Wu et al. Health Inf Sci Syst (2017) 5:11
https://doi.org/10.1007/s13755-017-0031-z

Health Information Science and Systems

RESEARCH

Statistical sleep pattern modelling for sleep quality assessment based on sound events

Hongle Wu[1*], Takafumi Kato[2], Masayuki Numao[1] and Ken-ichi Fukui[1]

Abstract

A good sleep is important for a healthy life. Recently, several consumer sleep devices have emerged on the market claiming that they can provide personal sleep monitoring; however, many of them require additional hardware or there is a lack of scientific evidence regarding their reliability. In this paper we proposed a novel method to assess the sleep quality through sound events recorded in the bedroom. We used subjective sleep quality as training label, combined several machine learning approaches including kernelized self organizing map, hierarchical clustering and hidden Markov model, obtained the models to indicate the sleep pattern of specific quality level. The proposed method is different from traditional sleep stage based method, provides a new aspect of sleep monitoring that sound events are directly correlated with the sleep of a person.

Keywords: Sleep quality, Sound data, Self-organizing map, Hierarchical clustering, Hidden Markov model

Introduction

Sleep is an important physiological state of the human body. Almost one third of the time in a person's life is spent sleeping. The quality of sleep is very important to a person's health. Therefore, sleep monitoring technology has become an indispensable content in modern personal sleep management [7].

In clinical treatment and research, almost all the methods of sleep quality assessment are based on sleep stages, which are normally scored through the polysomnography (PSG) recordings. PSG is the primary tool for sleep study [9]. PSG monitors body functions through many methods, including electroencephalography for the brain, electrooculography for eye movements, electromyography for muscle activity, and electrocardiography for heart rhythm, and is mainly used in medical science and treatment by doctors [19, 24]. Due to its professional property and financial cost, PSG usage is limited to only clinics. Hence, instead of using PSG to score sleep stages, several works tried to estimate sleep stages through other device recordings, however, the accuracy is not satisfactory [11, 35], a sleep quality assessment based on unfaithful sleep stages estimation is also unreliable. Therefore, recently the quality assessment approaches based on sleep stages are difficult to find a balance between cost and reliability.

On the other hand, as far as we know, many types of sleep disorder are respectively related to a distinctive type of sound, such as snoring, teeth grinding, limb movement and sleep talking. Meanwhile, the environmental sound inside or outside the bed room will also directly impact the sleep. Hence, in this paper we proposed a novel method to assess the sleep quality through sound events recorded in the bedroom. We combined several machine learning approaches including kernelized self-organizing map (SOM), hierarchical clustering (HC) and hidden Markov model (HMM), trained the models to indicate the sleep pattern of specific quality level. The training data was labelled by subjective sleep quality which obtained by self-rating questionnaire fulfilled by experiment subject. The proposed method is different from traditional sleep stage based method, provides a new aspect of sleep monitoring that sound events were directly correlated with one's sleep.

*Correspondence: wuhongle@gmail.com
[1] Department of Architecture for Intelligence, The Institute of Scientific and Industrial Research, Osaka University, Suita, Japan
Full list of author information is available at the end of the article

The main features of our method are as follows:

Non-invasive The sound data can be recorded by any recording device placed near the user's bed during sleep; hence, no burden is added to the user.
No additional cost Any off-the-shelf equipment with a microphone, including a smartphone, recording pen, or personal computer, can be used as the recording device.
Scientifically reliable We collaborated with medical experts in this study, a questionnaire was designed by the experts to evaluate the subjective sleep quality of experiment subjects, which made our training data with sufficient reliability.

In the previous work [41], we applied various kinds of SOM [21] algorithms to the extracted sound clips of sleep-related events to obtain cluster maps, and proved the reliability and feasibility of kernelized SOM for sleep-related sound data analysis. However, one of the problems of their method is cluster detection, even they could find the best match unit on the cluster map for every input data, it is difficult to find out the major clusters on the cluster map without manually annotation of the input data, which will cost lots of time. In this paper, we applied HC on the cluster map from Kullback–Leibler kernel SOM (KL-KSOM) [13], HC is a method that seeks to build a hierarchy of clusters, builds nested clusters by merging or splitting them successively. We calculated the distances between cells on cluster map, and detected the hierarchical structure of cells by HC. According to the property of SOM, by setting an appropriate metric on cells splitting, the cells were divided into several major clusters, and each major cluster of the cells mainly indicates a different kind of sleep related events, also every input vector can be assigned to a Best Match Unit (BMU, the nearest cell to the input vector) on the cluster map. Therefore this divided cluster map can be used as a virtual classifier for input sound event. We call this classifier the virtual classifier since it assigns input data into a certain cluster. Although we do not know the exact event type for each cells cluster, but the output from this virtual classifier is necessary and sufficient to form a categorized data sequence for the following HMM modelling.

In this study, the data set for experiments was consisted of 36 whole night sound recordings with 18 in good quality and 18 in poor quality, and we classified sound events that extracted from these recordings by the aforementioned virtual classifier. After we got these categorized sound events sequences which represents sleep pattern, the HMMs of good and poor sleep quality were trained respectively. Generally, HMM is used for structured predictions, for example on sequence data like speech [31], protein [38]. Also, there are works for classification with HMM [25]. In this study, the HMMs of good and poor sleep quality were used to predict the sleep quality. The likelihoods between an input sound event sequence and HMMs are calculated as input vectors, then several classification methods are applied, including support vector machines (SVM), adaptive Boosting (Adaboost), majority decision, etc.

We verified our method by 10-fold cross validation, the results revealed this novel approach of sleep quality assessment is feasible.

Methodology

Overview

In this section, we introduce the key methodologies applied in this study. Our method process includes following steps:

Sound recording Recording sound data by recording device, and converting sound format data to text format data through sound processing software.
Self-rating questionnaire In the morning, each subject fulfilled a self-rating questionnaire including questions about sleep quality of last night.
Data labelling The sound recordings were labelled by good or poor sleep quality based on the answers from the questionnaire, which are used only in training phase.
Events extraction Sound clips of events were extracted from the sound recordings, burst extraction algorithm ("Burst extraction algorithm 2.2" section) is applied.
Input data preprocessing Applying FFT to obtain the frequency power spectrum of each sound clips as input vector.
Clustering Using the frequency power spectrum of each data point, which is a vector of discretized frequencies as input vectors, we applied KL-KSOM ("Clustering by KL-KSOM" section 2.3) to get the cluster map. Then agglomerative HC ("Categorizing by HC" section 2.4) was used on the cells of the cluster map to reflect hierarchical structure of the map. Two steps of KL-KSOM and HC is effective, as KL-KSOM firstly captures the manifold of data distribution in the high dimensional spectrum feature space, which is very complex, and convert into simple two dimensional space which preserves the data distribution as much as possible that makes easier to identity a few numbers of major clusters (event type) by HC.
Sound events categorizing By selecting an appropriate stop-criteria for agglomerative HC through silhouette [34], the cells was divided into several major clusters, thus we obtained a virtual classifier as aforementioned. Classification was performed on all extracted

sound events, then a sequence with categorized data points was obtained for each sound recording.

Modelling by HMM According to the sleep quality labels on sound recordings, data sequences obtained from last step were divided into good and poor data sequence sets. The multinomial HMMs ("Modelling by hidden Markov model 2.5" section) for good or poor sleep quality were trained respectively by corresponding data sequence set.

Classification based on HMMs The likelihoods between an input sound event sequence and obtained HMMs are used as input data for sleep quality level classification ("Classification based on HMMs 2.6" section)

Evaluation 10-fold cross validation was used to evaluate the accuracy of classification.

Burst extraction algorithm

The first step to be followed after recording the sound is to determine the useful events inside an all-night-long sound recording. Manually searching the events will waste considerable time and is definitely unacceptable. In this study, we used the method in [13] to differentiate the steady noise from other types of sound events including sleep disorder symptoms, such as snoring, teeth grinding, or body movement, and environmental sound, such as air-conditioner operation or outdoor traffic. The sound events were extracted by the statistical burst extraction method [20].

By using Kleinberg's method, we no longer need to consider the size of the sliding window or amplitude threshold. Furthermore, by introducing the cost function, this method can extract an event that has been broken apart due to brief gaps during a single event; threshold methods are basically unable to perform this extraction.

Let z_t $(t = 1, \ldots, t_{end})$ be the amplitude at time t, and sound signals are assumed to be generated from a Gaussian probability density function:

$$f_j(z_t) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{(z_t-\mu)^2}{2\sigma_j^2}\right\} \quad (j = 0, \ldots, L), \tag{1}$$

where $\mu = \sum_t z_t/t_{end}$ is the mean for all sound signals, σ_0(steady state) is the variance of all sound values, and $\sigma_j = s^j\sigma_0(j \geq 1$, burst state). Here, $s > 1$ is a parameter that controls the resolution of bust levels. This model assumes that the different burst levels of the sound signals are generated by the different variances of the Gaussian functions. Let $\mathrm{Cost}_j(t)$ be a necessary cost for z_t to be state j, then the burst extraction algorithm is as follows:

Step 1 Initialize costs at $t = 0$ as $\mathrm{Cost}_j(0) = 0$ $(j = 0)$ and $\mathrm{Cost}_j(0) = \infty\,(j \geq 1)$.

Step 2 $t \to t + 1$.

Step 3 Calculate $\mathrm{Cost}_j(t)$ for $j = 0, \ldots, L$ by the following equation:

$$\mathrm{Cost}_j(t) = -\ln f_j(z_t) + \min_{0\leq l\leq \nu}\{\mathrm{Cost}_l(t-1) + \tau(l,j)\}, \tag{2}$$

where j is a state at t and l is a state at $t-1$. In addition, $\tau(l,j)$ is the transition cost from state l to j given by

$$\tau(l,j) = \begin{cases} (j-l)\gamma \ln t_{end} & \text{if } j > l \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $\gamma > 0$ is a parameter that controls the effect of transition cost.

Step 4 Continue Steps 2 and 3 until $t = t_{end}$.

Step 5 Estimate the optimal state sequence that gives the minimum cost using the Viterbi algorithm. The Viterbi algorithm traces in the reverse direction from the last signal t_{end}, i.e., the Viterbi algorithm starts from $state^*(t_{end}) = \arg\min_{0\leq j\leq L} \mathrm{Cost}_j(t_{end})$, and is iterated repeatedly until $t = 1$, choosing a previous optimal state as $state^*(t-1)$, which gives the current optimal state $state^*(t)$.

After calculating the optimal burst levels, sound events are obtained by extracting areas where the burst level is greater than 1 $(j \geq 1)$.

Clustering by KL-KSOM

The SOM [22] is an artificial neural network and originally a model of associative memory, but has recently been widely used for visual data mining, for example, in exploratory analysis support of documents [23], for the monitoring of machinery [37], and for application to medical care or economics. Data distribution in the high dimensional feature space can be captured by SOM and converted into low dimensional space, which makes it easier to identity a few numbers of major clusters among the data.

In this study, we used the frequency spectrum as input vector. However, the standard SOM uses Euclidean distance as a similarity measure of data points, so the distribution structure of a frequency spectrum cannot be captured since each discrete point is treated as an independent variable. The authors in [13] proposed the use of Kullback–Leibler (KL) divergence to introduce a distribution structure into a similarity measure of frequency spectrum of acoustic emission events and obtained a good effect. In this study, KL-KSOM was used to cluster the sleep-related sound events. Moreover, each spectrum data is normalized as $\sum_{k=1}^{\nu} x_k = 1$ for the requirement of the KL kernel.

The KL kernel function is defined as:

$$K_{KL}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp(-\beta JS(\boldsymbol{x}_i, \boldsymbol{x}_j)), \tag{4}$$

$$\begin{aligned} JS(\mathbf{x}_i, \mathbf{x}_j) &= KL(\mathbf{x}_i, \mathbf{x}_j) + KL(\mathbf{x}_j, \mathbf{x}_i) \\ &= \sum_{k=1}^{v} \left\{ x_{i,k} \log \frac{x_{i,k}}{x_{j,k}} + x_{j,k} \log \frac{x_{j,k}}{x_{i,k}} \right\}, \end{aligned} \tag{5}$$

where $KL(\mathbf{x}_i, \mathbf{x}_j)$ is the KL divergence, which is the distance between probability distributions, $JS(\mathbf{x}_i, \mathbf{x}_j)$ denotes the Jensen-Shannon divergence, which symmetrizes the KL divergence, and $\beta > 0$ is a scaling parameter.

The basic concept of the kernel SOM is the same as that of the SOM. However, in the kernel SOM, the reference vector is updated in an indirect manner because the reference vector in the mapped space cannot be calculated.

By replacing $\mathbf{x}$ in the updating formula of a reference vector in the standard batch type SOM by a mapping $\phi(\mathbf{x})$, the following updating formula can be obtained:

$$\mathbf{m}_i(t+1) := \gamma \sum_{n} h_{c(\mathbf{x}_n),i} \phi(\mathbf{x}_n), \tag{6}$$

where t is an iteration step, and γ is a regularization term $\gamma = 1/\sum_j^n h_{c(\mathbf{x}_j),i}$. However, since $\phi(\mathbf{x}_n)$ cannot be calculated, the ith reference vector is updated using the dissimilarity to all data points $\forall n\ d_{i,n}$, as follows:

$$\begin{aligned} d_{i,n}(t+1) &\equiv ||\phi(\mathbf{x}_n) - \mathbf{m}_i(t+1)||^2 \\ &= K(\mathbf{x}_n, \mathbf{x}_n) - 2\gamma \sum_{j}^{n} h_{c(\mathbf{x}_j),i} K(\mathbf{x}_n, \mathbf{x}_j) \\ &\quad + \gamma^2 \sum_{k}^{n} \sum_{l}^{n} h_{c(\mathbf{x}_k),i} h_{c(\mathbf{x}_l),i} K(\mathbf{x}_k, \mathbf{x}_l). \end{aligned} \tag{7}$$

The following describes the algorithm of the batch type KL-SOM:

Step 1 Initialize all dissimilarity between reference vectors and data points $\forall i, n\ d_{i,n}$ randomly and set the iteration step as $t = 1$.
Step 2 Search the best matching units $\{c(\mathbf{x}_1), \ldots, c(\mathbf{x}_N)\}$ for all inputs by the nearest neuron:

$$c(\mathbf{x}_n) = \arg \min_{i=1,\ldots,M} d_{i,n}, \tag{8}$$

Step 3 Exit if the best matching units $\{c(\mathbf{x}_1), \ldots, c(\mathbf{x}_N)\}$ were not changed or the iteration reached $t = t_{max}$.
Step 4 Update the dissimilarity of each reference vector to all inputs $\forall n\ d_{i,n}$ by Eq. (7).
Step 5 Decrease the neighborhood radius σ and increase the iteration counter $t \to t + 1$. Then, return to Step 2.

Categorizing by HC

HC algorithms organize a data set into a hierarchical structure according to a similarity measure. It is based on the belief that nearby objects are more related than objects that are farther away [33]. HC is applied in this study instead of other methods like K-Means because it is typically used to obtain major clusters in SOM [40].

HC algorithms connect objects based on their similarity to form clusters, which is usually represented using a dendrogram. HC algorithms differ in the choice of similarity measures, the linkage criterion (distance between clusters), and whether the process is agglomerative (bottom-up) or divisive (top-down). Agglomerative HC starts with singleton clusters and then recursively merges appropriate clusters, and divisive HC starts with one cluster containing all objects and recursively splits appropriate clusters [3].

Since the kernel function was introduced into the KL-KSOM, the similarity between cells on the cluster map is unable to be calculated. In this work, the similarity between cell a and b is calculated by the following formula:

$$\begin{aligned} d_{a,b} \equiv ||\mathbf{m}_a - \mathbf{m}_b||^2 &= \gamma_a^2 \sum_{i}^{n} \sum_{j}^{n} h_{c(\mathbf{x}_i),a} h_{c(\mathbf{x}_j),a} K(\mathbf{x}_i, \mathbf{x}_j) \\ &\quad - 2\gamma_a \gamma_b \sum_{k}^{n} \sum_{l}^{n} h_{c(\mathbf{x}_k),a} h_{c(\mathbf{x}_l),b} K(\mathbf{x}_k, \mathbf{x}_l) \\ &\quad + \gamma_b^2 \sum_{u}^{n} \sum_{v}^{n} h_{c(\mathbf{x}_u),b} h_{c(\mathbf{x}_v),b} K(\mathbf{x}_u, \mathbf{x}_v), \end{aligned} \tag{9}$$

where $\mathbf{x}_i (i = 1, \ldots, n)$ is all the input vector training the KL-KSOM map, γ is a regularization term $\gamma_n = 1/\sum_j^n h_{c(\mathbf{x}_j),i}$, and h, K is same as in Eq. (7).

After the similarities between cells in KL-KSOM cluster map were calculated, we applied agglomerative HC algorithm with ward criterion. By selecting an appropriate stop-criteria for agglomerative HC through silhouette [34], the cells was divided into several major clusters. According to the property of SOM, we assumed that each major cluster of the cells mainly indicates a different kind of sleep related events.

According to SOM algorithm, every input vector can be assigned to its BMU on the SOM cluster map, which is the nearest cell to the input vector. In this work, because of the kernelization, instead of traditional Euclidean distance, besides n input vectors $\mathbf{x}_i (i = 1, \ldots, n)$ training the KL-KSOM map, the similarity from a new input vector $\mathbf{x}_{n+1}$ to the cell i on the map will be calculated as follow:

$$\begin{aligned} d_{i,n+1} &\equiv ||\phi(\mathbf{x}_{n+1}) - \mathbf{m}_i||^2 \\ &= K(\mathbf{x}_{n+1}, \mathbf{x}_{n+1}) \\ &\quad - 2\gamma \sum_{j}^{n+1} h_{c(\mathbf{x}_j),i} K(\mathbf{x}_{n+1}, \mathbf{x}_j) \\ &\quad + \gamma^2 \sum_{k}^{n+1} \sum_{l}^{n+1} h_{c(\mathbf{x}_k),i} h_{c(\mathbf{x}_l),i} K(\mathbf{x}_k, \mathbf{x}_l), \end{aligned} \tag{10}$$

where γ is a regularization term $\gamma = 1/\sum_j^{n+1} h_{c(\mathbf{x}_j),i}$, and h, K is same as in Eq. (7).

Then, the new input vector can be assigned to the major cluster that the BMU belonging to as well, where BMU of the new input can be calculated by:

$$c(\mathbf{x}_{n+1}) = \arg\min_{i=1,\ldots,M} d_{i,n+1}. \tag{11}$$

Therefore the virtual classifier for input sound event was created. Classification was performed on all extracted sound events, then a sequence with categorized data points was obtained for each sound recording. Although we do not know the exact event type for each cells cluster, but the output from this virtual classifier is necessary and sufficient to form a categorized data sequence for the following HMM to generate a model to indicate the characteristic of sleep, also as known as sleep pattern in this study.

Modelling by hidden Markov model

The HMM is a generative probabilistic model, in which a sequence of observable $\mathbf{X}$ variable is generated by a sequence of internal hidden state $\mathbf{Z}$. The hidden states can not be observed directly. The transitions between hidden states are assumed to have the form of a (first-order) Markov chain. They can be specified by the start probability vector $\mathbf{5}$ and a transition probability matrix $\mathbf{A}$. The emission probability of an observable can be any distribution with parameters $\mathbf{\Theta}_i$ conditioned on the current hidden state (e.g. multinomial, Gaussian). The HMM is completely determined by $\mathbf{5}, \mathbf{A}$ and $\mathbf{\Theta}_i$ [31].

According to the sleep quality labels on sound recordings, data sequences obtained from KL-KSOM and HC were divided into good and poor data sequence sets. Multinomial HMM was applied on these data sequences to study the sleep pattern since it is an appropriate tool for sequence data modelling, also the likelihood between a model and an observed sequence is a proper metric on the similarity comparison of time series data. In this study, the likelihood was calculated by the log-likelihood function.

In our work, determining the number of hidden states of HMM is challenging, theoretically it should refer the number of different state of sleep. As we know, sleep occurs in cycles [30], proceeds in cycles of rapid eye movement (REM) and Non-REM (NREM). The American Academy of Sleep Medicine (AASM) divides NREM into three stages: N1, N2, and N3 [36]. However, the distinctions between these sleep stages are somewhat arbitrary, and the physiological boundaries between them are blurred and continuous. Hence, it is difficult to determine the exact number of hidden states of HMM. In order to improve the accuracy of classification, we trained HMMs on different numbers of hidden state, including 2, 3, 4 and 5 hidden states. In other words, we obtained 4 HMMs for good sleep and other 4 for poor sleep, 8 HMMs in total. During the experiment, we found out that the best number of hidden states is different modelling good or poor sleep quality.

Classification based on HMMs

To classify the sleep quality level of a new obtained sound recording, we firstly extract the clips of sleep-related sound events, categorize each events and form as a data sequence. Then we calculate likelihoods between the data sequence and HMMs obtained from previous section. In this study, several classification methods are applied:

SVM SVM [10] with two different kinds of input data are used in this study: 1. Likelihoods between input data sequence and 8 HMMs formed a 8-dimensional vector as input; 2. Event counts on three major clusters formed a 3-dimensional vector as input. The latter is a typical framework in classification. We applied event counts vector as input to make a comparison and demonstrate the significance of time sequential property in the sleep quality assessment.

Adaptive Boosting (*Adaboost*) Adaboost [12] applied same likelihoods vector input as SVM, with decision trees as the weak learners.

Majority decision The easiest way of determine the class of a data sequence is comparing its likelihoods to two HMMs from different sleep quality level, and choosing the greater side. Since it is difficult to determine the hidden state number of HMMs, we decided to make this comparison on 3, 4 and 5 hidden state HMMs respectively and choose the final class by majority decision. For example, if a data sequence was close to poor sleep HMM on 3 hidden state but close to good sleep HMMs on 4 and 5 hidden state, it will be classified as good.

Likelihood summations We simply sum likelihoods from 2, 3, 4 and 5 hidden states HMMs of good or poor sleep quality respectively, and choose the greater side.

Experiment

Overview

We first applied the KL-KSOM to the extracted sound data, obtained the cluster map as result. Then HC was

Table 1 Questionnaire for sleep quality

	Question	Answer options				
1	How long it took until falling asleep last night comparing to usual?	**A** Very long	**B** Long	**C** Same	**D** Short	**E** Very short
2	How many times you woke up last night comparing to usual?	**A** Very many	**B** Many	**C** Same	**D** few	**E** Very few
3	The sleep duration of last night comparing to usual.	**A** Very long	**B** Long	**C** Same	**D** Short	**E** Very short
4	How was the sleep depth of last night comparing to usual?	**A** Very deep	**B** Deep	**C** Same	**D** Light	**E** Very light
5	Overall, how was the sleep of last night comparing to usual?	**A** Very good	**B** Good	**C** Same	**D** Poor	**E** Very poor

applied on the cells in the cluster map to get the hierarchical structure of cells. By selecting an appropriate stop-criteria for agglomerative hierarchical clustering by silhouette coefficient, the cells was divided into several major clusters, and we obtained a virtual classifier as aforementioned for sleep sound events. After getting this classifier, classification was performed on all extracted sound events, then for every night's sound recording, a sequence with categorized data points was obtained.

The data sequences were labelled as good or poor sleep according to the subjective sleep quality from questionnaire, trained the HMMs for good or poor sleep quality respectively by corresponding data sequences.

In the end, we built several sleep quality classifiers based on these HMMs and evaluated the performance via 10-fold cross validation.

Experimental setting

The data used in this study were prepared by the Graduate School of Dentistry of Osaka University. The study protocol was approved by the research ethics committee. Written informed consent was obtained from all subjects. All subjects were asked to sleep in a specific room (Fig. 1) from 22:30 to 6:30. The recording device included LA1250 (Ono Sokki)[1] and R-4 Pro (Roland).[2] A microphone was placed at a distance of 50 cm from the subjects heads. The sound data were recorded on a single channel (mono) at a sampling rate of 48 kHz. In addition, all subjects were measured by PSG simultaneously.

Subjects were asked to fulfill a self-rating questionnaire after waking up during the experiment. The questions regarding sleep quality are showed in Table 1. Questions 1–4 are general sleep quality evaluation criteria, question 5 is the overall self-rating of sleep quality. Based on the statistics on these questionnaires, we found that subjects who answered "Very good" or "Good" on question 5 are more likely to had a short falling asleep period, few awaking times, long sleep duration and deep sleep depth, and vice versa, this is consistent with the observation of the factors affecting the quality of sleep in medicine [39]. Based on this founding, the sound recordings from subject with answer "Very good" or "Good" for the question 5 were regarded as good sleep data, on the contrary, recordings with "Poor" or "Very poor" were regarded as poor sleep data. Based on this rule, we selected 36 sound recordings from 36 different subjects with 18 in good quality and 18 in poor quality. The age range of these subjects is 20–29, and gender distribution was balanced.

The sound data is stored and processed on a Linux server with two Intel Xeon 12-Core 2.7GHz CPU and 128GB memory.

Event extraction

Based on the burst extraction method, from 36 recordings, we obtained a total of 39,105 sound events, with hyper-parameters of $L = 6$, $s = 1.5$, and $\gamma = 100$. The hyper-parameters were tuned manually in a certain range to extract as more useful events as possible, and keep the extracted useless noise in an acceptable amount. FFT was applied to the extracted sound data to obtain the frequency power spectrum. From 20 Hz to 20 kHz, at intervals of 20 Hz, 1000 discretized points as an input for KL-KSOM were obtained for every sound data. Figure 2

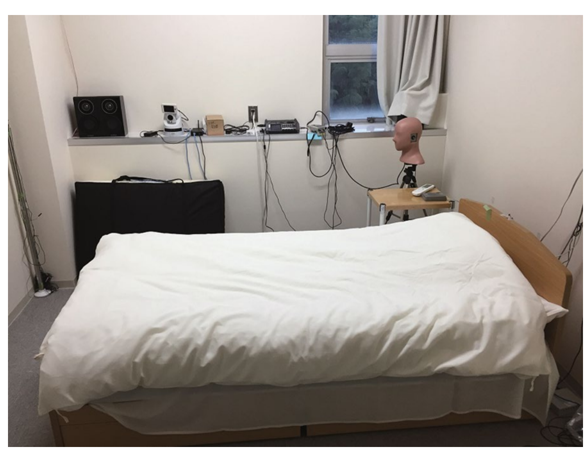

Fig. 1 Experiment room

[1] https://www.onosokki.co.jp/English/hp_e/products/keisoku/s_v/la1200.html.

[2] http://proav.roland.com/products/r-4_pro/.

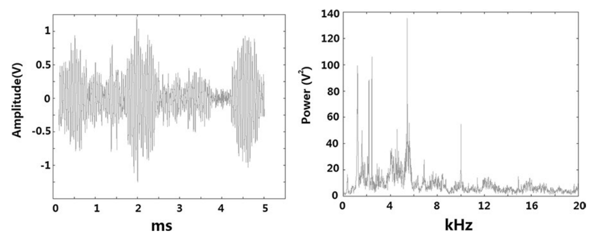

Fig. 2 An example of an extracted sound wave and frequency spectrum

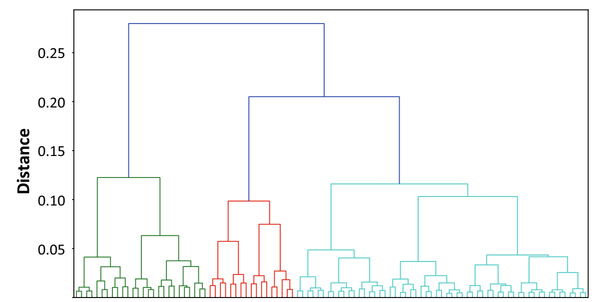

Fig. 3 Dendrogram by HC

shows an example of an extracted sound event and its preprocessed frequency spectrum.

Sound events categorizing by KL-KSOM and HC

In the first part of this experiment, 5000 extracted data were randomly selected and combined into one dataset for KL-KSOM training. The number of cells was set to 10×10 with a two-dimensional regular grid. In general, the number of neurons is not sensitive to these results, in that SOM captures the data distribution in the feature space.

After obtained the KL-KSOM cluster map, the similarities between cells on the map were calculated by Eq. (9). Then, we applied agglomerative HC algorithm with Ward's criterion, the dendrogram is shown as Fig. 3. The stop-criteria for agglomerative HC was determined through silhouette, the silhouette coefficient on different stop-criteria is showed in Fig. 4. In this experiment, 0.15 was selected as the stop-criteria, thus cells was divided into three clusters as shown in Fig. 5.

We determined BMU for extracted sound events through Eq. (11), each sound event was then assigned to one of the three clusters. As aforementioned, we assumed that each cluster mainly indicates a different kind of sleep related events, we checked several specific events we observed from sound recordings, and found similar events were mainly categorized into same cluster. Therefore, all the events with time stamps extracted from one same sound recording formed a categorized data sequence to indicate that subject's sleep pattern.

Sleep quality classification by HMM

According to sleep quality level from self-rating questionnaire, the data sequences obtained from last step were divided into good and poor quality sequence sets, each containing 18 sequences. We trained multinomial HMMs with 2, 3, 4 and 5 hidden states for good or poor sleep quality respectively by corresponding sequence set, in other words, 4 pairs of HMMs were generated. The likelihoods between a new input data sequence and HMMs were calculated through the log-likelihood function. Radial basis function kernel was used in SVM and we tuned the hyper-parameter of the kernel through nested cross-validation with grid search approach.

To evaluate the classifiers, 10-fold cross validation was performed. The results shown in Table 2 revealed this novel approach of sleep quality assessment is feasible as we achieved 77% accuracy in maximum, and SVM used likelihoods vector as input made a significant improvement in accuracy. Also, we checked the accuracy of SVM method with 2-dimensional input vector from 2, 3, 4, 5 and 6 hidden states HMMs respectively, and as shown in Table 3, we found out data with good sleep quality got best accuracy on 5 hidden states models and poor ones on 3 hidden states models. The highest scores are highlighted with bold fonts. According to the experiment result, we found that SVM with likelihoods as input performed much better than the one with event counts as input, which indicated time sequential property is important in quality assessment of sleep.

The matrices of transition probabilities of 3 and 5 hidden states HMMs on sound event are shown as Fig. 6a and sleep stage as Fig. 6b. Sleep stages were scored by medical experts based on PSG data recorded simultaneously, and HMMs of sleep stage sequences were trained for the comparison to that of sound events. However, we found there is no significant difference on sleep stage sequence HMMs between good and poor sleep quality (Fig. 6b). On the contrary, the HMMs of sound events from different sleep quality level have obvious difference (Fig. 6a). This evidence is interesting that sleep stage sequence is useless for assessing sleep quality.

By comparing good and poor models of 5 hidden state HMM on sound events (Fig. 6a), the good model is stable as self-loop probabilities are high. Also some transitions are completely do not appear. These properties are reasonable from the aspect of sleep science (e.g., N3 $\rightarrow$ Wake does not happen), this property also appears in HMM on sleep stage sequence (Fig. 6b). In contrast, transition probabilities in the poor model on sound event are varied, which implies poor sleep do not have specific sleep pattern related to sounds.

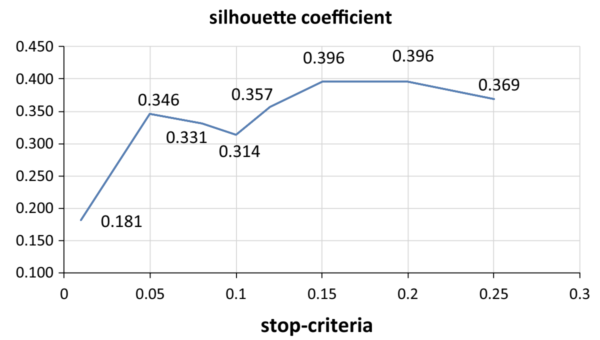

Fig. 4 Silhouette coefficient on different stop-criteria

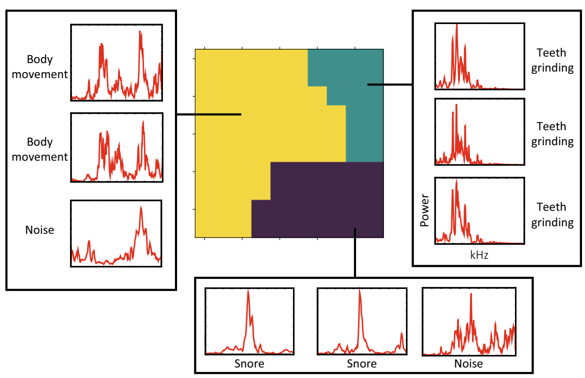

Fig. 5 Major clusters on KL-KSOM cluster map with frequency spectrum of event examples from each cluster

Discussion

The age range of the subjects is not general since all of subjects are university students, but with the scope of data collection enlarging, this problem will be solved. Furthermore, sometimes there is a long time interval between two sound events, it usually happened on quiet subjects, in the future work, we will try to insert virtual events into these intervals, we assume it will make the distribution of events on the timeline more balanced and the entire sleep process can be reflected more accurately.

Table 3 SVM classification accuracy by input data from different hidden states number HMMs

Test data	Number of hidden states				
	2	3	4	5	6
Good quality data	0.572	0.693	0.722	**0.757**	0.722
Poor quality data	0.667	**0.754**	0.667	0.652	0.652
Mean	0.619	0.723	0.694	0.704	0.687

Currently we are still focus on single user application, for multi-user scenarios, also known as "Cocktail Party Problem [17] for sound based method, it can be solved by place multiple devices on different place in the room, for example: both sides of the bed, and extract different sound sources based on aspect and phase difference.

Related work

Regarding sleep quality assessment, Pittsburgh Sleep Quality Index (PSQI), a self-report questionnaire, is a popular method that assesses sleep quality over a 1-month time interval [6]. However, the limitation is obvious, the variation of scores is highly dependent on the subject completing them, also as a relatively new measure, it has not received enough investigation to determine the entirety of the psychometric measures [28].

According to American Academy of Sleep Medicine, the sleep stage scoring based on PSG has long been considered as the "gold standard" of sleep study [4]. 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. There have been a handful of investigations of the correlation between perceived sleep quality and PSG-based sleep stage [1, 5, 18, 32]. There are some consensuses from these researches, for example: poor sleep quality estimates are associated with reduced Stage N1 and more Stages N3. However, in these researches, sleep quality was still assessed based on sleep stage scoring, the direct correlation between physiological signals and sleep quality has not been established.

Table 2 Classification accuracy of different methods

Method	Mean accuracy		
	Total data	Good sleep quality data	Poor sleep quality data
SVM on likelihood	**0.775**	**0.767**	**0.783**
SVM on event count	0.483	0.531	0.435
Adaboost	0.615	0.583	0.647
Majority decision	0.722	0.757	0.687
Likelihood summations	0.694	0.667	0.722

a HMMs trained on sound events

Good — 5 hidden states models

From \ To	1	2	3	4	5
1	0.9496	0	0.0432	0.0071	0
2	0	0.7241	0.2394	0.0365	0
3	0.0211	0.131	0.8414	0	0.0065
4	0	0	0.0036	0.9941	0.0023
5	0	0	0.0336	0	0.9664

Good — 3 hidden states models

From \ To	1	2	3
1	0.6228	0.0083	0.3689
2	0.0108	0.9819	0.0073
3	0.5915	0.0092	0.3993

Poor — 5 hidden states models

From \ To	1	2	3	4	5
1	0.3924	0	0.363	0	0.2439
2	0.018	0.5408	0.0024	0.4382	0
3	0.1403	0.0127	0.8113	0	0.0357
4	0.0062	0.4514	0.011	0.5312	0
5	0.1646	0	0.1305	0.0024	0.7025

Poor — 3 hidden states models

From \ To	1	2	3
1	0.8731	0.0193	0.1076
2	0.024	0.9757	0
3	0.1511	0.0019	0.847

b HMMs trained on sleep stage sequences

Good

From \ To	1	2	3	4	5
1	0.7915	0.2085	0	0	0
2	0	0.7432	0.2483	0.0085	0
3	0.0216	0.0423	0.905	0.0031	0.028
4	0.0208	0	0	0.9792	0
5	0.0059	0.016	0.0166	0.0176	0.9439

Poor

From \ To	1	2	3	4	5
1	0.7728	0.1742	0.053	0	0
2	0	0.8915	0.0669	0.0405	0.0011
3	0	0.0359	0.9641	0	0
4	0.0602	0	0	0.9398	0
5	0	0.0165	0.0086	0.0259	0.949

Fig. 6 Transition probability matrices of HMMs. **a** HMMs trained on sound events. **b** HMMs trained on sleep stage sequences

Besides PSG, in the academic field of sleep analysis, various studies using other methods trying to simplify the operation, such as infrared thermography [14], water filled mat [29] and Kinect [27] have been proposed. These methods still require additional professional equipment to record the sleep data and specialized knowledge to use the equipment; the data collection work is limited within the scope of medical specialists. Our method, by contrast, can be applied through any off-the-shelf sound recording device including a smartphone or a personal computer, therefore greatly reduces the cost of data collection and making large-scale data collection possible.

Currently, there are many products on the market that aim to make sleep assessment portable at a reduced cost. ZEO[3] is a popular PSG-based home sleep analysis product. Besides traditional PSG, actigraphy has also been used as an alternative tool; there are many actigraphy-based products including Beddit[4] and Fitbit.[5] The accuracy of these devices is still controversial, according to [26], medical experts do not suggest to use the results from these consumer equipment for medical research, which means they are not reliable enough; authors in [35] made comparisons between PSG scored sleep stages and outputs of several consumer sleep devices, which showed high degree of inconsistency; similar discussion can also be found in [11]. Another problem of these products is that they are invasive to users, which means that users have to wear an additional device or place a device on their bed during sleep. According to a recent survey, many people are resistant to wearing a device during sleep [8]. Even if users accept to wear the device, it is not easy to properly place the sensors in the correct position.

Moreover, additional devices add extra financial burden to the user. The efforts in the market to reduce the cost are mostly through mobile apps. Mobile apps use a smartphone's built-in sensors, and hence, users do not need to purchase additional hardware. There are some academic publications regarding smartphone application for sleep analysis. Gu et al. proposed a method for scoring sleep quality by a smartphone application named Sleep Hunter [15], and Hao et al. developed an application called iSleep [16]. Gu used not only sound data but also data from the accelerometer and light sensor, which limited the range of the available equipment. Hao used only sound data; however their ground truth is another high-quality sound data, which lacks medical reliability. Currently, neither Sleep Hunter nor iSleep can be found in any application store. Moreover, we investigated two popular applications: Sleep as Android[6] and Sleep Cycle alarm clock[7]; however, no academic proof or accuracy evaluation for their outputs exists, which is consistent with [2], that the authors mentioned very few of the apps are based on published scientific evidence.

Conclusion

In this paper we proposed a novel approach to assess the sleep quality through sound data. We combined several machine learning approaches including kernelized SOM, hierarchical clustering and HMM, obtained the models to indicate the sleep pattern of specific quality level. The

[3] https://en.wikipedia.org/wiki/Zeo,_Inc.

[4] http://www.beddit.com/.

[5] https://www.fitbit.com/.

[6] http://sleep.urbandroid.org/.

[7] http://www.sleepcycle.com/.

proposed method is different from traditional sleep stage based method, provides a new aspect of sleep quality assessment.

According to the experiment, the classifier by HMMs obtained a feasible result, which empirically warrants our approach on the assessment of personal sleep quality by sound data. In the future work, we will try to further improve the accuracy of our method and integrate it into smartphone application for daily use.

Author details
[1] Department of Architecture for Intelligence, The Institute of Scientific and Industrial Research, Osaka University, Suita, Japan. [2] Department of Oral Physiology, Graduate School of Dentistry, Osaka University, Suita, Japan.

Acknowledgements
This research is partially supported by the Center of Innovation Program from Japan Science and Technology Agency, JST, the Grant-in-Aid for Scientific Research (B)(#25293393) from the JSPS, and Challenge to Intractable Oral Diseases from Osaka University Graduate School of Dentistry.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 25 July 2017 Accepted: 16 October 2017
Published online: 30 October 2017

References

1. Åkerstedt T, Billiard M, Bonnet M, Ficca G, Garma L, Mariotti M, Salzarulo P, Schulz H. Awakening from sleep. Sleep Med Rev. 2002;6(4):267–86.
2. Behar J, Roebuck A, Domingos JS, Gederi E, Clifford GD. A review of current sleep screening applications for smartphones. Physiol. Meas. 2013;34(7):R29–46.
3. Berkhin P. A survey of clustering data mining techniques. In: Grouping multidimensional data. Berlin: Springer; 2006. p. 25–71.
4. Berry RB, Brooks R, Gamaldo CE, Harding SM, Marcus C, Vaughn B. The AASM manual for the scoring of sleep and associated events. Rules, Terminologyand Technical Specifications. Darien, IL: AmericanAcademy of Sleep Medicine; 2012.
5. Bonnet MH, Johnson LC. Relationship of arousal threshold to sleep stage distribution and subjective estimates of depth and quality of sleep. Sleep. 1978;1(2):161–8.
6. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28(2):193–213.
7. Chen Z, Lin M, Chen F, Lane ND, Cardone G, Wang R, Li T, Chen Y, Choudhury T, Campbell AT. Unobtrusive sleep monitoring using smartphones. In: 2013 7th International conference on pervasive computing technologies for healthcare and workshops. IEEE; 2013. p. 145–52.
8. Choe EK, Kientz JA, Halko S, Fonville A, Sakaguchi D, Watson NF. Opportunities for computing to support healthy sleep behavior. In: CHI'10 extended abstracts on human factors in computing systems. New York: ACM; 2010. p. 3661–6.
9. Chokroverty, S. Sleep disorders medicine: basic science, technical considerations, and clinical aspects. Boston: Butterworth-Heinemann; 2013.
10. Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–97.
11. Dickinson DL, Cazier J, Cech T. A practical validation study of a commercial accelerometer using good and poor sleepers. Health Psychol Open 2016;3(2):2055102916679012.
12. Freund Y, Schapire RE. A desicion-theoretic generalization of on-line learning and an application to boosting. In: European conference on computational learning theory. New York: Springer; 1995. p. 23–37.
13. Fukui K, Akasaki S, Sato K, Mizusaki J, Moriyama K, Kurihara S, Numao M. Visualization of damage progress in solid oxide fuel cells. J Environ Eng. 2011;6(3):499–511.
14. Fukumura H, Okada S, Makikawa M. Estimation of sleep stage using SVM from noncontact measurement of forehead and nasal skin temperature. BME. 2012;50(1):131–7. doi:10.11239/jsmbe.50.131.
15. Gu W, Yang Z, Shangguan L, Sun W, Jin K, Liu Y. Intelligent sleep stage mining service with smartphones. In: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing. New York: ACM; 2014. p. 649–60.
16. Hao T, Xing G, Zhou G. isleep: unobtrusive sleep quality monitoring using smartphones. In: Proceedings of the 11th ACM conference on embedded networked sensor systems (SenSys '13). New York: ACM; 2013. p. 4:1–4:14. doi:10.1145/2517351.2517359.
17. Haykin S, Chen Z. The cocktail party problem. Neural Comput. 2005;17(9):1875–902.
18. Hoch CC, Reynolds CF, Kupfer DJ, Berman SR, Houck PR, Stack JA. Empirical note: self-report versus recorded sleep in healthy seniors. Psychophysiology. 1987;24(3):293–9.
19. Kato T, Masuda Y, Yoshida A, Morimoto T. Masseter emg activity during sleep and sleep bruxism. Arch Ital Biol. 2011;149(4):478–91.
20. Kleinberg J. Bursty and hierarchical structure in streams. Data Min Knowl Discov. 2003;7(4):373–97.
21. Kohonen T. The self-organizing map. Neurocomputing. 1998;21(1):1–6.
22. Kohonen T. Essentials of the self-organizing map. Neural Netw. 2013;37:52–65.
23. Kohonen T, Kaski S, Lagus K, Salojärvi J, Honkela J, Paatero V, Saarela A. Self organization of a massive document collection. IEEE Trans Neural Netw. 2000;11(3):574–85.
24. Lavigne G, Rompre P, Montplaisir J. Sleep bruxism: validity of clinical research diagnostic criteria in a controlled polysomnographic study. J Dent Res. 1996;75(1):546–52.
25. Li J, Najmi A, Gray RM. Image classification by a two-dimensional hidden Markov model. IEEE Trans Signal Process. 2000;48(2):517–33. doi:10.1109/78.823977.
26. Mantua J, Gravel N, Spencer R. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors. 2016;16(5):646.
27. Metsis V, Kosmopoulos D, Athitsos V, Makedon F. Non-invasive analysis of sleep patterns via multimodal sensor input. Pers Ubiquitous Comput. 2014;18(1):19–26.
28. Mollayeva T, Thurairajah P, Burton K, Mollayeva S, Shapiro CM, Colantonio A. The pittsburgh sleep quality index as a screening tool for sleep dysfunction in clinical and non-clinical samples: a systematic review and meta-analysis. Sleep Med Rev. 2016;25:52–73.
29. Noh T, Serizawa Y, Kimura T, Yamazaki K, Hayasaka Y, Itoh T, Izumi S, Sasaki T. The assessment of sleep stage utilizing body pressure fluctuation measured by water mat sensors. J Adv Sci. 2009;21(1&2):27–30.
30. Parmeggiani PL. Systemic homeostasis and poikilostasis in sleep: Is REM sleep a physiological paradox? Singapore: World Scientific; 2011.
31. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE. 1989;77(2):257–86.
32. Riedel B, Lichstein K. Objective sleep measures and subjective sleep satisfaction: how do older adults with insomnia define a good night's sleep? Psychol Aging. 1998;13(1):159–63. doi:10.1037//0882-7974.13.1.159.
33. Rokach L, Maimon O. Clustering methods. In: Data mining and knowledge discovery handbook. Heidelberg: Springer; 2005. p. 321–352.
34. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53–65.
35. Russo K, Goparaju B, Bianchi MT. Consumer sleep monitors: is there a baby in the bathwater? Nat Sci Sleep. 2015;7:147–57.
36. Silber MH, Ancoli-Israel S, Bonnet MH, Chokroverty S, Grigg-Damberger MM, Hirshkowitz M, Kapen S, Keenan SA, Kryger MH, Penzel T, et al. The visual scoring of sleep in adults. J Clin Sleep Med. 2007;3(2):121–31.
37. Simula O, Kangas J. Process monitoring and visualization using self-organizing maps. Neural Netw Chem Eng. 1995;6:371–84.

38. Sonnhammer EL, Von Heijne G, Krogh A, et al. A hidden Markov model for predicting transmembrane helices in protein sequences. ISMB. 1998;6:175–82.
39. Tanaka H, Shirakawa S. Sleep health, lifestyle and mental health in the japanese elderly: ensuring sleep to promote a healthy brain and mind. J Psychosom Res. 2004;56(5):465–77.
40. Vesanto J, Alhoniemi E. Clustering of the self-organizing map. IEEE Trans Neural Netw. 2000;11(3):586–600.
41. Wu H, Kato T, Yamada T, Numao M, Fukui K. Personal sleep pattern visualization via clustering on sound data. In: Proceedings of AAAI workshops at the 31st AAAI conference on artificial intelligence; 2017. p. 592–9.