conditions by the built-in light sensor of smartphones. The sample rate of the light sensor is set to be 100Hz, which is the same as the configuration of Sleep Hunter. The three types of illumination conditions are also leveraged as sleep-related features for sleep stage detection.

According to our survey, for most smartphones, the light sensor is usually installed in the front face of the phone. As a result, if the phone is facing toward the ground, the illumination sensing module could be paralyzed because the illumination samples cannot reflect the real illumination conditions around the phone. Indeed, this phenomenon occurs frequently during sleep given that some unconscious body movements would occasionally change the position of smartphones.

To overcome this challenge, we build a light-weight hierarchical illumination intensity sensing scheme. Fig. 8 illustrates the work flow of this sensing scheme. Firstly, Sleep

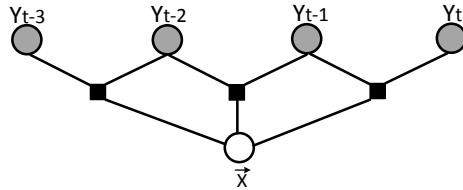


Fig. 9. Conditional random field

Hunter employs the proximity sensor to detect whether the light sensor is blocked or not. If it is not blocked, Sleep Hunter calculates the average illumination intensity l_{cur} in the detection phase, and then determines the current illumination condition by comparing l_{cur} with the two preset thresholds: β_1, β_2 . On the other hand, if the light sensor is blocked, then Sleep Hunter locates the latest record of the illumination condition when the light sensor is not blocked, and treats it as the current illumination condition until the light sensor recovers. Since the illumination condition is hard to change in a short time, we design an energy saving mechanism. If the average illumination condition is stable for the third consecutive detection phase, Sleep Hunter would stop the light sensor for 10min.

5.4 Sleep Duration (SD) and Personal Factor (PF)

Apart from those associated physical activities and sleep environment, the transition of sleep stage is affected by chronological rules statistically during the sleep process. According to the clinical study in [11], the first REM sleep stage usually occurs about 70 to 90 minutes after we fall asleep. A complete sleep cycle takes 90 to 110 minutes on average. Moreover, the first sleep cycle each night contains the relatively short REM phase and the relatively long period of deep sleep. As time goes by, the duration of REM increases while that of deep sleep decreases. By morning, people spend nearly all their sleep time in light sleep and REM. According to this chronological property, Sleep Hunter takes its running time as sleep duration and regards it as a sleep-related feature. Moreover, the sleep stage is also affected by the personal physiological status. For example, the proportion of deep sleep stage decreases with the increase of the user's age [21]. Since the age of sleeper is an important physiological factor impacting sleep [22], [23], Sleep Hunter obtains the user's age from the registration information and further takes it with the sleep duration as features for the sleep stage detection.

6 SLEEP STAGE DETECTION

In this section, we present the details of the sleep stage detection scheme.

We propose a linear-chain conditional random field (CRF) [24] to integrate the aforementioned features and make further inference. CRFs are discriminative models that predict the global probability of a sequence of random variables. The probability structure of a CRF depends on a log-linear combination of observable features and dependence of hidden variables, which is depicted as a bipartite factor graph. CRFs have been widely used in audio, speech, language processing and health sensing [25], [26]. The rationale of CRF applied here lies twofold. Firstly, as depicted in Fig. 1, the occurrence of sleep stages during sleep forms a sequence. This process can be characterized by the CRF model.

Secondly, sleep-related events have dependent relationships. For example, cough or snore often leads to transient asphyxia, which decreases the oxygen capacities, and in turn, results in body movements during sleep. CRFs, compared with HMMs, are better suited for sequences that have long interdependencies and therefore may have better performance in our applications.

6.1 Building up the Detection Model

Fig. 9 gives the structure of CRF model we used. The shaded nodes (Y_1, \dots, Y_{t-1}, Y_t) indicate the hidden sleep stage variables during sleep. $Y_t \in \{\text{light sleep, deep sleep, REM}\}$ is an output of CRF model, which represents the sleep stage in the detection phase t . The unshaded node $\vec{X} = \{X_1, \dots, X_{t-1}, X_t\}$ denotes the observable sleep-related features occurred in the sleep process. $X_t = \{N_B(t), N_A(t), N_I(t), N_D(t), N_P\}$ represents the feature vector at detection phase t . The explanation of each item, which is the input of model, is listed as follows. $N_B(t)$: the number of occurrences of micro body movement and macro body movement during the detection phase t . $N_A(t)$: the number of occurrences of sleep-related acoustic events during the detection phase t . $N_I(t)$: the illumination condition during the detection phase t . $N_D(t)$: sleep duration. N_P : age of user.

The factor $\psi_t(Y_{t-1}, Y_t, \vec{X}, t)$ of CRF shown as black boxes in Fig. 9, returns a positive real valued number that represents the compatibility of the observable feature vector \vec{X} to the current hidden sleep stage Y_t , and the dependence between Y_{t-1} and Y_t . The standard log-linear form of the factor is listed as Equation 1,

$$\psi_t(Y_{t-1}, Y_t, \vec{X}, t) = \exp\left(\sum_{i=1}^m \lambda_i f_i(Y_{t-1}, Y_t, \vec{X}, t)\right) \quad (1)$$

where λ_i is the parameter used to indicate the real-valued weight of feature functions in CRF. f_i is a positive real-valued feature function used for characterizing its arguments, which is set manually according to our training data. m is the number of feature functions. Assuming the length of the observation sequence is $n + 1$, the conditional distribution $p(\vec{Y}|\vec{X})$ based on the factor $\psi_t(Y_{t-1}, Y_t, \vec{X}, t)$ is listed as follows,

$$p(\vec{Y}|\vec{X}) = \frac{1}{Z_{\vec{X}}} \exp\left(\sum_{t=1}^n \sum_{i=1}^m \lambda_i f_i(Y_{t-1}, Y_t, \vec{X}, t)\right) \quad (2)$$

where $Z_{\vec{X}}(\vec{X})$ is a normalized factor defined in [24] to normalize probability $p(\vec{Y}|\vec{X})$ in $[0,1]$.

Given the training data, we can estimate the weight λ_i of each feature function f_i by maximizing the conditional log-likelihood of the labelled sequences [27]. For fast training, the parameter estimation is based on the limited-memory BFGS [28]. CRF leverages Viterbi Algorithm [29] to predict the current hidden sleep stage Y_t according to the feature functions and their weights.

7 EXPERIMENTS

In this section, we detail the system implementation and evaluation results.

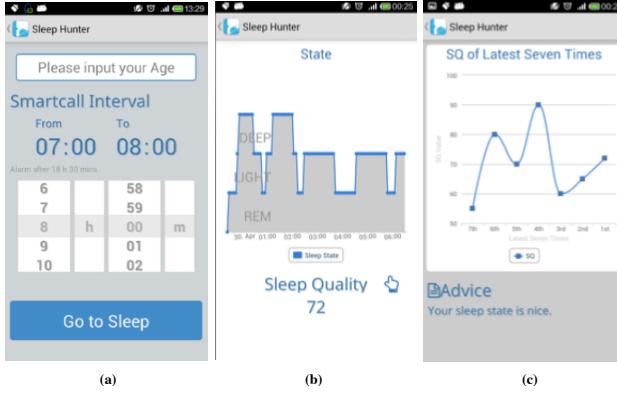


Fig. 10. User Interface of Sleep Hunter. (a) The Screen of Personal Information Collection and Smart Call Setting. (b) The Screen of Sleep Stages Tracking. (c) The Screen of Sleep Score Recording.

7.1 Prototype Implementation

We validate the architecture and algorithms of Sleep Hunter through a prototype implemented on Android OS platform. The current version is a two-layer construction which consists of around 5000 lines of code. The first layer is made up of five detection modules to monitor acoustic event, body movement, illumination condition, sleep duration and personal factor respectively. The second layer is a CRF module, which is used to predict the sleep stages. Moreover, the result of the sleep stages serves the applications of Sleep Quality Report and Smart Call Service. The sample rates of accelerometer, microphone and light sensor are set to be 100 Hz, 16KHz and 100 Hz respectively to prevent the distortion.

The total size of the application file is about 1.7 MB and occupies around 3.1 MB storage after it is launched. The displaying of user interface and each module of background processing functions are handled with multithreading to ensure the operation efficiency.

The screen shots are shown in Fig. 10. The screen (a) provides the methods for users to input their ages and set the wake up period. The screen (b) illustrates a fine grained sleep stage conditions during the night. Users could learn their sleep condition concretely by scanning it. The screen (c) provides a historical records of the sleep score in latest seven days, which helps the user have a general view of his/her sleep quality.

The application is easy to operate. Before sleep, the user only needs to initiate the application and put the Smartphone besides his/her pillow. Alternatively, the user could set the period of morning call when he/she wants to wake up. When Sleep Hunter detects a suitable time, it will sing a soft song to wake the sleeper up. The user could scan his/her sleep stages of the night and check the sleep score after getting up.

TABLE 3
Information of experimental smartphones

	OS Platform	RAM	CPU	Battery
Galaxy S4	Android OS 4.2.2	2GB	quad-core 1.638 GHz	2600mAh
Samsung Note2	Android OS 4.1	2GB	quad-core 1.638 GHz	3100mAh
HTC G14	Android OS 2.3	2GB	dual-core 1.228 GHz	1520mAh
MIUI 2SC	Android OS 2.3	2GB	quad-core 1.700 GHz	2000mAh

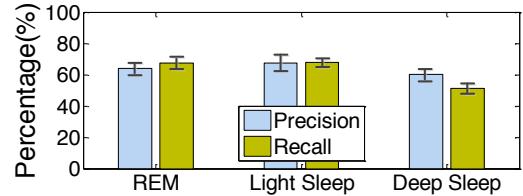


Fig. 11. The results of cross validation

7.2 Experiment setups

To give a comprehensive evaluation, we install Sleep Hunter on four different types of smartphones including: Samsung Galaxy S4 I9508, Samsung Note 2, HTC G14, and MIUI 2SC. All these smartphones are equipped with necessary sensors for sleep stage detection. The hardware configurations of these phones are detailed in Table 3. As Sleep Hunter is independent of platforms, we envision Sleep Hunter to be easily extended to other mobile operating systems like WP8 and iOS.

7.3 Training Data

We collect 90 sets of nocturnal sleep data as the training data to learn the sleep-related features by the CRF model. A total of 45 volunteers from 10 years old to 60 years old participate in the experiments and each of them contributes 2 sets of nocturnal sleep data. We divide these volunteers' ages into 5 stages by every 10 years and there are 9 participants in each stage. During experiments, these volunteers sleep alone in a quiet room and each of them sleeps at least 6 hours within a general period from 22:00 to 9:00 the next day. Moreover, every participant wears Zeo and runs Sleep Hunter in his/her smartphone simultaneously during sleep. The smartphone is placed beside the participant's head.

Considering that there is no absolute ground truth to detect sleep stage and the operations of other professional medical equipment are complicated, we leverage the result of Zeo, which is based on EEG, as the ground truth. Though Zeo is not a professional medical sleep monitor, its accuracy is around 75% [30], which is comparable to the accuracy (about 80%) of those polysomnography-based research works [1], [31], [32]. Therefore, it is reasonable to regard the results of Zeo as our ground truth to train the CRF model and measure the detection performance of Sleep Hunter.

Actually, Zeo detects four sleep stages: wake, REM, light sleep and deep sleep. Since both physiological properties and physical activities of wake and REM are similar, and the duration of wake rarely occurs in the sleep process based on our experiments, we regard wake stage as REM in Sleep Hunter, which makes little negative influence on the sleep stage detection.

Fig. 11 shows the 10-fold cross-validation [19] based on the training data. For REM and light sleep, the average precisions and

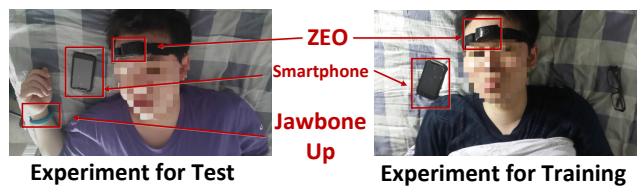


Fig. 12. Experiment Cases.

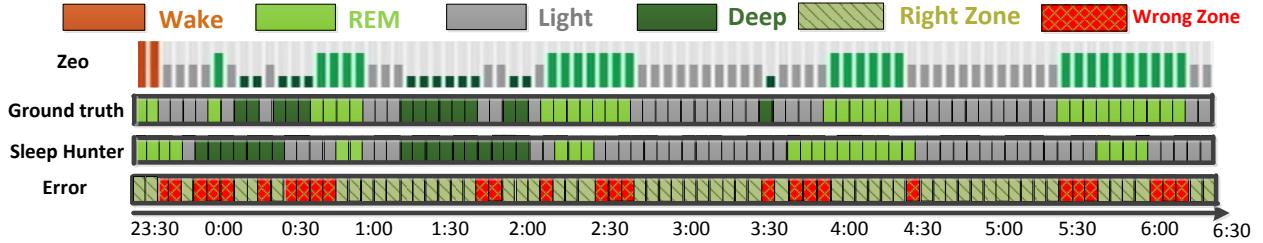


Fig. 13. Sleep stage tracking of one user during a night

recalls lie above 60%. For deep sleep, its precision is around 63% and its recall is about 52%. Moreover, the standard deviations of precision and recall of each stage are less than 5%, demonstrating the performance of CRF is stable and robust.

7.4 Testing Data

The testing data is collected over 30 sets of nocturnal sleep data from 15 volunteers with different genders and careers across various age stages as those of 45 training participants. In each age stage, we collect sleep data from 3 volunteers, with each contributing 2 sets of nocturnal sleep data. The sleep periods are also consistent with those of participants in the training data. To compare the detection performance of Sleep Hunter with other actigraphy-based products, apart from wearing Zeo and placing a smartphone installed Sleep Hunter beside his/her head as the 45 training participants, each volunteer is also asked to wear Jawbone Up [7] on his/her wrist and installs Sleep As Android [8] in the smartphone. Jawbone Up and Sleep As Android, the two representative and widely used products for sleep stage detection, run the whole sleep process with Sleep Hunter. Since Jawbone Up is a comfortable and effortless bracelet [33], its small size and light weight make it hard to be felt by sleepers during sleep. Moreover, Sleep As Android is an app running on smartphone. The sleep environments of the 15 testing participants, therefore, are similar to those of the 45 training participants. Fig. 12 illustrates the experimental cases for testing and training respectively.

7.5 Micro-benchmarks

This section presents a set of micro-benchmarks that evaluate the performance of sleep stage detection. We also evaluate the effectiveness of feature and model in Sleep Hunter.

7.5.1 Performance of Sleep Stage Detection

We measure the sleep stage detection performance of Sleep Hunter based on the testing dataset. The overall detection performance is shown in Fig. 14. The percentages in black blocks point to the precisions and recalls of the three types of sleep stages and the overall detection accuracy. The result of light sleep, whose precision reaches 66.74% and recall arrives at 68.71%, shows the

outstanding detection performance. REM also exhibits satisfying detection performance. Its precision arrives at 63.67% and recall reaches 66.11%. The precision and recall of the deep sleep stage seem slightly weaker than those of the other stages. The values of these two indices are 60.00% and 50.73% respectively. We can explain this phenomenon by the values in the confusion matrix in Fig. 14. The values in white and grey blocks are the corresponding numbers of the three types of sleep stages in our testing data. We can learn the total amount of deep sleep is much less than REM and light sleep. It shows that the occurrence of deep sleep stage in the entire sleep process is much less than those of the other two sleep stages. Since the deep sleep happens rarely, it is harder for the classifier to learn a competitive result of this sleep stage than those of others. The values in grey blocks, however, still demonstrate that most sleep stages are detected accurately. The detection accuracy of system runs up to 64.55%, showing the overall detection performance of Sleep Hunter is satisfying.

Moreover, we also evaluate the accumulative time error of each sleep stage detected by Sleep Hunter over a whole sleep process. Fig. 15 summarizes this kind of time error accumulated over the entire 6-9 hours of sleep based on our testing dataset. We can learn that 80% accumulative time errors of deep sleep stages stay under 45 minutes. The median error of this sleep stage is only around 25 minutes. Similarly, we also see that 80% accumulative time errors of REM are under 60 minutes and its median accumulative time error reaches about 30 minutes. The upper bound of the 80% accumulative time errors of light sleep arrives at around 75 minutes, which is the longest time in the three sleep stages. Its median error is around 50 minutes. Even though the precision and recall of deep sleep are little weaker than those of the others, its accumulative time error is the lowest in the sleep stages. It can also be explained that there are less occurrences of deep sleep than those of light sleep and REM during sleep. In Fig. 15, most accumulative time errors of the three sleep stages are shorter than one hour. Since the sleep duration lasts for 6 hours to 9 hours in the testing experiments, the fact of less than one hour accumulative time error demonstrates that the detection performance of Sleep Hunter is outstanding.

Fig. 13 shows a user sleep tracking instance in our testing dataset. We carefully compared the detection results of Sleep

Ground Truth	Predictions			Recall
	REM	Light Sleep	Deep Sleep	
REM	538	206	39	68.71%
Light Sleep	246	630	77	66.11%
Deep Sleep	61	108	174	50.73%
	63.67%	66.74%	60.00%	64.55%
	Precision	Accuracy		

Fig. 14. Performance of sleep stage detection

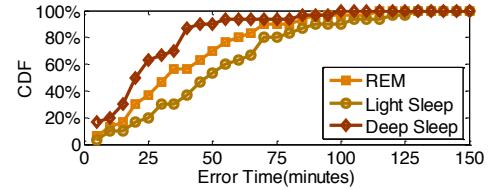


Fig. 15. Accumulative time error

TABLE 4
Evaluation of sleep-related features

Features	REM		Light Sleep		Deep Sleep	
	Precision	Recall	Precision	Recall	Precision	Recall
BM	39.62%	34.91%	37.84%	47.11%	30.12%	28.27%
BM+AE	45.41%	39.67%	47.83%	49.31%	38.34%	33.27%
BM+AE+IC	46.13%	41.81%	49.10%	52.27%	42.91%	35.84%
BM+AE+IC+SD	60.89%	67.99%	63.36%	59.15%	57.96%	46.53%
BM+AE+IC+SD+PF	63.67%	68.71%	66.74%	66.11%	60.00%	50.73%

Hunter with those of Zeo, which we regard as the ground truth. The sleep stage detection errors are labelled as the wrong zone in the last boxed line. The right zone indicates the right detection of Sleep Hunter. Since we regard wake stage as REM in Sleep Hunter, the wake stages detected by Zeo are labelled as REM in the ground truth.

In Fig. 13, we can observe that the detection errors occur frequently in the following two cases: one is during the transition of two different sleep stages, while the other is in the sleep stages with a short lifespan (*i.e.*, less than 10 minutes). In the former case, as physical activities of sleepers do not occur immediately when a new sleep stage begins, Sleep Hunter may fail to realize the sleep transitions and make error predictions accordingly. For the sleep stages whose durations are short, physical features may not be displayed obviously, which cannot be sensed by Sleep Hunter and thus induces mistakes. Actually, this phenomenon is also consistent with the results of the testing experiments over 30 sets of nocturnal sleep data, which could be seen as the main sources of error in Sleep Hunter.

From Fig. 13, we can see that although Sleep Hunter makes misjudgement occasionally, it still obtains 57 right zones accounting for 285 minutes, which occupies nearly 70% lifespan of the sleep process.

7.5.2 Evaluation on Features

Table 4 exhibits the trend of detection performance by adding each feature incrementally. The contributions of features for classification are reliant on the design of feature functions in CRF, which predict hidden variables by incorporating observable features together. Apparently, with the increasing number of features, the precision and recall of each sleep stage of the CRF model improves, demonstrating the rationality of our configurations of the feature functions and selected features.

7.5.3 Model Comparison

In order to justify the advantages of CRFs over HMMs in Sleep Hunter, we compare the detection accuracy of these two models in Fig. 16. We report the average detection results, including detection precision and recall in different sleep stages. From the Fig. 16, we can learn that CRFs outperforms HMMs in terms of

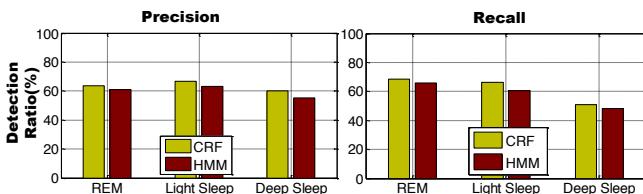


Fig. 16. Detection precision and recall of HMM and CRF in different sleep stages

TABLE 5
CPU share and processing time

Phones	Body Movement	Acoustic Events	Light	Duration	CRF	Total	CPU Share
Galaxy S4	12.72s	77.75s	1.61s	0.69s	1.55s	79.00s	6%
Note2	12.21s	77.87s	1.87s	0.58s	1.63s	80.50s	5%
HTC G14	14.78s	80.12s	2.37s	1.02s	2.81s	82.93s	9%
MIUI 2SC	12.31s	78.19s	1.58s	0.59s	2.74s	80.90s	8%

the mean precision and mean recall over all the three sleep stages, which demonstrates the rationale of CRFs in Sleep Hunter.

7.6 Macro-benchmarks

This section presents a set of macro-benchmarks that compare the performance of Sleep Hunter with other works.

7.6.1 Performance Comparison

To our best knowledge, there is no obvious baseline for the sleep stage detection performance in actigraphy-based work, and thus we compare the detection performance of Sleep Hunter with Jawbone Up and Sleep As Android, two representative actigraphy-based products monitoring sleep stage and having been widely used in the market. Since these two devices only detect the light sleep stage and the deep sleep stage, we compare these two with those of Zeo.

The average values of precision and recall of these products are calculated based on our testing dataset, which are shown in Fig. 17. Even though Sleep Hunter detects one more sleep stage than the other two products, its detection ability is still much better than the other two products, which clearly demonstrates that Sleep Hunter is superior to the existing actigraphy-based detection systems of sleep stage.

7.7 System Overhead

7.7.1 System Delay and CPU Share

In this section, we leverage four types of experimental smartphones to measure the system delay and the CPU share of Sleep Hunter. Since Sleep Hunter predicts the sleep stage every 5 minutes, it makes classification by the data recorded from each submodule in the last 5 minutes while it is running. Accordingly, the system delay is determined by the time cost to deal with the data in the last 5 minutes. As the personal factor submodule collects information from the user's registration, we mainly analyze the time consumptions of other four submodules and the whole system. Table 5 illustrates the average delay of the corresponding submodules. The system delays for the smartphones range from 79s to 82.93s. Specifically, the acoustic event detection module occupies the major part of time consumption. Such cost is caused by parsing acoustic primitive data and computing its frequency spectrum. Given these submodules in the first layer run in parallel, the system delay of

Device	Stage	Light Sleep		Deep Sleep	
		Precision	Recall	Precision	Recall
Sleep Hunter	Light Sleep	66.74%	66.11%	60.00%	50.73%
Jawbone UP	Light Sleep	37.74%	65.14%	34.62%	29.03%
Sleep As Android	Light Sleep	25.71%	32.14%	36.36%	49.61%

Fig. 17. Performance comparison

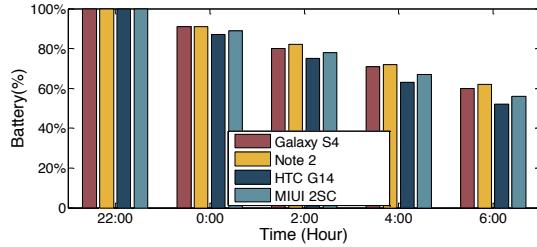


Fig. 18. Battery lifetime trace

Sleep Hunter equals the largest time consumption of submodules in the first layer plus the time cost of the CRF model in the second layer. Since the sleep stage hardly changes in a short time (*i.e.*, less than 2 minutes) and the smart call service could be provided as long as it lies in the wake-up period that lasts one hour in Sleep Hunter, the system delay around 80s makes little negative influence on the performance of our system.

Furthermore, we can observe that the CPU shares of Sleep Hunter with 4 types of smartphones in Table 5. The CPU share for G14 is little larger than those of other three smartphones, which may be caused by the different properties of CPU. Despite such a difference, the CPU share of Sleep Hunter stays stable from 5% to 9%, indicating that Sleep Hunter occupies negligible CPU resource for daily use.

7.7.2 Battery life

Considering the usage period of Sleep Hunter lasts for at least 5 hours, the battery consumption directly impacts its applicability. To record the battery conditions, a volunteer is asked to run Sleep Hunter on our four types of smartphones simultaneously from 23:00 to 7:00 next day. Since Sleep Hunter is a service that could run in backstage, we lock the screen during this process and shut down all the applications except Sleep Hunter and a battery tracing application, which is installed in each smartphone to record the rest battery storage every two hours. Fig. 18 illustrates the battery storage condition of the experimental smartphones. Based on the bar chart, each smartphone consumes around 10% on average every two hours. With time passing by, the power consumption increases gradually, and finally ends up at the level about 60%. Fig. 19 shows the distribution of battery consumption during the experimental process in detail. Sleep Hunter only consumes less than 40% of the whole energy cost, most power is consumed by Android OS. Namely, Sleep Hunter only costs less than 4% power every two hours. Accordingly, the energy consumption of Sleep Hunter is not significant and the battery storage of most smartphones could be affordable for users to execute it during the whole sleep process.

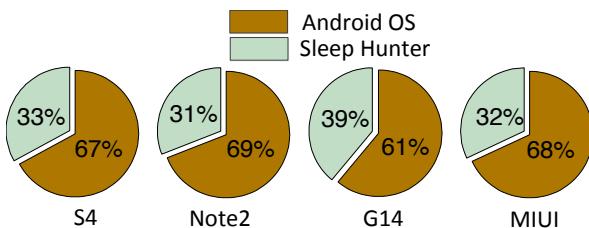


Fig. 19. Battery consumption distribution

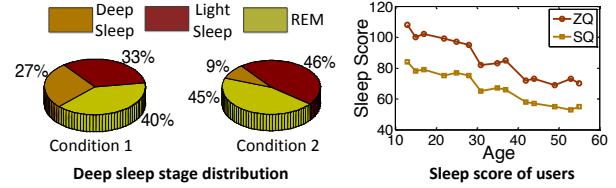


Fig. 20. Sleep quality report

7.8 Case study

7.8.1 Sleep Quality Report

Since sleep quality is determined by the percentage of different sleep stages during the whole sleep process, we can measure a user's sleep quality by Sleep Hunter. We design a sleep quality measuring method based on the evaluation approach 'ZQ' of Zeo [34]. 'SQ' is a score shown in Eq. 3 for measuring the sleep quality of users in Sleep Hunter. Its structure and parameters are designed according to 'ZQ', where *REM*, *Deep* and *Light* represent their durations (minutes) in a sleep process. The range of 'SQ' is from 50 to 100.

$$SQ = \left[\frac{(REM \times 0.5 + Light \times 0.75 + Deep) \times 100}{REM + Light + Deep} \right] \quad (3)$$

The right figure in Fig. 20 illustrates the average 'SQ' and 'ZQ' of 2 sets of nocturnal sleep data for the 15 users in the testing dataset. Though differences exist in the structures of 'SQ' and 'ZQ', the general variance trends of these two indices are correlated with the growth of age.

We analyze the sleep quality by 'SQ' in the following parts.

We observed that the occurrence of deep sleep is usually affected by its starting time and duration. According to our observation of the testing dataset, we figure out that if the deep sleep stage occurs in the first hour after the user falls asleep and lasts more than 10 minutes, the possibility of its recurrence is high and the sleep quality is usually appealing. Otherwise, the percentage of deep sleep would be low and the sleep quality is disappointing. The left figure in Fig. 20 shows the distribution of sleep stages in the two different conditions based on our testing dataset. Condition 1 stands for the distribution of sleep stages in a sleep process when the deep sleep stage occurs in the first hour and lasts for more than 10 minutes. Condition 2 represents the opposite case. Clearly, the percentage of deep sleep in Condition 1 is significantly larger than that in Condition 2. Based on Eq. 3, 'SQ' in Condition 2 is usually lower than that in Condition 1. It reflects the fact that if a user cannot fall into deep sleep within one hour or remains in this stage in a short time, the user's sleep quality will be low. Based on our knowledge of medical research [35], if the brain waves of a person remain rapid for a long time, it is hard for him/her to transform the fast brain waves to slow brain waves. It is because fast brain waves often make individuals

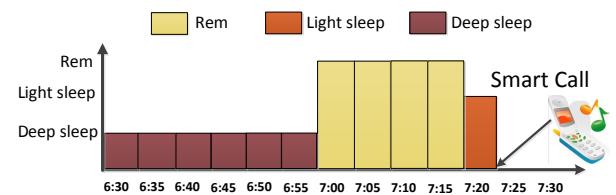


Fig. 21. Smart call service

TABLE 6
Performance of smart call service

Case	Explanation	Sleep Hunter	Sleep Android	As
Case 1	There are actual some light sleep stages in the preset wake-up period and the app detects these light sleep stages rightly.	86.67%	42.33%	
Case 2	There are actual some light sleep stages in the preset wake-up period but the app does not detect these light sleep stages in the right time or does not find any light sleep stages at all.	6.67%	36.33%	
Case 3	There are no light sleep stage in the preset wake-up period actually but the app mistakes some other sleep stages as light sleep.	6.67%	21.33%	
Case 4	There are no light sleep stage in the preset wake-up period actually and the app also does not find them.	0%	0%	

excited, the long period excitement would disturb their normal rest habits.

As a point of interest, we also make an analysis on the relationship between age and sleep quality. From the right figure in Fig. 20, we observe that ‘SQ’ decreases with the increase of the age generally. More specifically, we find two rapid drops in Fig. 20. The first sudden drop occurs at the age of around 30. ‘SQ’ falls from the score around 75 to the score about 65. According to our observation, we think the 10-point drop stems from users’ work and family stress. At this age stage, most individuals have their own families to take care of and carry more workload than they used to do in school. High pressure, therefore, leads to the poor sleep quality. The second sudden drop occurs at the age of 40. ‘SQ’ slips from the score above 65 to the score near 55. It is likely that many users are approaching menopause during this age stage, when their physiological functions are experiencing changes. Some climacteric syndromes cause adverse impacts on the users and then decrease their sleep qualities. Given that the gender and career of the 15 participants in the testing dataset are different, our sleep quality report, therefore, is convincing and representative to some extent. We note that all conclusions are speculative based on our small sample size, but our observations are consistent with other sleep research studies [36], [37], [38].

According to the observations above, we could make two recommendations for users. 1) Since the beginning of sleep highly impacts the quality of the whole sleep process, people may benefit from light activities for relaxation before going to bed. For example, listening to some soft music, making several deep breaths or taking a warm bath can greatly help people relax before sleep. 2) People in their 30s are largely threatened by the pressure from work. Individuals at this age may benefit from some adjustments to achieve a greater balance between work and life. Seeking advice from more experienced people and cultivating a positive attitude may help them release some pressure off. For those at their 40s, health deals, moderate but habitual exercise could help them improve sleep quality.

7.8.2 Smart Call

Sleep Hunter can also provide smart wake-up services based upon the sleep stage detection, since individuals feel much more refreshed when they are woken up in light sleep [35]. Users are allowed to set a one-hour period when they want to be woken, and then Sleep Hunter wakes the user up when it detects the current sleep stage is light sleep during the preset period. If Sleep Hunter detects no light sleep for the duration, it wakes users up at the end of this period. Fig. 21 depicts an illustration of a real smart call case in our testing experiments. The user sets the wake-up

period from 6:30 a.m. to 7:30 a.m., and then Sleep Hunter will automatically wake the user up based on the detection result of sleep stages in the period. When Sleep Hunter detects that the user stays in a light sleep at 7:25 a.m., Sleep Hunter sings a soft song to wake him/her up.

We evaluate the smart call performance of Sleep Hunter and Sleep As Android by the users in our testing experiments. Each user is asked to select a wake-up period before going to bed. Table 6 gives the occurrence rates of these two apps under four different cases. Compared with Sleep As Android, the larger occurrence rate in Case 1 and smaller occurrence rates in Case 2 and Case 3 show that the smart call service of Sleep Hunter performs much better. Moreover, 86.67% occurrence rate of Case 1 demonstrates that Sleep Hunter is able to provide the smart call service accurately. There are mainly two reasons behind this. The first one is the remarkable sleep stage detection performance of Sleep Hunter. The second one is that more light sleep stages and REM occur and less deep sleep stages happen during the latter period of sleep. Since Sleep Hunter has shown excellent detection performances of light sleep and REM, smart call service, therefore, exhibits outstanding.

8 DISCUSSION

The sleep stage detection accuracy of Sleep Hunter is 64.55%, which seems hardly comparable to those of polysomnography-based devices. Therefore, Sleep Hunter cannot take place of those professional medical devices for the high-accurate sleep detection. However, Sleep Hunter, as an app, pays more attention to providing a pervasive and less intrusive sleep service using commodity smartphones. By Sleep Hunter, users can be informed of their general sleep condition conveniently and be woken up properly.

9 RELATED WORK

Since sleep quality has being a crucial issue of each individual, and the insufficient sleep would lead to numerous diseases including diabetes, endocrine dyscrasia, depression, stomach problem, immunity decline [39], [40], [41], [42], [43], [44], a diverse range of research works [1], [45], [46], [47], [48] have been focused on it from both medical and consumer usage scenarios.

Below, we summarize the state-of-the-art research areas related to our work from the following two categories.

Polysomnography-based work The polysomnography is regarded as the gold standard to measure the sleep quality in clinical field. It usually relies on some certain biomedical signals such as brain wave, muscle tone and eye movement to assess the

sleep quality. For example, most medical research works on sleep quality leverage electroencephalograph (EEG) to monitor brain waves and then recognize sleep stages. In [1], [31], [32], authors extract the electroencephalograph features from polysomnography (PSG) and leverage unsupervised learning approaches to predict sleep stages.

However, these polysomnography-based work usually requires to be assisted by the professional devices. For example, Zeo [5], which is a product based on EEG, leverages a brain wave sensor built in a headband to monitor sleeper's electroencephalograph, and then the EEG recordings are sent to the user's smartphone via Bluetooth. Zhang *et al.* [49] design a real-time system to monitor the user's sleep condition. It ameliorates users' sleep qualities by exploiting a pulse oximeter to detect the pulse oxygen saturation (SPO2) of the human body during their sleep processes. Such special equipment, professional technologies and expensive cost make them impractical for people in daily use.

Compared with such works, Sleep Hunter concentrates on physical activities rather than biomedical signals. Moreover, it does not need special additional devices for detection.

Actigraphy-based work Actigraphy provides a simpler approach depends on physical activities to predict sleep quality. Compared with the polysomnography, it is much more convenient for people in daily use. With the flourish development of pervasive computing, a number of mobile devices and technologies have been leveraged to sensing human behaviors intelligently [50], [51], [52], [53], [54], [55]. Some sensor technologies also assist people to monitor sleep by Actigraphy.

iSleep [12] measures the sleep quality by recording some certain sleep-related acoustic events and evaluates it by the Pittsburgh Sleep Quality Index (PSQI) [2]. Bai *et al.* [56] predict the sleep quality by observing users' daily activities with smartphones. In [57], the authors leverage pervasive sensors to record the sleep disruptors for users. The authors in [58] monitor sleep by the RFID sensors installed with accelerometers. In [47], the authors explore a series of opportunities to support healthy sleep behaviors by ubiquitous computing. Several research works predict the sleep quality by calculate the influence from the external factors such as the daily activity, sleeping environment, location and family settings [45], [59].

Many Smartphone Apps such as Sleep As Android [8], Sleep Journal [60], YawnLog [61] and numerous consumer-oriented and wearable wrists including Jawbone Up [7], FitBit [62], YawnLog [61] and WakeMate [63] and so on predict sleep stages and measure sleep quality based on physical activities including body movement and snore simply. For example, when many body movements or acoustic events happen in a period, they regard the user stays in light sleep. On the contrary, the user is assumed to step into deep sleep as long as little body movement and ambient noise occur in the period.

Different from these works, Sleep Hunter evaluates the user's sleep quality by measuring the durations of different sleep stages in a sleep process rather than recording some certain sleep-related activities. It incorporates sleep-relative events from different perspectives and leverages a statistical model to predict the sleep stage, which provides a fine-grained detection performance of sleep stages without any assistant devices. Moreover, compared with the wearable intelligent wrists, Sleep Hunter does not require users to wear , providing a non-intrusive method for free use.

10 CONCLUSION

Recent advances in sensor technology and machine learning technique empower machine to intelligently understand human behaviors. This paper guides this opportunity into an application that automatically detects sleep stage transitions of sleepers for sleep quality monitoring. The core idea is to leverage built-in sensors on commodity phones to sense sleep-related events, and further predict the dwelling time of each sleep stage by a statistical model based on these observable events. We implement Sleep Hunter on Android platform and test it with data collected over 30 sets of nocturnal sleep data. The results show that our system achieves desirable detection accuracy.

REFERENCES

- [1] M. Längkvist, L. Karlsson, and A. Loutfi, "Sleep stage classification using unsupervised feature learning," *Advances in Artificial Neural Systems*, vol. 2012, p. 5, 2012.
- [2] J. S. Carpenter and M. A. Andrykowski, "Psychometric evaluation of the pittsburgh sleep quality index," *Journal of psychosomatic research*, vol. 45, no. 1, pp. 5–13, 1998.
- [3] J. Trinder, J. Kleiman, M. Carrington, S. Smith, S. Breen, N. Tan, and Y. Kim, "Autonomic activity during human sleep as a function of time and sleep stage," *Journal of sleep research*, vol. 10, no. 4, pp. 253–264, 2001.
- [4] "Polysomnography," <http://en.wikipedia.org/wiki/Polysomnography>.
- [5] "Zeo Sleep Manager Pro," <http://www.digifit.com/Zeo/>.
- [6] "Actigraphy," <http://en.wikipedia.org/wiki/Actigraphy>.
- [7] "Jawbone Up," <https://jawbone.com/up>.
- [8] "Sleep As Android," https://play.google.com/store/apps/details?id=com.urbandroid.sleep&hl=zh_EN.
- [9] W. Webb and H. Agnew, "Sleep stage characteristics of long and short sleepers." *Science*, 1970.
- [10] S. Ancoli-Israel, R. Cole, C. Alessi, M. Chambers, W. Moorcroft, and C. Pollak, "The role of actigraphy in the study of sleep and circadian rhythms. american academy of sleep medicine review paper," *Sleep*, vol. 26, no. 3, pp. 342–392, 2003.
- [11] G. Jean-Louis, D. F. Kripke, S. Ancoli-Israel, M. R. Klauber, and R. S. Sepulveda, "Sleep duration, illumination, and activity patterns in a population sample: effects of gender and ethnicity," *Biological psychiatry*, vol. 47, no. 10, pp. 921–927, 2000.
- [12] T. Hao, G. Xing, and G. Zhou, "isleep: unobtrusive sleep quality monitoring using smartphones," in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2013, p. 4.
- [13] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. T. Campbell, "Soundsense: scalable sound sensing for people-centric applications on mobile phones," in *Proceedings of the 7th international conference on Mobile systems, applications, and services*. ACM, 2009, pp. 165–178.
- [14] D. Li, I. K. Sethi, N. Dimitrova, and T. McGee, "Classification of general audio data for content-based retrieval," *Pattern recognition letters*, vol. 22, no. 5, pp. 533–544, 2001.
- [15] J. Saunders, "Real-time discrimination of broadcast speech/music," in *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96 Vol 2. Conference Proceedings., 1996 IEEE International Conference on*, vol. 2, no. 2. IEEE, 1996, pp. 993–996.
- [16] B. Kedem, "Spectral analysis and discrimination by zero-crossings," *Proceedings of the IEEE*, vol. 74, no. 11, pp. 1477–1493, 1986.
- [17] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator," in *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, vol. 2. IEEE, 1997, pp. 1331–1334.
- [18] D. Boswell, "Introduction to support vector machines," 2002.
- [19] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *IJCAI*, vol. 14, no. 2, 1995, pp. 1137–1145.
- [20] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *Advances in information retrieval*. Springer, 2005, pp. 345–359.
- [21] K. L. Lichstein, H. H. Durrence, B. W. Riedel, D. J. Taylor, and A. J. Bush, *Epidemiology of sleep: Age, gender, and ethnicity*. Psychology Press, 2004.

- [22] M. M. Ohayon, M. A. Carskadon, C. Guilleminault, and M. V. Vitiello, "Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan," *SLEEP-NEW YORK THEN WESTCHESTER-*, vol. 27, pp. 1255–1274, 2004.
- [23] M. A. Carskadon, W. C. Dement *et al.*, "Normal human sleep: an overview," *Principles and practice of sleep medicine*, vol. 4, pp. 13–23, 2000.
- [24] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.
- [25] E. Fosler-Lussier, Y. He, P. Jyothi, and R. Prabhavalkar, "Conditional random fields in speech, audio, and language processing," *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1054–1075, 2013.
- [26] E. C. Larson, M. Goel, G. Boriello, S. Heltshe, M. Rosenfeld, and S. N. Patel, "Spirosmart: Using a microphone to measure lung function on a mobile phone," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 280–289.
- [27] I. J. Myung, "Tutorial on maximum likelihood estimation," *Journal of Mathematical Psychology*, vol. 47, no. 1, pp. 90–100, 2003.
- [28] A. Skajaa, "Limited memory bfgs for nonsmooth optimization," *Master's thesis*, 2010.
- [29] G. D. Forney Jr., "The viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
- [30] J. R. Shambroom, S. E. Fabregas, and J. Johnstone, "Validation of an automated wireless system to monitor sleep in healthy adults," *Journal of Sleep Research*, vol. 21, no. 2, pp. 221–230, 2012.
- [31] E. Oropesa, H. L. Cycon, and M. Jobert, "Sleep stage classification using wavelet transform and neural network," *International computer science institute*, 1999.
- [32] F. Ebrahimi, M. Mikaeili, E. Estrada, and H. Nazeran, "Automatic sleep stage classification based on eeg signals by using neural networks and wavelet packet coefficients," in *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*. IEEE, 2008, pp. 1151–1154.
- [33] "Jawbone Up review: An easy-to-wear and insightful fitness pal," <http://www.cnet.com/products/jawbone-up/>.
- [34] "Whats Your Bulletproof ZQ Score? The Zeo Hack Every Sleep Hacker Needs To Know," <http://www.bulletproofexec.com/zeo-hack/>.
- [35] M. Beloved, *Meditation Pictorial*. Michael Beloved, 2011.
- [36] Y. Q. J. C. Marquie, J. Foret, "Effects of age, working hours, and job content on sleep: a pilot study," *Experimental aging research*, vol. 25, no. 4, pp. 421–427, 1999.
- [37] P. Polo-Kantola, R. Erkkola, K. Irljala, H. Helenius, S. Pullinen, and O. Polo, "Climacteric symptoms and sleep quality," *Climacteric*, vol. 2, no. 4, pp. 293–294, 1999.
- [38] P. Eichling and J. Sahni, "Menopause related sleep disorders," *J Clin Sleep Med*, vol. 1, no. 3, pp. 291–300, 2005.
- [39] A. N. Vgontzas, D. Liao, S. Pejovic, S. Calhoun, M. Karataraki, and E. O. Bixler, "Insomnia with objective short sleep duration is associated with type 2 diabetes a population-based study," *Diabetes Care*, vol. 32, no. 11, pp. 1980–1985, 2009.
- [40] G. Livingston, B. Blizzard, and A. Mann, "Does sleep disturbance predict depression in elderly people? a study in inner london," *British Journal of General Practice*, vol. 43, no. 376, pp. 445–448, 1993.
- [41] D. J. Gottlieb, N. M. Punjabi, A. B. Newman, H. E. Resnick, S. Redline, C. M. Baldwin, and F. J. Nieto, "Association of sleep time with diabetes mellitus and impaired glucose tolerance," *Archives of internal medicine*, vol. 165, no. 8, pp. 863–867, 2005.
- [42] C. L. Drake, T. Roehrs, and T. Roth, "Insomnia causes, consequences, and therapeutics: an overview," *Depression and anxiety*, vol. 18, no. 4, pp. 163–176, 2003.
- [43] D. F. Kripke, L. Garfinkel, D. L. Wingard, M. R. Klauber, and M. R. Marler, "Mortality associated with sleep duration and insomnia," *Archives of general psychiatry*, vol. 59, no. 2, pp. 131–136, 2002.
- [44] N. T. Ayas, D. P. White, J. E. Manson, M. J. Stampfer, F. E. Speizer, A. Malhotra, and F. B. Hu, "A prospective study of sleep duration and coronary heart disease in women," *Archives of Internal Medicine*, vol. 163, no. 2, pp. 205–209, 2003.
- [45] Z. Chen, M. Lin, F. Chen, N. D. Lane, G. Cardone, R. Wang, T. Li, Y. Chen, T. Choudhury, and A. T. Campbell, "Unobtrusive sleep monitoring using smartphones," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on*. IEEE, 2013, pp. 145–152.
- [46] J. S. Bauer, S. Consolvo, B. Greenstein, J. Schooler, E. Wu, N. F. Watson, and J. Kientz, "Shuteye: encouraging awareness of healthy sleep recommendations with a mobile, peripheral display," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2012, pp. 1401–1410.
- [47] E. K. Choe, S. Consolvo, N. F. Watson, and J. A. Kientz, "Opportunities for computing technologies to support healthy sleep behaviors," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 3053–3062.
- [48] S. Kim, J. A. Kientz, S. N. Patel, and G. D. Abowd, "Are you sleeping?: sharing portrayed sleeping status within a social network," in *Proceedings of the 2008 ACM conference on Computer supported cooperative work*. ACM, 2008, pp. 619–628.
- [49] J. Zhang, Q. Zhang, Y. Wang, and C. Qiu, "A real-time auto-adjustable smart pillow system for sleep apnea detection and treatment," in *Proceedings of the 12th international conference on Information processing in sensor networks*. ACM, 2013, pp. 179–190.
- [50] S. Chen, A. Pande, and P. Mohapatra, "Sensor-assisted facial recognition: An enhanced bio-metric authentication system for smartphones," in *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. ACM, 2014, pp. 109–122.
- [51] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, "Moodscope: building a mood sensor from smartphone usage patterns," in *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 2013, pp. 389–402.
- [52] N. Roy, H. Wang, and R. R. Choudhury, "I am a smartphone and i can tell my users walking direction," in *Proceedings of the 12th international conference on Mobile systems, applications, and services*. ACM, 2014.
- [53] H. L. Chu, V. Raman, J. Shen, A. Kansal, V. Bahl, and R. R. Choudhury, "I am a smartphone and i know my user is driving." in *COMSNETS*, 2014, pp. 1–8.
- [54] Y. Wang, J. Yang, Y. Chen, H. Liu, M. Gruteser, and R. P. Martin, "Tracking human queues using single-point signal monitoring," in *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. ACM, 2014, pp. 42–54.
- [55] S. Nawaz and C. Mascolo, "Mining users' significant driving routes with low-power sensors," *arXiv preprint arXiv:1409.2264*, 2014.
- [56] Y. Bai, B. Xu, Y. Ma, G. Sun, and Y. Zhao, "Will you have a good sleep tonight?: sleep quality prediction with mobile phone," in *Proceedings of the 7th International Conference on Body Area Networks*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2012, pp. 124–130.
- [57] M. Kay, E. K. Choe, J. Shepherd, B. Greenstein, N. Watson, S. Consolvo, and J. A. Kientz, "Lullaby: a capture & access system for understanding the sleep environment," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 226–234.
- [58] E. Hoque, R. F. Dickerson, and J. A. Stankovic, "Monitoring body positions and movements during sleep using wisps," in *Wireless Health 2010*. ACM, 2010, pp. 44–53.
- [59] D. Aliakseyeu, J. Du, E. Zwartkruis-Pelgrim, and S. Subramanian, "Exploring interaction strategies in the context of sleep," in *Human-Computer Interaction-INTERACT 2011*. Springer, 2011, pp. 19–36.
- [60] "Sleep Journal App," <http://sleeptrackerapp.blogspot.com/>.
- [61] "YawnLog," <http://beta.yawnlog.com>.
- [62] "FitBit," <http://www.fitbit.com>.
- [63] "WakeMate," <http://www.wakemate.com>.