iSleep: Unobtrusive Sleep Quality Monitoring using Smartphones*

Tian Hao
Michigan State University

Guoliang Xing
Michigan State University

Gang Zhou
College of William and Mary

ABSTRACT

The quality of sleep is an important factor in maintaining a healthy life style. To date, technology has not enabled personalized, in-place sleep quality monitoring and analysis. Current sleep monitoring systems are often difficult to use and hence limited to sleep clinics, or invasive to users, e.g., requiring users to wear a device during sleep. This paper presents iSleep – a practical system to monitor an individual's sleep quality using off-the-shelf smartphone. iSleep uses the built-in microphone of the smartphone to detect the events that are closely related to sleep quality, including body movement, couch and snore, and infers quantitative measures of sleep quality. iSleep adopts a lightweight decision-tree-based algorithm to classify various events based on carefully selected acoustic features, and tracks the dynamic ambient noise characteristics to improve the robustness of classification. We have evaluated iSleep based on the experiment that involves 7 participants and total 51 nights of sleep, as well the data collected from real iSleep users. Our results show that iSleep achieves consistently above 90% accuracy for event classification in a variety of different settings. By providing a fine-grained sleep profile that depicts details of sleep-related events, iSleep allows the user to track the sleep efficiency over time and relate irregular sleep patterns to possible causes.

1. INTRODUCTION

Sleep plays an important role in our overall health. Having insufficient amount of sleep can easily cause fatigue and lack of concentration during the day. Besides the amount of sleep, the quality of sleep is also an important factor in maintaining a healthy life style. Clinical studies show that sleep is related to many serious diseases including diabetes, obesity and depression [16] [27].

*This work is supported in part by the NSF under grant CNS-0954039 (CAREER), CNS-1250180 and ECCS-0901437. This study is approved by the Institutional Review Board (IRB) of Michigan State University.

SenSys'13, November 11–15, 2013, Rome, Italy.
Copyright 2013 ACM 978-1-4503-1169-4 ...$15.00.

To date, technology has not enabled personalized, in-place sleep quality monitoring and analysis. Polysomnography (PSG) is the primary clinical tool for sleep monitoring [13]. It can provide a quantitative profiling of sleep to diagnose sleep disorders. However, due to the need of various sensors, PSG-based sleep quality measurement is usually limited to clinical settings. Actigraphy has been studied as an inexpensive alternative to assess sleep and wakefulness based on body movement [8]. Several portable sleep assessment products are designed based on PSG or actigraphy technologies, including ZEO [7], Sleep Tracker [5] and fitbit [1]. However, they are invasive to users as they require a device to be worn by the user during sleep. A recent in-depth survey of 230 participants suggested that, although most people are interested in using technology to track their sleep quality, many are resistant to the idea of having to wear a device during sleep [14].

This paper presents iSleep – a practical system to monitor an individual's sleep quality using off-the-shelf smartphone. iSleep is very easy to use and truly unobtrusive: the user just needs to start iSleep app and place the phone somewhere close to the bed (e.g., on a night stand). iSleep uses the built-in microphone of the smartphone to detect the events that are closely related to sleep quality, including body movement, couch and snore. Based on the detected events, iSleep infers quantitative measures of sleep quality based on actigraphy and Pittsburgh Sleep Quality Index (PSQI) [12] which are two well-established scoring criteria in sleep literature. We have released an initial version of iSleep on the Google Play Store [2]. Within 6 days, iSleep was installed by more than 100 users from 9 countries on various Android devices. By providing a detailed sleeping profile, iSleep enables the user to be aware of irregular sleep patterns like restlessness caused by extensive snoring which are otherwise hard to find. Moreover, as an unobtrusive, portable, in place monitoring tool, iSleep can track sleep quality quantitatively over a long period of time, which helps healthcare provider diagnose trends related to certain diseases.

The design of iSleep faces several challenges such as highly diverse acoustic profiles of sleep from person to person and in different environments. We carefully analyze the acoustic data collected from real sleep experiments and choose several statistical acoustic features that can differentiate environment noise and various sleep-related events. To improve the robustness of detection, iSleep tracks the ambient noise characteristics and updates the noise model adaptively. Finally, iSleep adopts a lightweight decision-tree-based algorithm to classify various sleep-related events and derive quantitative sleep quality measures. We have evaluated iSleep extensively in a long-term experiment that involves 7 participants and total 51 nights of sleep, as well as using the data collected from the Android phones that downloaded and installed iSleep from Google Play Store. Our results show that iSleep achieves consistently above 90% classification accuracy for various events, across different subjects and in a variety of different sleep environments.

2. RELATED WORK

According to AASM (American Academy of Sleep Medicine), the sleep stage scoring based on polysomnography (PSG) has long been considered as the "gold standard" of sleep study [20]. A polysomnogram typically requires the recording of multiple channels including electroencephalography (EEG), electromyography (EMG), electrocardiography (ECG) or heart rate, respiratory effort, air flow, oxygen saturation and etc. [13]. The result of PSG includes a collection of indices such as sleep onset latency, total sleep time and etc, which are considered together to infer the sleep quality. Due to the need of various sensors, PSG-based sleep quality measurement is usually limited to sleep clinics.

Actigraphy has been studied as an inexpensive alternative to assess human sleep and wakefulness [8] based on the subject's body movements overnight. The basic idea is that the state of sleep and wake can be inferred from the amount of body movement during sleep [8]. Through processing the logged acceleration data, epoch-by-epoch (usually 30 second or 1 minute) sleep/wake predictions are calculated. Several algorithms [18] [29] [15] have been proposed to derive sleep quality from actigraphy. The average accuracy of predicting sleep/wake state is around 90% (reported 88% in [15] and 94-96% in [31]).

A widely used subjective sleep quality assessment method is through PSQI (Pittsburgh Sleep Quality Index) [12], which is a self-rated questionnaire to assess the sleep quality and disturbance over a long-term interval. In PSQI, a set of sleep measures are collected, including sleep latency, sleep duration, sleep disturbance and etc. PSQI has been shown useful in numerous studies [9] [11] over a variety of populations. However, the accuracy of PSQI is highly variable and is often impeded by the inaccuracy of subject's memory and perception.

Several commercial personal sleep assessment products are currently available. Watch PAT [6] detects respiratory disturbances during sleep by monitoring peripheral arterial tone (PAT). The users are required to attach a probe to their finger during sleep. ZEO [7] is a popular sleep monitoring product that infers sleep stages using three EEG sensors contained in a head band worn by the user during sleep. Several actigraphy-based products such as Sleep Tracker [5] and fitbit [1] require the user to wear the device containing accelerometer during sleep.

Recently, several research efforts aimed at developing low-cost sleep assessment systems. In [28], a wearable neck-cuff system for real-time sleep monitoring is designed based on oximetry sensor, microphone and accelerometer. Instead of directly measuring the sleep, SleepMiner [10] predicts the sleep quality based on the user's daily context information such as sound, light, postures, and positions. In [19], the body position and movements during sleep are monitored using accelerometers attached to bed mattress. A dense pressure sensitive bedsheet for sleep posture monitoring is proposed in [23]. However, these systems incur nontrivial monetary costs of hardware or professional installation.

Several Android and iOS Apps such as *Sleep as Android* [3] and *Sleep Cycle* [4] can measure sleep quality. All of them exclusively rely on the actigraphy-based methods that monitor body movements overnight using smartphones. However, sleep-related events such as cough and snore can not be reliably detected based on acceleration. For example, snore is the sound caused by the vibration of respiratory structures while sleeping due to obstructed air movement, and is not necessarily associated with body motion. Moreover, since the motion data is collected through the built-in accelerometer, the phone must be put on the bed, which not only is inconsistent with the habit of most users, but also may obstruct the individual's body movement.

iSleep leverages the existing body of work on acoustic signal processing (e.g. SoundSense [24] and StressSense [25]). However, iSleep employs novel techniques to address the challenges specific to sleep-related event classification, including highly diverse acoustic profiles of sleep.

3. SYSTEM REQUIREMENTS AND CHALLENGES

iSleep is designed to be a "sleep diary" that provides the user real-time, fine-grained feedback to their sleep quality on a daily basis [1]. Specifically, iSleep is designed to meet the following requirements: (1) Since iSleep operates over night while the user is asleep, it needs to be unobtrusive. It should minimize the burden on the user, and the user should not feel any kind of uncomfort when using the system. (2) iSleep needs to provide fine-grained measurement, such as overall sleep efficiency and the occurrences of events that may interrupt sleep, such as cough and snore. Such fine-grained sleep profiling helps the user understand what factors affect their sleep quality. (3) iSleep needs to deliver robust monitoring accuracy across different users, smartphones and sleep environments. (4) The users' privacy needs to be strictly protected. Due to the inherently private nature of sleep, any concern (or even suspension) of privacy breach may prevent the adoption of sleep monitoring technology like iSleep. For instance, the system should process the sensor samples on the fly and only keep sleep-related data (e.g., the number/loudness of snores), instead of sending any raw sensor samples to a remote server, because they may capture sensitive information such as audio of sleep talks, conversations before/after sleep and etc.

To meet these requirements, three major challenges need to be addressed in developing iSleep. First, in order to effectively monitor the sleep quality in an unobtrusive manner, iSleep samples and analyzes acoustic signals from the built-in microphone to detect sleep-related events. Therefore, the user only needs to leave the phone somewhere close to the bed (up to several meters). However, the built-in microphone of smartphone is designed for capturing close vocals, and usually has low sensitivity. Moreover, many sleep-related events only generate low-intensity sound. For example, the intensity of the sound from a roll-over movement is typically only several *dB* higher than that of ambient noise.

[1]iSleep is not designed or certified for clinical use, although the monitoring results provided by iSleep could be potentially useful for professional diagnosis of sleep-related disease such as insomnia [8].

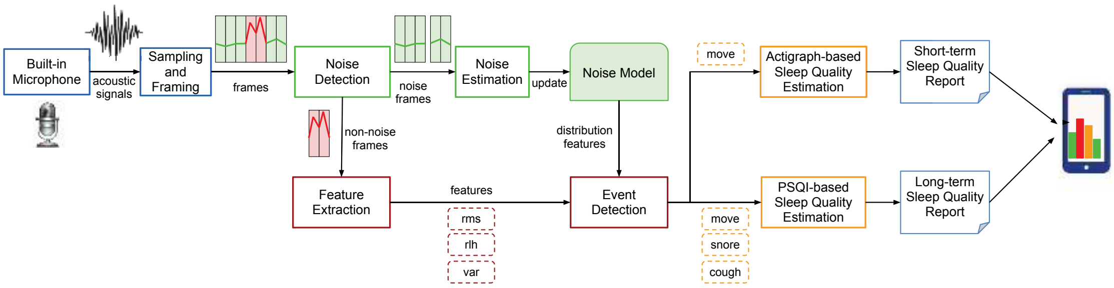


Figure 1: The architecture of iSleep system.

Second, iSleep needs to detect sleep-related events in a robust manner across different users and environments. For instance, different people likely snore in different ways in terms of sound frequency and loudness. Moreover, even the same person may snore differently due to the change of body position overnight. Noises from appliances such as fans may also have a major impact on the acoustic event detection accuracy.

Lastly, in order to preserve users' privacy, iSleep does not store or transmit raw sound samples. Instead, sound data is processed locally on the smartphone in real-time, while only the event detection results such as the number of occurrences of snore/cough/body movement are kept and shown to the user. To capture the features of various acoustic events, the microphone must be sampled at a high rate. Due to the resource constrains of smartphones, the acoustic processing algorithms must be extremely lightweight in order to process the data in real-time, while maintaining satisfactory event detection accuracy.

4. SYSTEM OVERVIEW

Keeping the above challenges in mind, we aim to build a light-weight sleep quality monitoring system that is reliable in detecting sleep-related events across different users and environments. Fig. 1 shows the architecture of the iSleep system. First, the acoustic signal is continuously sampled at the frequency of 16 kHz from the microphone, and segmented into *frames*. Second, the acoustic frames are fed to *Noise Detection*, where the system determines whether a frame only contains the sound of ambient noise. The model of ambient noise is then updated based on detected noise frames. As a result, the system is able to adapt to the changes of ambient noise. Third, acoustic features such as root mean square and variance will be extracted from the frames that potentially contain events of interest. The extracted features, along with the updated ambient noise model, are fed to the *Sleep Event Detection*, where sleep-related events such as snoring, coughing and body movement will be detected.

iSleep derives both short-term (one-night) and long-term sleep quality from sleep-related events according to two well-established sleep scoring criteria: actigraphy and Pittsburgh Sleep Quality Index (PSQI) [12]. iSleep uses actigraphy to estimate the sleep/wake states overnight and then computes a metric called *sleep efficiency*, which is the ratio of actual sleep time to total in-bed time. Compared with other quality measures such as sleep stages, sleep efficiency provides a quantitative and more intuitive feedback to users. In addition to one-night sleep efficiency, iSleep employs PSQI to estimate long-term sleep quality over multiple nights. PSQI

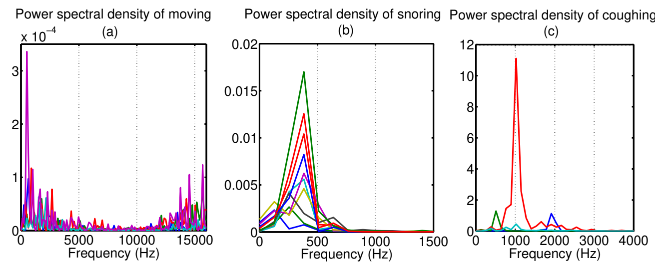


Figure 2: (a), (b) and (c) show the power spectral density of typical moving, snoring and coughing events, respectively.

is a self-rated questionnaire which assesses sleep quality and disturbances over a long time interval. Based on the detected events such as snoring and coughing, iSleep is able to estimate the answers to several PSQI questions such as "During the past month, how often have you had trouble sleeping because you cough or snore loudly?". Then a sleep quality score can be calculated based on scoring rules specified by PSQI.

As an unobtrusive, portable tool for in-place sleep monitoring, Sleep has a potential in helping users improve their sleep quality and stay healthy in many ways. For example, by providing a sleeping profile that depicts details of sleep-related events, iSleep allows the user to track the sleep efficiency over time, relate bad sleep to possible causes like extensive snores which are otherwise hard to identify, and help their healthcare providers diagnose trends related to certain diseases. Moreover, the fine-grained sleep events detected by iSleep can greatly improve the fidelity of subjective, questionnaire-based sleep assessment tools like PSQI whose utility is otherwise impeded by the inaccuracy of subject's memory and perception.

5. SYSTEM DESIGN

In this section, we describe the design of iSleep. First, we discuss the sleep-related events that iSleep can detect, and the acoustic features used to detect those events. Next, we describe how to estimate ambient noise. Lastly, we discuss sleep-related event classification. Our design is based on careful analysis of real data of a long-term experiment that involves 7 subjects and total 51 nights of sleep. The details of the experimental setting are described in Section 7.

5.1 Sleep Events and Feature Extraction

Since most people sleep in a relatively quiet environment

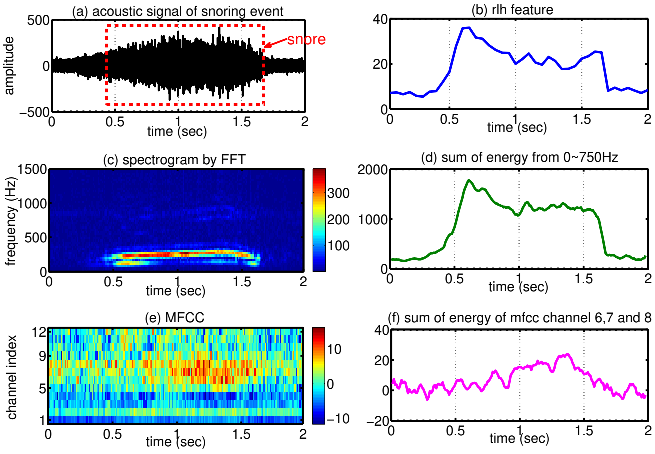


Figure 3: Comparison of different frequency-domain features including *rlh*, FFT and MFCC. The comparison is based on a 2-second recording containing a snoring event. (a) shows the acoustic signal of the recording. The part within the red rectangle represents the snoring event. (b) shows the *rlh* extracted from this recording. (c) and (d) show the spectrogram calculated by FFT and the summation of low-frequency (0-750Hz) energy, respectively. (e) and (f) show the energy of 12 MFCC channels and the total energy from channel 6 to 8, respectively.

at night, iSleep categorizes the possible sounds during sleep into sleep-related events, ambient noise, and other sounds such as those caused by cars/trains passing by. Specifically, the sleep-related events of interest include body movement, snoring and coughing. Our key insight is that, even though the acoustic profiles of sleep events are highly dependent on each individual, they have distinguishable features in terms of energy and frequency. For example, the dominant frequency of snoring is much lower than that of other events. iSleep requires user to manually start and close the app when he goes to bed and gets up, respectively. The duration of sleep is hence equal to the running time of iSleep. We note that it is possible to infer the time that the user goes to bed or gets up by combining accelerometer microphone readings. However, this requires the smartphone to constantly sample sensors, resulting in high energy consumption.

In order to build a light-weight classifier that can adapt to different individuals and environments, we choose three features based on the key characteristics of each sleep-related event. The first feature is root mean square (rms), which captures the loudness of sound. Let f be a frame that consists of n samples of acoustic amplitude $s_1, s_2, ..., s_n$. In our implementation, each frame contains 1,600 acoustic samples collected at 16 kHz. The rms of the frame is given by

$$rms(f) = \sqrt{\frac{s_1^2 + s_2^2 + ... + s_n^2}{n}} \quad (1)$$

where $rms(f)$ denotes the value of rms for frame f. Fig. 2 shows the power spectrals of moving, snoring and coughing. We can see that their energy distributions are different. For example, most energy of snoring concentrates on low-frequency band ($0 \sim 750Hz$), while the energy of moving

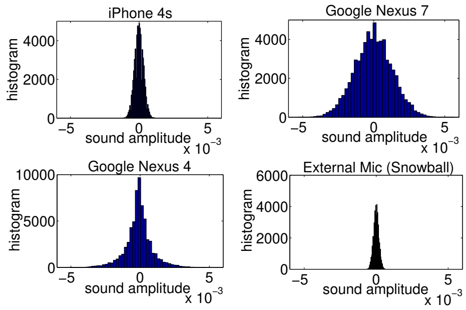


Figure 4: Histograms of recorder noise generated by different devices. The duration of acoustic data used is 4 seconds. The x-axis indicates the normalized amplitude.

distributes on both high and low frequency bands. Based on this observation, the second feature we choose is the ratio of low-band to high-band energies (rlh). The change of rlh over time reflects the transition of dominant frequency. The rise of rlh indicates that the proportion of low-band energy in the total energy is increasing. In other words, the dominant frequency is transiting in the direction of high to low. For example, the dominant frequency of snoring is significantly lower than that of the ambient noise and other activities. As a result, the rise of rlh can be used to effectively detect snoring. In order to compute the rlh of frame f, we need to calculate the energy of frame f in both low and high frequency bands. The low-band frame f^l is calculated by applying a low-pass filter as follows,

$$s_i^l = s_{i-1}^l + \alpha \times (s_i - s_{i-1}^l) \quad (2)$$

where s_i^l is the i-th acoustic sample of the low-band frame f^l, and the default value of α is set to be 0.25. The high-band frame f^h is computed by applying a high-pass filter as follows,

$$s_i^h = \alpha \times (s_{i-1}^h + s_i - s_{i-1}) \quad (3)$$

where s_i^h is the i-th acoustic sample of the high-band frame f^h, and the default value of α is set to be 0.25. Then the rlh of frame f, $rlh(f)$, is given by

$$rlh(f) = \frac{rms(f^l)}{rms(f^h)} \quad (4)$$

Fast Fourier Transform (FFT) and Mel-frequency cepstral coefficients (MFCC) [21] are two commonly used frequency-domain features for acoustic analysis. However, compared with them, rlh has significantly lower computation overhead, while yielding the same detection performance for our application. The computation of normalized rlh within a frame (containing 1600 samples) requires around 11200 addition and 8000 multiplication operations, while that of FFT needs over 54000 addition and 36000 multiplication operations [17]. Fig. 3 shows the comparison among the three aforementioned features based on a two-second recording of snoring. We can see that, rlh (Fig. 3(b)) yields a similar result as FFT (Fig. 3(d)), which can be used to clearly detect the snoring. The performance of MFCC (Fig. 3(f)) is inferior to that of rlh. Designed for speech recognition, MFCC is more sensitive to vocal sound, which have very different

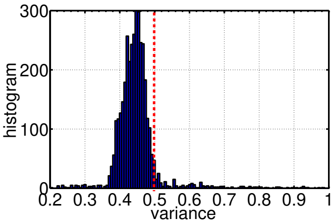


Figure 7: The histogram of variance of $\overline{std}$ (Equ. 5) within 4-second noise window. The data is extracted from real experiments, containing $3,644$ noise windows ($14,576$ seconds).

characteristics from snoring. The third feature is variance (var), which reflects how far the amplitudes of acoustic signals within the frame are spread out. For example, the var of a frame associated with body movement is typically much lower than that of snoring or coughing.

5.2 Noise Characterization

In a typical scenario, the events related to sleep quality, including body movement and snoring, are rare while most sounds are noise generated by various sources. Thus noise identification is critical to the accuracy of detecting sleep-related events. In this paper, we refer to the ambient sounds that last for a relatively long period as noises, as opposed to the short-duration sounds that are caused by a sleeping individual. Specifically, acoustic noise is caused by two major sources. The first is the background noise generated by the recording unit itself while recording. The level of recorder noise of built-in microphone of smartphone is much higher than that of external standalone microphone. In addition, the recorder noise level varies with different smartphones.

Fig. 4 shows the amplitude distributions of recorder noises collected by sampling the microphones of 4 different devices (iPhone 4s, Nexus 7, Nexus 4 and an external conference microphone) in a quiet room for 4 seconds. The sampling frequency is 16 kHz and the value of each sample is scaled to $[-1, 1]$ from its original 16-bit representation. We can observe that the noise level of external conference microphone is substantially lower than those of the built-in microphones of smartphones. Among the smartphones, iPhone 4s generates the lowest level of recorder noise.

Another common source of noise is appliances that are operating overnight (e.g., fan or air-conditioner). Fig. 5 shows the distribution of three features extracted from different types of noises. Three groups of audio clips are used, which are noise recorded by Nexus 4, noise with A/C operating recorded by Nexus 4, and noise recorded by Nexus 7. The total length of each audio clips is around 140 minutes (84000 frames), and they are extracted from data collected from different users during sleep. We can observe that the mean and standard deviation across different groups differ substantially. For example, the operation of A/C results in a wider distribution in all three features (rms, rlh and var). However, a key observation is that most values in all three cases fall within the range of $mean \pm 3 \times std$, regardless of the types of the noises or smartphones used.

5.3 Noise Model Estimation

The result in Section 5.2 suggests that, the key feature that differentiates noise from event-related sound is its relatively stable variance. This is due to the fact that the noise does not vary substantially within a short duration (i.e., a few seconds). This observation allows us to design a simple yet robust sleep-event detection method based on the noise profile. The basic idea is to detect events based on the thresholds of features that are calculated using the noise measurement. To adapt to different environments, the system continuously detects and updates the current noise model, which is used to calculate the thresholds used to detect and classify sleep events.

Specifically, iSleep first detects noise from a sequence of frames with stable standard deviations. It involves two steps. First, the system calculates the standard deviation $std_i = std(f_i)$, where f_i denotes the i-th frame, which captures the stability of acoustic signal within a frame. However, the standard deviation varies with different devices and noise profiles. Therefore, in order to to improve the robustness of noise detection, we normalize the standard deviation of each frame within a T-second window (containing 40 frames in our implementation) as follows:

$$\overline{std_i} = \frac{std_i - std_{mean}}{std_{mean} - std_{min}} \tag{5}$$

where std_{mean} and std_{min} denote the mean and minimum standard deviation within the current window W. Second, the system calculates the variance of the normalized standard deviation within the window W. Fig. 7 shows the histogram of the variance based on $3,644$ noise windows collected from real experiments conducted by different subjects. We can see that the variances of most noise windows are grouped within [0.4,0.5]. More than 95% variances are below 0.5. Therefore, we use 0.5 as a threshold to detect noise. Specifically, the frames within a noise window will be considered as noise if the variance is lower than 0.5.

Fig. 6(b) plots the normalized standard deviation of different noises. We can see that, even though they have different acoustic amplitude distribution (shown in Fig. 6(b)), the normalized standard deviations are similar. Since their variances are lower than the preset threshold 0.5, they will be classified as noise. The histogram in Fig. 8 shows the distribution of acoustic signals for a duration of 4 seconds. It contains a slight body movement lasting from around 2.8s to 3.8s. As the sound of the movement is very slight, its distribution is close to that of noise without operating appliance. However, we can observe that the normalized standard deviation has a variance of 1.75, which clearly reflects the movement event. Therefore, these frames will be classified as frames of interest and fed into the feature extraction component.

A typical scenario at night is that the fan of A/C or heater is automatically turned on and off, leading to a changing noise profile. Fig. 9 shows the scenario where the air conditioner is turned on. Fig. 9(b) shows the variance over the past 4 seconds. In the detection result shown in Fig. 9(c), we can observe that only a short period corresponding to the transition is detected as non-noise sound. As such noise misclassification only occurs in occasional transient states of appliances, it does not affect the accuracy of sleep-related event detection. Fig. 10 shows another typical scenario where the sounds of sleep events are included. We can see that only the parts without human generated sounds are detected as noise. Therefore, our noise detection algorithm is able to effectively detect changing noise.

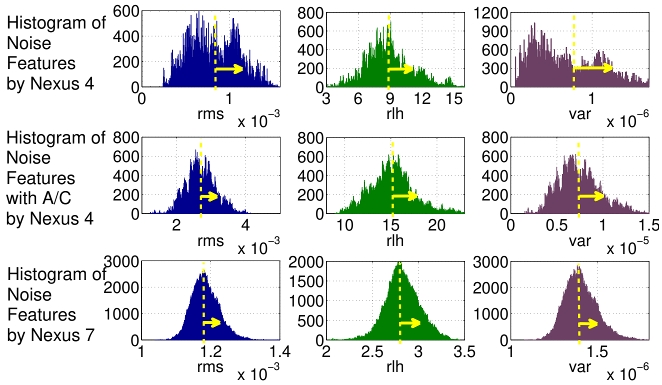


Figure 5: Histograms of three noise features recorded by different devices. The dotted lines indicate the mean value and the length of the arrow indicates the value of standard deviation. The duration of each audio clip is around 140 minutes, containing about 84,000 frames.

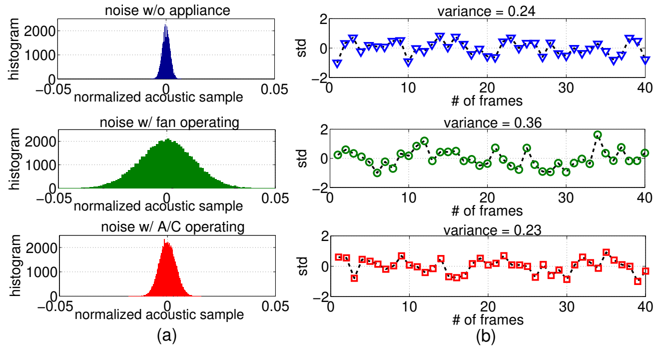


Figure 6: (a) The histograms of sound intensity of different types of noises. (b) The normalized standard deviation of each frame. The acoustic data is collected by iPhone 4s for a duration of 4 seconds.

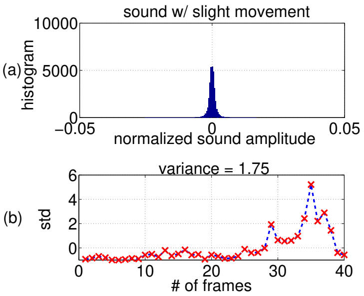


Figure 8: (a) The histogram of sound intensity of a 4-second recording containing body movement. (b) The normalized standard deviation of each frame. The acoustic data is collected by iPhone 4s for a duration of 4 seconds.

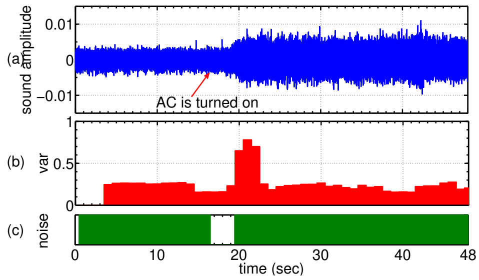


Figure 9: The process of detecting changing noise. (a) shows the amplitude of acoustic signals sampled in the bedroom at night. Around the 19-th second, the air conditioner was turned on. (b) shows the variance of the signal over the last 4 seconds. (c) shows the result of noise detection.

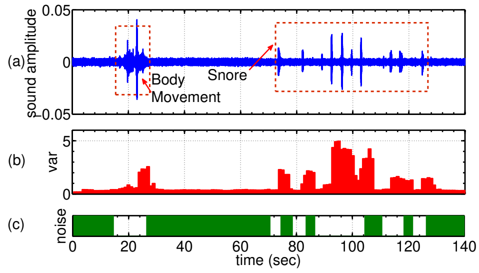


Figure 10: The process of detecting noise when the sounds of sleep events are included. (a) shows the sound wave of a 140-second audio collected by Nexus 4 with the sound of body movement and snoring. (b) shows the variance of the sound signal over the last 4 seconds. (c) shows the result of noise detection.

After a sequence of frames is detected as noise frames, iSleep calculates three features for each of the frames. In order to estimate the current noise, iSleep computes the mean and standard deviation ($mean(rms)$, $std(rms)$, $mean(rlh)$, $std(rlh)$, $mean(var)$ and $std(var)$) for each feature. Then, each newly calculated distribution feature F_{new} will be used to update the current corresponding feature (F_{cur}) according to an Exponential Moving Average (EMA) algorithm as follows,

$$F_{cur} = F_{cur} + \beta \times (F_{new} - F_{cur}) \tag{6}$$

In our implementation, the default value of β is set to be 0.5. The EMA algorithm ensures that estimated noise model adapt to the changing noise profile. After the features of current noise are updated, they will be stored and used in the event detection process.

5.4 Event Detection

The main objective of sleep-related event detection is to achieve robust performance across different users, smartphone platforms and environments. iSleep adopts an adaptive event detection algorithm that adapts to the estimated noise model. First, acoustic features are extracted and normalized for each frame that is not detected as noise frame. Then, based on the normalized features ($\overline{rms}$, $\overline{rlh}$ and $\overline{var}$), frames are classified. Lastly, we apply operators in mathematical morphology [30] to the classification results to filter out false-positive and false-negative errors.

5.4.1 Feature Extraction

iSleep normalizes the features of each frame based on the current measurement of noise. Such a calibration process allows the system to adapt to different devices and environments. For example, the rms value of frame f is normalized as follows,

$$\overline{rms}(f) = \frac{rms(f) - mean(rms)}{std(rms)} \tag{7}$$

where $\overline{rms}(f)$ is the normalized rms, $mean(rms)$ and $std(rms)$ are the current distribution features associated with rms ex-

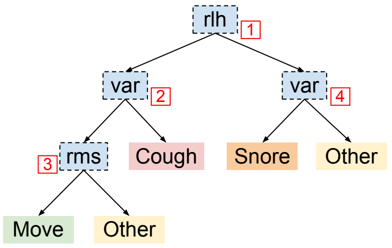


Figure 11: The decision tree for event detection. The splitting conditions are shown in Fig. 14(a), 14(b) and 14(c).

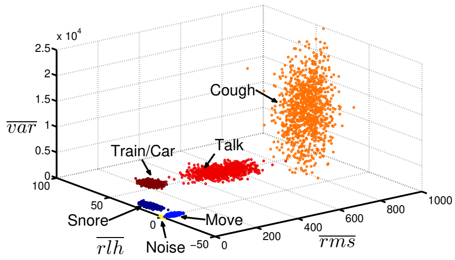


Figure 12: The sound feature vectors of different events in feature space.

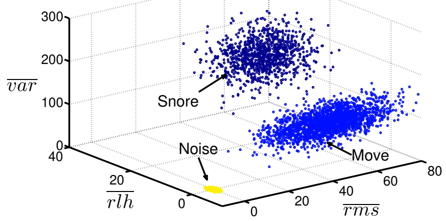
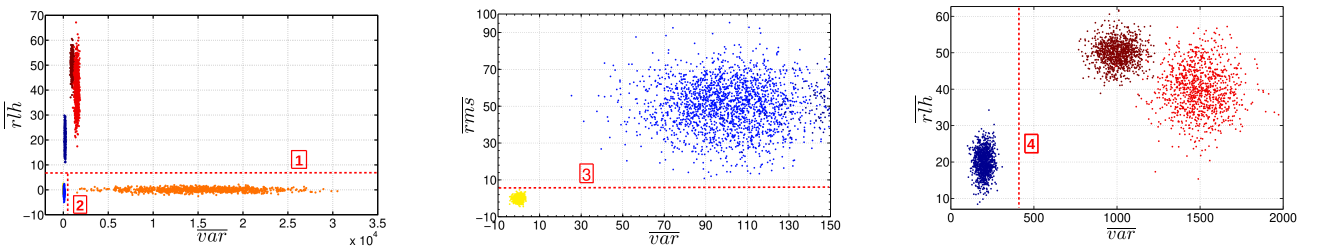


Figure 13: The sound feature vectors of snore, move and noise in the feature space.



(a) (b) (c)

Figure 14: The splitting conditions of the decision tree in Fig. 11. (a) shows conditions [1] and [2], (b) shows condition [3], and (c)shows condition [4].

tracted from the noise. Likewise, rlh and var are also normalized using the corresponding distribution features of the noise and the results are denoted as $\overline{rlh}(f)$ and $\overline{var}(f)$, respectively.

Fig. 12 shows the distribution of three normalized features. It is plotted from the data collected from 7 subjects using 3 different devices during a one-week experiment. Fig. 13 shows the zoom-in plot after excluding coughing and taking events. The frames associated with each event are manually labeled and extracted. Then the three features of each frame are calculated and normalized using the current model of the noise. Four test subjects live close to the railway, and one subject lives in an apartment with automatic central A/C. In addition, the three devices have different noise profiles. However, Fig. 12 and 13 show that, after normalization, the same events are grouped together and different groups are clearly separable.

5.4.2 *Event Classification*

Motivated by the results in Fig. 12 and 13, we design a decision-tree based classifier (shown in Fig. 11) to detect sleep-related events. Decision tree is robust to errors and widely used for classification problems with attribute-value pairs and discrete output values [26]. Fig. 14 shows the splitting conditions of the decision tree. The dotted rectangles indicate the splitting features, and the leaf nodes denote the classification results. The splitting features and thresholds are determined based on the information gain calculated using entropy. Specifically, the entropy of a node T is given by

$$Entropy(T) = -\sum_j p(j) \cdot log(p(j)) \tag{8}$$

where $p(j)$ is the relative frequency of class j at node T. Since iSleep focuses on detecting three sleep events, therefore, $j = 3$. After splitting node T into k nodes (T_1, T_2, ... T_k), the information gain is given by

$$G = Entropy(T) - [\sum_{j=1}^{k} \frac{n_i}{n} Entropy(T_j)] \tag{9}$$

where n_j is the number of samples in node T_j and n is the number of samples in node T. For each splitting, the system chooses the split that maximizes the information gain.

Next, we describe the classification process in detail. First, the non-noise frames are split into two groups according to the $\overline{rlh}$, which captures the dominant frequency. Since $\overline{rlh}$ is the ratio between low-band and high-band energy, high $\overline{rlh}$ means low dominant frequenc. As we can observe in Fig. 14(a), this splitting condition is able to separate the sounds into two groups; one group includes sounds of noise, movement and coughing with relatively high frequency, and another group includes other sounds with lower dominant frequencies.

Then the frames in the high-frequency group are further split into two groups based on $\overline{var}$, shown in Fig. 14(a). Since $\overline{var}$ reflects how far the intensities of acoustic signals within the frame are spread out, it is able to separate frames caused by coughing from those caused by movement and noise in the high-frequency group. Therefore, the system is able to detect cough frames at this stage. Likewise, as shown

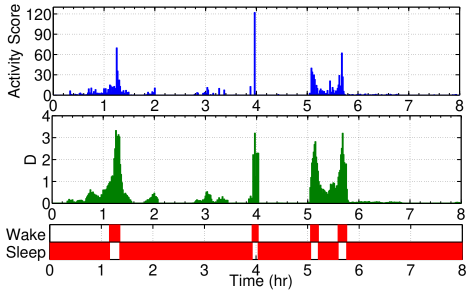


Figure 15: Actigraphy-based sleep/wake estimation based on recording of 8 hours.

in Fig. 14(c), in the low-frequency sound group, frames containing snore can also be detected according to $\overline{var}$.

In the third branch shown in Fig. 11, we aim to split the frames caused by movement and noise according to $\overline{rms}$. Both sounds have relatively high dominant frequency and low variance. As shown in Fig. 14(b), since the sound intensity of movement is higher than that of noise, $\overline{rms}$ is able to separate the frames caused by movement from those caused by noise. As a result, the movement frames are detected at this stage.

5.4.3 *Handling Mis-classifications*

As the occurrence of most events results in clustered frames, isolated event frames are likely false positives. Therefore, after event frames are detected, we apply the opening operator in mathematical morphology [30] to filter out isolated frame events. Mathematical morphology is widely used in the identification of geometrical structures in image processing. Specifically, single or continuous event frames can be filtered out if the number of these continuous frames is less than the operator diameter (default value is 5). We apply the closing operator [30] to the resultant frame sequence after applying the opening operator, in order to connect those event areas with narrow gaps between them. This is because the narrow gap between two continuous event frame clusters is likely false negative. Specifically, if the length of the gap is less than the diameter of closing operator, the frames within the gap will be classified as event frames. Finally, we apply dilation operator [30] with the diameter of 2 frames to the continuous event frames. This will result in an expansion of 2 frames on both ends of the event frame sequences. The purpose of dilation is to ensure that the "edge" of this event is included.

Fig. 16 shows the event detection process in a typical scenario. The duration of the acoustic signal is 85 seconds, where the body movement of the user (18-26th second) is followed by a sequence of snoring events (38-83th second). Figure 16(a), (b) and (c) show the normalized features. We can observe that the increase of $\overline{rlh}$ clearly reflects the snoring event. The movement events usually have the similar $\overline{rlh}$ with noise, but higher $\overline{rms}$ and $\overline{var}$. In Fig. 16(d), the first plot shows the classification result for each frame. We can see that the movement and snoring events are detected correctly, but several noise frames are misclassified as event frames. The second plot in Fig. 16(d) shows the event detection result after we apply the opening, closing and dilation operators. We can see that the isolated misclassified frames are removed, and the gap between two sequences of movement frames are closed. However, the snoring event at

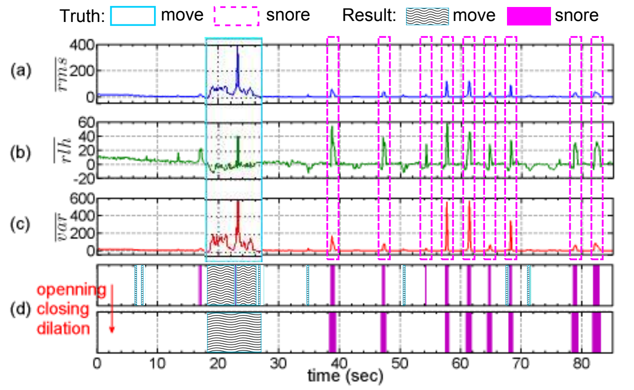


Figure 16: The event detection process of a 85-second acoustic signal that contains sound of body movement and snoring. (a), (b) and (c) show the features for each frame, respectively. (d) shows the event detection results before and after opening, closing and dilation operations.

around 58 second is also filtered out. This is mainly because this particular snoring event is extremely short with very low intensity. As a result, only a single frame within this snoring event is detected as snoring frame. This frame is removed in the opening operation, causing a false negative error. Note that the detected snore events are used to calculate PSQI scores, which only require coarse-grain information about sleep-related events. Specifically, the detected snore will be used to automatically answer the question: 'During the past month, how often have you had trouble sleeping because you cough or snore loudly'. The options include 'not during the past month', 'less than once a week', 'once or twice a week' and 'three or more times a week'. As a result, a few misclassifications during a particular sleep will not affect the choice of answer. Therefore, misclassifying a weak snore or cough event is acceptable because its impact on the final sleep quality assessment is negligible.

5.5 Sleep Scoring

iSleep uses the detected sleep-related events to derive quantitative measures of sleep quality based on two criteria. One is actigraphy that only requires information about body movement of a whole-night sleep. The other is PSQI, where all the detected events are considered jointly for estimating the sleep quality.

In our implementation of actigraphy-based estimation, iSleep adopts similar method proposed in [31], where the sleep/wake state of a minute is determined by taking 4 previous minutes and 2 following minutes into account. The model takes the form:

$$\begin{aligned} D = &P(W_{-4}A_{-4} + W_{-3}A_{-3} + W_{-2}A_{-2} + W_{-1}A_{-1} \\ &+ W_0A_0 + W_{+1}A_{+1} + W_{+2}A_{+2}) \end{aligned} \tag{10}$$

where P is a scale factor for the entire equation, W_{-i}, W_0 and W_{+i} represent the weighting factor for the previous minute, current minute and following minute, and A_{-i}, A_0 and A_{+i} indicate the activity scores for the previous minute, current minute and following minute, respectively. If $D \geq 1$,

A1	Time to go bed at night
A2	Minutes taken to fall asleep
A3	Get-up time in the morning
A4	Hours of actual sleep per night
B1	Cannot sleep within 30 minites
B2	Wake up in the middle of the night or early morning
B3	Cannot breath comfortably
B4	Cough or snore loudly

Table 1: Metrics from PSQI that iSleep uses to estimate the sleep quality.

Duration of Sleep	A4
Sleep Disturbance	B1, B2, B3, B4
Sleep Latency	A2, B1
Sleep Efficiency	A1, A3, A4

Table 2: The left column is the components in PSQI that can be derived from detected events. The right column is the metrics that are used to calculate the corresponding component score.

the state of the current is determined as wake, whereas $D \leq 1$ means the current minute is in sleep state. The model used in iSleep adopts the weighting factors suggested in [31]. It takes the following form:

$$D = 0.125(0.15A_{-4} + 0.15A_{-3} + 0.15A_{-2} + 0.08A_{-1} + 0.21A_0 + 0.12A_{+1} + 0.13A_{+2}) \quad (11)$$

where the activity score A is the number of frames associated with body movement in each minute. Fig. 15 shows the prediction of wake/sleep state over a 8-hour recording during sleep. The first plot shows the calculated activity scores for each minute. The second plot shows the calculated D value by using Eqn. 10. The last plot is the sleep/wake estimation result. We can see that the user changes from sleep state to wake state 4 times throughout the night. The duration of each wake state lasts around 10 minutes. The sleep efficiency is defined as the ratio of actual sleep time to total in-bed time:

$$Sleep\ Efficiency = \frac{T_{sleep}}{T_{sleep} + T_{wake}} \quad (12)$$

For the long-term sleep quality estimation, iSleep calculates the scores of 4 components listed in Table 2 from PSQI. In order to calculate the component scores, iSleep measures the metrics listed in Table. 1 based on the detected events. However, some of the metrics used to calculate the score of *Sleep Disturbance* can not be measured by iSleep. For example, some of them are related to the bedroom's temperature, whether having dream, or feeling pain during sleep. As a result, instead of using the 9 metrics, iSleep uses only 4 metrics that can be inferred from the detected events and scales the score by multiplying $\frac{9}{4}$. The scoring rules of the other components are the same as specified in PSQI.

6. IMPLEMENTATION

iSleep is implemented on Android 4.2.2 Jelly Bean. The application file has a size of around 1 MB and takes 2.7 MB storage on the phone after installed. It requires about 20 MB RAM allocation while running. The displaying and processing functions are implemented in separate threads to ensure the timeliness of acoustic sampling and processing.

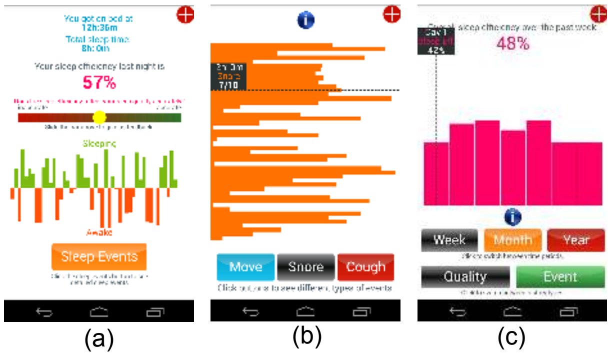


Figure 17: The user interface of iSleep. (a) The screen showing sleep efficiency and sleep states over night. (b) The screen showing the sleep events detected overnight and the normalized loudness of each event. (c) The screen showing the history of sleep efficiencies and events.

iSleep samples the built-in microphone at 16 KHz. The samples are buffered and segmented into frames with the duration of 0.1 second. Based on the variance, iSleep detects non-noise frames and noise-frames. Noise frames are used to estimate current noise distribution. Then, according to the noise distribution and features extracted from the non-noise frames, iSleep detects sleep-related events and saves them for further processing. Lastly, for each night, iSleep uses actigraphy to generate a short-term sleep quality report according to the movement events. For each week, iSleep estimates the long-term sleep quality according to PSQI and all the detected sleep events. To protect the users' privacy, iSleep only buffers the raw acoustic signal for the last 4 seconds and the buffered data is erased after the feature extraction.

We have released an initial version of iSleep on the Google Play Store [2]. The screen shots are shown in Fig. 17. The application is easy to use and understand. Before sleep, the user only needs to start the app and put the phone on the nightstand within 6 feet of the bed. iSleep prevents the CPU from sleeping, so that it can still keep running after the screen is turned off by pressing the power button. After getting up, the user needs to stop the monitoring to see the sleep efficiency and detected sleep events. Within 6 days of release, iSleep has been installed by around 100 users from more than 9 countries on various Android devices. The feedbacks collected from the Google Play Store and the app show that users like to use iSleep to track their sleep quality and be aware of their sleep events.

7. EVALUATION

In this section, we evaluate the performance of iSleep using experiments. Our primary results show that iSleep is able to effectively capture various sleep events and accurately estimate the sleep quality of the user.

7.1 Experimental Setting

We conduct three sets of experiments to evaluate the performance of iSleep. Section 7.2 presents the experimental results of a long-term experiment that involves 7 subjects and total 51 nights of sleep. All of the subjects are college/graduate student volunteers ranging from 20 to 30 years

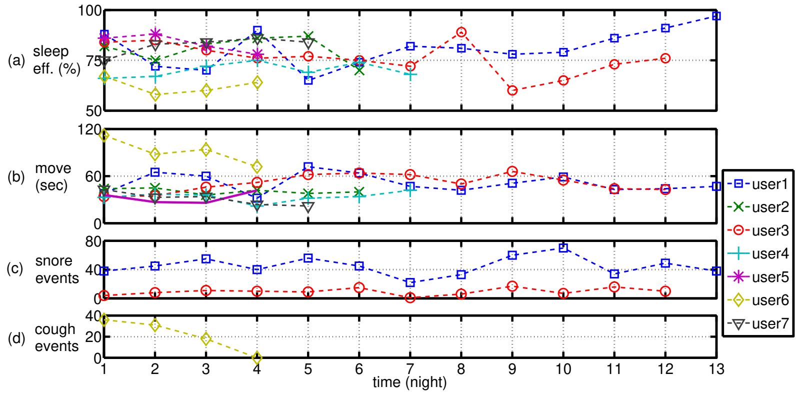


Figure 18: The sleep efficiency and sleep events of different users during the long-term experiment. (a) The sleep efficiency at each night. (b) The total time (in seconds) of body movement at each night. (c) The number of snoring events at each night. (d) The number of coughing events at each night.

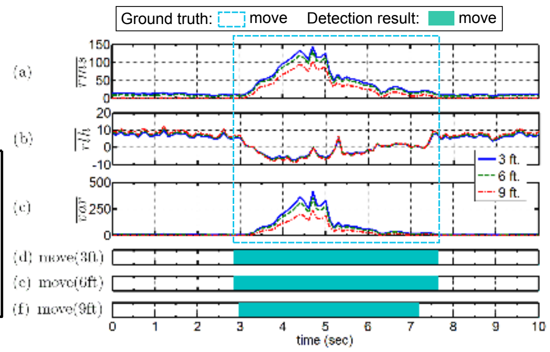


Figure 19: Event detection results based on a 10-second audio clip recorded at different distances (3, 6, and 9 feet). The movement events are labeled. (a), (b) and (c) show their features. (d), (e) and (f) are the detection results.

old, and 2 of the 7 subjects are male. Section 7.3 presents micro-benchmarks that evaluate the system performance under different settings and environmental factors. Section 7.4 evaluates iSleep based on the data collected from its Android application released on Google Play Store.

Two metrics are used to quantify the performance of sleep event detection. *Event detection accuracy* (EDA) evaluates the accuracy of snoring and coughing detection. It is defined as the ratio of the number of correctly detected events to the total number of events. A snoring or coughing event is successfully detected as long as a subset of the frames associated with it is correctly classified. For instance, a snoring event containing 10 frames is considered as detected if iSleep correctly classifies at least one of these frames. This is mainly because iSleep only considers the number of occurrences of these two types of events when calculating PSQI scores. Moreover, the duration of a single snoring or coughing event is relatively short (typically around 1 second). Therefore, for our application, it is not necessary to detect the exact duration of these events.

Another performance metric used in our evaluation is *frame detection accuracy* (FDA). Different from EDA, FDA quantifies the performance of body movement detection. It is defined as the percentage of the correctly classified movement frames in all frames associated with the movement. This is because the calculation of activity score D (Eqn. 11) in actigraphy is based on the number of frames associated with body movement in each minute. Therefore, it is important to evaluate iSleep's accuracy in detecting the duration of body movement.

Our both evaluation metrics are based on true positive results, because the classification algorithm yields very few false negative and false positive results. The reason for this is two fold. First, the acoustic features we chose are very effective in differentiating different events. As a result, a sleep event is rarely mis-classified as another type of event (false negatives results). As shown in Table. 3, over 51 nights of sleep, around 6-10% of frames associated with movement are misclassified as noise, 48 out of 1446 (3.3%) snore events are misclassified as other events. Second, adaptive noise modeling and the mathematical morphology adopted by iSleep can

sbj	nights	move(FDA)	snore(EDA)	cough(EDA)
1	13	91.7%	585/601(97.3%)	0/0
2	6	92.0%	0/0	0/0
3	12	89.8%	114/122(93.4%)	0/0
4	7	93.4%	0/0	0/0
5	4	94.1%	0/0	0/0
6	4	92.2%	0/0	85/85
7	5	90.1%	0/0	0/0
ttl	51	91.9%	699/723(96.7%)	85/85

Table 3: The event detection result based on the data collected from 7 subjects and total 51 nights of sleep.

effectively eliminate the noise frames that are mis-classified as sleep events (false positive results). Our long-term experiments show that, less than 20 minutes of various noises ($< 0.08\%$) are misclassified as sleep-related events, and no noise frames are misclassified as snore or cough events.

7.2 Long-term Experiment

In this section, we present the result of a long-term experiment. 7 participants (2 males, 5 females) are recruited for data collection. The duration of the data collection for each participant varies from 3 to 14 days. There are totally 51 nights of sleep during the experiment.

The experimental platform used in data collection is composed of three components. First, two smartphones/tablets are put on both sides of the subject to collect acoustic data during sleep. The distance between the phone and the subject is around 5 feet, unless otherwise specified. The recorded audio clips are stored on the phones and retrieved for analysis after the experiment. Second, an omnidirectional microphone (Snowball USB Microphone) is attached on the headboard to collect the high-quality audio as ground truth of snoring and coughing events. A small laptop is connected with the microphone to store recorded audio clips. In order to minimize the impact of the noise from laptop fan, we place the laptop 16 feet away from the bed and connect it with the microphone using a long cable. Third, in order to record the ground truth of body movements, an iPod touch is used to log the acceleration data of the user during sleep. In order to mitigate the impact of differences of mattresses,

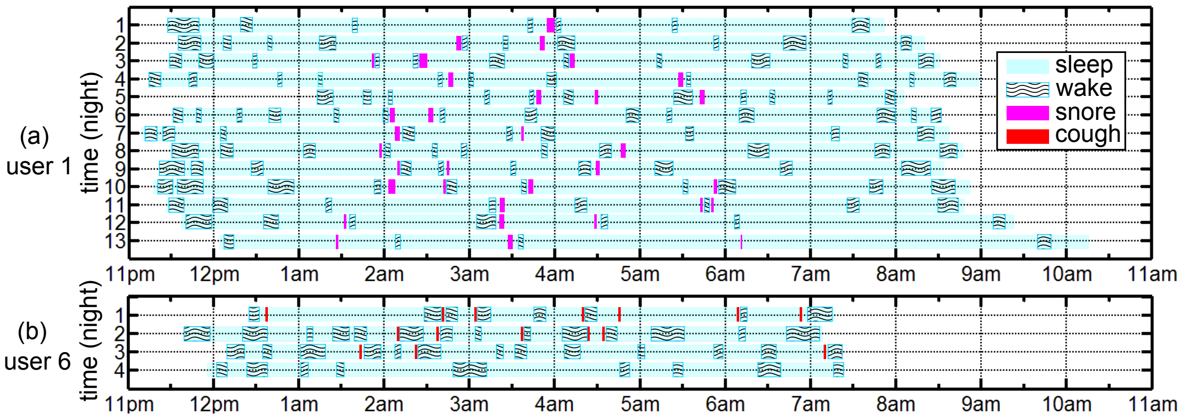


Figure 20: The sleep states and events of user 1 and 6 during the long-term experiment. The sleep states are calculated using the body movements based on actigraphy.

it is put inside a sport armband attached to the subject's lower leg.

After the data collection, we first synchronize the data captured by different devices. Then the snoring and coughing events are manually labeled from the high quality audio clips recorded by external microphones. The acceleration data collected by the iPod attached to the leg is used to obtain the ground truth for movement events. To evaluate the accuracy of sleep quality monitoring, each subject is asked to fill a PSQI questionnaire about how they feel about their sleeps during the period of data collection. The questionnaires are then used to correlate with the sleep quality measures calculated by iSleep.

Our evaluation is based on the data collected from 7 subjects and total 51-night sleeps. The overall detection results are shown in Table 3. We can see that the movement detection accuracies are relatively stable across different subjects, and the average FDA is 91.9%. The snoring detection accuracy is 97.3% and 93.4% for subject 1 and 3, respectively. The system achieves 100% coughing detection accuracy for subject 6.

Fig. 18 shows the sleep efficiency and sleep events detected by iSleep during the experiment. There are two snorers (user 1 and 3) out of seven subjects. Specifically, user 1 usually snores periodically (every 2 or 3 seconds) for a duration of around one minute. User 3 snores more sporadically. Another observation is that coughing events are detected for user 6 during the first three days of experiment. This is due to the fact that user 6 happened to catch a cold. The number of coughing events gradually decreases every night as the user recovers. We can see that the users who snore or cough during sleep are more likely to have more dynamic and lower sleep efficiency. The main reason is that snores and coughs are usually followed by body movements, which indicate wakefulness of the user.

Fig. 20 shows the detailed sleep states and events detected by iSleep of user 1 and 6. We have several interesting observations: (1) Most of the snoring and coughing events are accompanied by body movements, which cause the subject to transition from sleep state to wake state. (2) User 1 usually snores during 2 to 5 am. (3) At the fifth night, user 1 got in bed about 1.5 hours later than she usually does. Due to the reduction of sleep time, her sleep efficiency is significantly lower than the other 12 nights. (4) The low sleep efficiency of user 6 is caused by her relatively short sleep time (around 7 hours) and frequent body movement due to the cold symptoms. The subjects are enthusiastic about these patterns of their sleep discovered in the experiment, and expressed interests in adopting tools like iSleep for long-term, daily sleep monitoring. We also note that iSleep users are able to make these observations easily on their own through the interface of iSleep shown in Fig. 17.

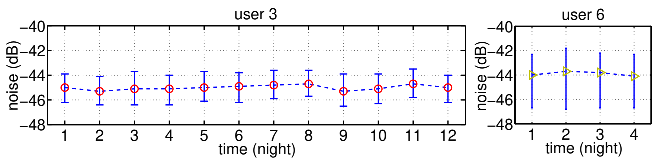


Figure 21: The noise levels (with 95% confidence interval) of two subjects during the long-term experiment.

We observed different noise profiles in the sleeping environments of participants. Fig. 21 shows the noise intensity at different nights for User 3 and User 6. Although the noise level for each subject is relatively stable over time, the noise of User 6 is louder and substantially more dynamic than that of User 3. The high-quality audio data of external microphone confirms that this was due to the louder A/C in User 6's home. Despite the substantial differences in sleeping environments of different subjects, iSleep achieved consistently high detection accuracy, as shown in Table 3.

In order to evaluate the performance of measuring long-term sleep quality, we compare the 4 component scores (shown in Table 2) that are obtained from the subjects' PSQI questionnaires and iSleep. According to the scoring rules of PSQI, each component is scored on a scale of 0 to 3, where 0 means better and 3 means worse. For example, the score of Sleep Efficiency is calculated as follows:

$$\begin{cases} \text{Sleep Efficiency} \geq 85\%, & \text{score}=0; \\ 75\% \leq \text{Sleep Efficiency} < 85\%, & \text{score}=1; \\ 65\% \leq \text{Sleep Efficiency} < 75\%, & \text{score}=2; \\ \text{Sleep Efficiency} < 65\%, & \text{score}=3. \end{cases}$$

iSleep calculates the component scores according to the same rules based on the detected sleep events. Table 4 shows the scores computed from subject questionnaires and by iSleep. We can observe that there is only one mismatch for Duration of Sleep, Sleep Disturbance, and Sleep Latency. And

Subject	Duration of Sleep		Sleep Disturbance		Sleep Latency	
	PSQI	iSleep	PSQI	iSleep	PSQI	iSleep
1	1	1	1	1	1	1
3	2	2	1	1	1	1
4	1	1	1	1	0	**1***
5	1	1	1	1	1	1
6	3	3	3	**2***	1	1
7	2	**1***	1	1	0	0

Table 4: The comparison of PSQI scores provided by the subjects and computed by iSleep. The component score ranges from 0 (best) to 3 (worst). The scores of iSleep that do not match those from subjects' PSQI questionnaires are labeled by *.

the score discrepancy for each mismatch is only one. This result demonstrates that iSleep can accurately predict users' answers to these questions based on objective assessment of sleep-related events. As a result, iSleep can be used as a reliable sleep diary that significantly reduces users' burden on remembering the details of their past sleeps.

7.3 Micro-benchmarks

This section presents a set of micro-benchmarks that evaluate the performance of iSleep under various distance and noise settings. We also evaluate iSleep in two-user scenarios and its processing overhead and impact on battery lifetime.

7.3.1 Impact of Distance

In real scenarios, users likely place their phones at different distances from the bed. In order to evaluate iSleep's performance with respect to the distance between the phone and the user, we put phones at 3, 6 and 9 feet away from the user during sleep, respectively. The evaluation for each distance is based on a one-night experiment that lasts about 6.5 hours containing movement and snore events. The result of movement detection is shown in Fig. 23. We can see that increasing the distance between the phone and the user leads to lower movement detection accuracy. When the distances are 3 feet and 6 feet, the mis-classifications are mainly caused by minor leg or arm movements with relatively low sound intensity. However, when the distance is 9 feet, the sound intensity of movement events is substantially reduced. Another observation is that, the FDAs of different devices are relatively consistent at the same distance. This is because the acoustic features used in classification are normalized by the current noise model, making the detection robust against the differences in microphones' sensitivities.

Fig. 19 shows the features and detection results of a 10-second audio clip captured from different distances. We can observe that the increase of distance leads to lower $\overline{rms}$ and $\overline{var}$. As a result, the frames on the edges of a movement event with low sound intensity are more likely to be misclassified as noise. However, the detection of snore events is not affected by distance because of the significantly higher sound intensity.

Next, we investigate the accuracy of recognizing two users under different device and distance settings. As discussed in Section ??, iSleep compares the $\overline{rms}$ calculated by two devices to differentiate the events of different users. We focus on body movement events here because they have lower intensity than other events and hence are more difficult to differentiate. The recognition accuracy is defined as the percentage of movement frames which are correctly associated with the user. Six pairs of devices are used for each setting of distance difference. For each pair, one device is put 3 feet away from the user, while the other is located at 4, 4.5, 5, 5.5 and 6 feet away, respectively.

Device pair	1 ft.	1.5 ft.	2 ft.	2.5 ft.	3 ft.
pair 1	60/62	58/58	66/66	61/61	60/60
pair 2	59/62	57/58	65/66	61/61	60/60
pair 3	60/62	57/58	66/66	61/61	60/60
pair 4	57/62	55/58	65/66	60/61	60/60
pair 5	60/62	57/58	65/66	60/61	60/60
pair 6	57/62	56/58	64/66	60/61	59/60

Table 5: The user recognition accuracy by taking the majority vote for each movement event. The details devices are shown in Fig. 22.

Fig. 22 shows the recognition accuracy based on $\overline{rms}$ of each frame. We can observe that, the accuracy raises with the difference of distances. Moreover, the pairs consisting of the same model of devices result in higher recognition accuracy (over 91%), because their microphones have similar sensitivity. However, since each sleep-related event is composed of a sequence of frames, the recognition accuracy can be improved by taking a simple majority vote of the frames. As shown in Table 5, the average recognition accuracy is improved to 98%. The mis-recognitions mainly occur on the movement events with a short duration, such as a slight leg jerking for less than one second. When the distance difference is 1.5 feet or further, iSleep can achieve a recognition accuracy of more than 95%.

7.3.2 Impact of Noise

We now examine the impact of noise on the performance of iSleep. The evaluation is based on the real data containing body movements, snoring, coughing, and noises from various appliances including a humidifier (around 9 feet away from the bed), a ceiling fan (around 8 feet above the bed) and the central A/C (two vents on the ceiling of the bedroom). The operational time of each of these appliances is at least 2.5 hours. iSleep can reliably detect all the snoring and coughing events under different noises. The result of movement detection is shown in Fig. 24. We can observe that the operation of appliances increases the average noise level, leading to an up to 10% drop in FDA. This is mainly because when the noise level rises, some movement frames with low sound intensity are mis-classified as noise. Specifically, iSleep can still achieve over 90% movement detection accuracy, while the ceiling fan and humidifier are operating.

Fig. 25 shows a typical event detection process in the presence of noise. The duration of the audio clip is 40 seconds, when the A/C starts operating at 0 second. We can observe that during the first 4 seconds, the $\overline{rlh}$ rises from around 0 to around 20, due to the low-frequency sound from A/C. Then the sound of the first 4 seconds is detected as noise, and used to update the noise model. As a result, the $\overline{rlh}$ falls back to around 0 at the 5th second. At the 9th second, the $\overline{rlh}$ rises again, due to the speed change of the A/C fan. iSleep detects sound from 10 to 14 seconds as noise, and updates the noise model at the 14th second.

7.3.3 Processing Time and Energy Consumption

We expect the smartphone to be connected to the charger when iSleep is used for the whole night. However, users may forget to charge the phone, or would like to use iSleep

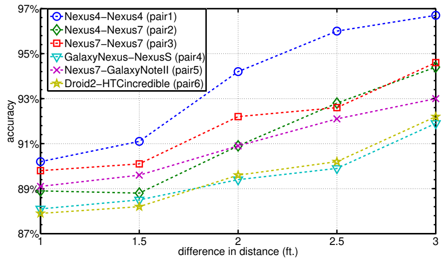


Figure 22: The accuracy of user recognition in two-user scenarios. In each experiment, six pairs of devices are used. The audio in each experiment contains movement events for a duration of around 10 minutes.

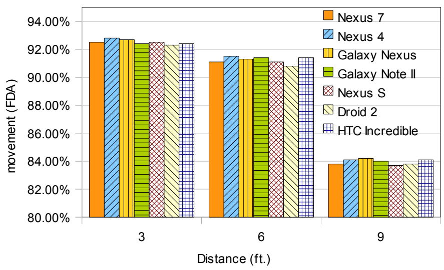


Figure 23: The impact of distance between the phone and the user on detection accuracy of movement. The acoustic data for each distance is collected from a 6.5-hour real experiment.

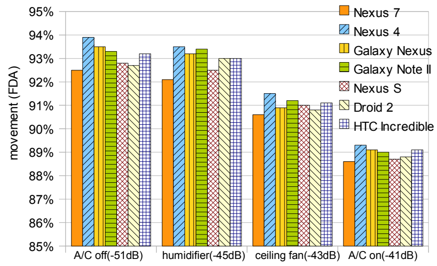


Figure 24: The impact of appliance noise on movement detection accuracy, based on the data from a real real experiment that lasts about 10 hours.

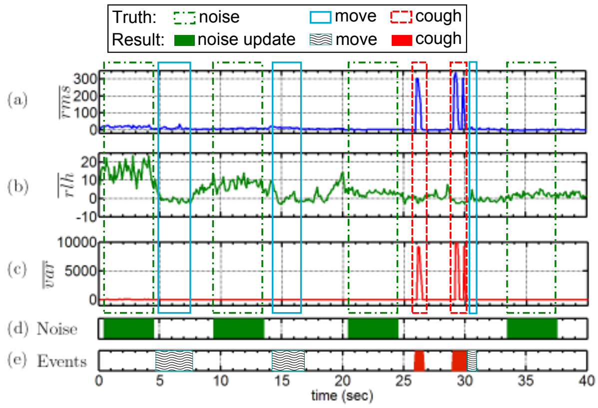


Figure 25: Event detection in the presence of operating A/C. (a), (b) and (c) show the acoustic features over time. (d) shows the detected noise frames that are used to update current noise model. (e) is the detection result.

during short naps without having to charging the phone. We now evaluate the processing time of each component of iSleep and the system energy consumption. The evaluation results based on data collected from 5 devices overnight are shown in Table 6. We can see that the feature extraction component consumes the most processing time among all components, since three features need to be computed for each frame. Thanks to the light-weight decision-tree based classifier, the event detection component only consumes around 0.2% of total CPU time. We also evaluate energy consumption of iSleep based on the battery usage data from the Android system. Since the screen is turned off, computation and microphone sampling are the major sources of power consumption. On average, iSleep consumes around 4% battery per hour (excluding the system consumption of Android). This result suggests that, a fully charged phone running iSleep likely survives a full night of usage without connecting to the charger.

7.4 Evaluation using iSleep App Data

This section presents a preliminary evaluation based on the data of real iSleep users. We collected data from the

Phones	ND	FE	ED	Total	% of CPU
Nexus 7	30ms	38ms	0.15ms	68.15ms	1.7%
Nexus 4	28ms	36ms	0.13ms	64.13ms	1.6%
Nexus S	67ms	88ms	0.27ms	155.27ms	3.9%
G. Nexus	40ms	53ms	0.15ms	93.15ms	2.3%
G. Note II	33ms	40ms	0.15ms	73.15ms	1.8%

Table 6: The average CPU time consumed by different components of iSleep to process 4-second acoustic data. (ND: noise detection, FE: feature extraction, ED: event detection).

Android phones that downloaded and installed iSleep from Google Play Store during the first week after the release of iSleep. Although there were more than 100 installs, as expected, many users opted out the data collection. The information collected include users' ratings on the accuracy of sleep efficiency computed by iSleep as well as the numbers of various events detected during each night (no raw acoustic data was collected). On the screen of monitoring results, iSleep shows a slide bar (see Fig. 17) that allows the user to rate the accuracy of the sleep efficiency measured by iSleep on a scale of 0 (not accurate) to 100 (very accurate). The average of 25 scores on sleep efficiency from users is above 85%. Fig. 26 shows the results of four users randomly chosen from those who participated in the data collection. We can see that, both the total in-bed time and actual sleep time are relatively consistent for the same user, reflecting the user's normal sleep behavior. A detailed analysis of the results also suggests that the shorter sleep time is usually caused by either snoring or extensive body movement. Another observation by correlating the sleep efficiency and user ratings is that, users are more likely to give low feedback scores when the measured sleep efficiency is low.

8. CONCLUSION AND FUTURE WORK

We have described the design, implementation, and evaluation of iSleep – a practical system to monitor an individual's sleep quality using off-the-shelf smartphone. Compared with existing solutions, iSleep is very easy to use and unobtrusive. iSleep uses the built-in microphone of the smartphone to detect the events that are closely related to sleep quality, including body movement, couch and snore, and infers quantitative measures of sleep quality based on actigraphy and Pittsburgh Sleep Quality Index (PSQI). We

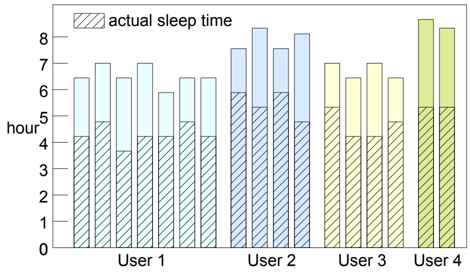


Figure 26: The actual sleep times and in-bed times of 4 iSleep App users.

have evaluated iSleep extensively in a long-term experiment that involves 7 participants and total 51 nights of sleep. Our results show that iSleep achieves above 90% accuracy for sleep-rated event classification in a different settings. The fine-grained sleep profile measured by iSleep also enabled users to track details of sleep events over time and discover irregular sleep patterns.

The high-rate microphone sampling is a major source of energy consumption. We will investigate an adaptive sampling scheme in which the microphone is sampled at a low rate, and only sampled at a higher rate when a potential event is detected. Environmental factors such as room temperature play an important role in quality of sleep. We plan to integrate iSleep with tools that can monitor sleep environments [22]. This will enable in-depth analysis of causes of interrupted sleep and irregular sleep patterns, providing important information for healthcare providers to find trends related to certain diseases.

9. REFERENCES

[1] Fitbit. `http://www.fitbit.com/`.
[2] https://play.google.com/store/apps/.
[3] Sleep as android. `https://sites.google.com/site/sleepasandroid/`.
[4] Sleep cycle. `http://www.sleepcycle.com/`.
[5] Sleep tracker. `http://www.sleeptracker.com/`.
[6] Watch pat. `http://www.itamar-medical.com/WatchPAT.html`.
[7] Zeo. `http://www.myzeo.com/sleep/`.
[8] S. Ancoli-Israel et al. The role of actigraphy in the study of sleep and circadian rhythms. In *Sleep*, volume 26, pages 342–392, 2003.
[9] J. Backhaus et al. Test-retest reliability and validity of the pittsburgh sleep quality index in primary insomnia. *Journal of psychosomatic research*, 53(3):737–740, 2002.
[10] 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 *BodyNets*, 2012.
[11] D. Buysse et al. Quantification of subjective sleep quality in healthy elderly men and women using the pittsburgh sleep quality index (psqi). *Sleep*, 14(4):331–338, 1991.
[12] D. Buysse, C. Reynolds, T. Monk, S. Berman, and D. Kupfer. The pittsburgh sleep quality index (psqi): A new instrument for psychiatric research and practice. In *Psychiatry Research*, volume 28, pages 193–213, 1989.
[13] A. Chesson et al. Practice parameters for the indications for polysomnography and related procedures. In *Sleep*, volume 20, pages 406–422, 1997.
[14] E. K. Choe et al. Opportunities for computing to support healthy sleep behavior. In *CHI*, 2010.
[15] R. Cole et al. Technical note automatic sleep/wake identification from wrist activity. In *Sleep*, volume 15, pages 461–469, 1992.
[16] D. Foley et al. Sleep disturbances and chronic disease in older adults: results of the 2003 national sleep foundation sleep in america survey. *Journal of psychosomatic research*, 56(5):497–502, 2004.
[17] A. Ganapathiraju, J. Hamaker, J. Picone, and A. Skjellum. A comparative analysis of fft algorithms. *the IEEE Transactions on Signal Processing*, 1997.
[18] J. Hedner et al. A novel adaptive wrist actigraphy algorithm for sleep-wake assessment in sleep apnea patients. In *Sleep*, volume 27, pages 1560–1566, 2004.
[19] E. Hoque, R. Dickerson, and J. Stankovic. Monitoring body positions and movements during sleep using wisps. In *Wireless Health*, 2010.
[20] C. Iber et al. The aasm manual for the scoring of sleep and associated events. In *Westchester, IL: American Academy of Sleep Medicine*, 2007.
[21] S. Imai. Cepstral analysis synthesis on the mel frequency scale. In *ICASSP*, volume 8, pages 93–96, 1983.
[22] 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 *UbiComp*, 2012.
[23] J. Liu, W. Xu, M.-C. Huang, N. Alshurafa, and M. Sarrafzadeh. A dense pressure sensitive bedsheet design for unobtrusive sleep posture monitoring. In *PerCom*, 2013.
[24] H. Lu, W. Pan, N. Lane, T. Choudhury, and A. Campbell. Soundsense: scalable sound sensing for people-centric applications on mobile phones. In *MobiSys*, 2009.
[25] H. Lu, M. Rabbi, G. Chittaranjan, D. Frauendorfer, M. Mast, A. Campbell, D. Gatica-Perez, and T. Choudhury. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *UbiComp*, 2012.
[26] T. Mitchell. *Machine Learning*. McGraw-Hill.
[27] J. Parish. Sleep-related problems in common medical conditions. *Chest Journal*, 135(2):563–572, 2009.
[28] M. Rofouei, M. Sinclair, R. Bittner, T. Blank, N. Saw, G. DeJean, and J. Heffron. A non-invasive wearable neck-cuff system for real-time sleep monitoring. In *BSN*, 2011.
[29] A. Sadeh, K. Sharkey, and M. Carskadon. Activity-based sleepąłwake identification: An empirical test of methodological issues. In *Sleep*, volume 17, pages 201–207, 1994.
[30] P. Soille. *Morphological image analysis: principles and applications*. Springer-Verlag New York, Inc., 2003.
[31] J. Webster et al. An activity-based sleep monitor system for ambulatory use. *Sleep*, 5(4):389–399, 1982.