**Personal Sleep Pattern Visualization via Clustering on Sound Data**

**Hongle Wu**  
The Institute of Scientific and Industrial Research  
Osaka University  
Mihogaoka 8-1, Ibaraki, Osaka 567-0047, JAPAN

**Takafumi Kato, Tomomi Yamada**  
Graduate School of Dentistry  
Osaka University  
Yamada-Oka 1-8, Suita, Osaka 565-0871, JAPAN

**Masayuki Numao, Ken-ichi Fukui**  
The Institute of Scientific and Industrial Research  
Osaka University  
Mihogaoka 8-1, Ibaraki, Osaka 567-0047, JAPAN

---

**Abstract**

The quality of a good sleep is important for a healthy life. Recently, several sleep analysis products have emerged on the market; however, many of them require additional hardware or there is a lack of scientific evidence regarding their clinical efficacy. We proposed a novel method via clustering of sound events for discovering the sleep pattern. This method extended conventional self-organizing map algorithm by kernelized and sequence-based technologies, obtained a fine-grained map that depicts the distribution and changes of sleep-related events. We introduced widely applied features in sound processing and popular kernel functions to our method, evaluated their performance, and made a comparison. Our method requires few additional hardware, and by visualizing the transition of cluster dynamics, the correlation between sleep-related sound events and sleep stages was revealed.

---

**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 (Chen et al. 2013).

In medical treatment, the primary tool for sleep study is polysomnography (PSG) (Chokroverty 2013). 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 (Kato et al. 2011) (Lavigne, Rompre, and Montplaisir 1996). Due to its professional property and financial cost, PSG usage is limited to only clinics. Hence, instead of using PSG to study the sleep quality, we are looking for an approach to make the assessment more economical and more practical, also in the same time, keeping the accuracy within an acceptable range.

Currently, there are many products on the market that aim to make sleep assessment portable at a reduced cost. ZEO\(^1\) 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\(^2\) and Fitbit\(^3\). One of the problems 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 (Choe et al. 2010). Even if users accept to wear the device, it is not easy to properly place the sensors in the correct position. Also, according to (Mantua, Gravel, and Spencer 2016), medical experts do not suggest to use the results from these consumer equipment for medical research, which means they are not reliable enough.

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. However, according to (Behar et al. 2013), very few of the apps are based on published scientific evidence.

To solve the problems mentioned above simultaneously, and considering that many types of sleep disorder are respectively related to a distinctive type of sound, such as snoring, tooth grinding, limb movement and sleep talking, this paper proposes a method for sleep analysis based on clustering of sound data. The main features of our method are as follows:

**Fine-grained sleep process visualization:** We propose a novel algorithm to cluster the sleep-related events on spatio-temporal dimensions. The transition of sleep state is visualized on the cluster map, which provides a clear and easy way to understand the analysis report.

**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.

---

\(^1\)https://en.wikipedia.org/wiki/Zeo,_Inc.  
\(^2\)http://www.beddit.com/  
\(^3\)https://www.fitbit.com/
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 validated: The consistency of this method with the medical evidence from PSG proves its reliability.

We extracted sound clips of events from the recorded sound data, applied Fast Fourier Transform (FFT) to get the frequency spectrum as input vectors, and then applied self-organizing map (SOM) (Kohonen 1995) algorithm to the data to obtain cluster maps. In our previous work (Wu et al. 2016), we calculated the Euclidean distance between frequency spectrum as the only similarity measure between sound events in standard SOM, in this paper, in order to make a comparison, we applied Mel Frequency Cepstral Coefficient (MFCC) (Davis and Mermelstein 1980) which is a feature widely used in automatic speech recognition as another metric. Besides the standard SOM, kernel SOM (Fukui et al. 2011) was also used. Since the Euclidean distance applied in the standard SOM treats each discrete point as an independent variable, in (Wu et al. 2016), Kulbak-Leibler (KL) kernel was introduced through kernel SOM as a similarity measure in order to capture the distribution structure of a frequency spectrum. In this paper, to make a comparison, we tried radial basis function (RBF) kernel and polynomial kernel besides KL kernel. According to experiment results, the KL kernel SOM (KL-KSOM) obtained the best effect.

In order to visualize the transition of cluster dynamics, we introduced sequence-based self-organizing map (ShSOM) (Fukui et al. 2008). ShSOM is an extended SOM that introduces the sequencing weight function (SWF) in SOM. By converting the spatio-temporal neighborhood into the topological neighborhood via the neighborhood function, it is able to visualize the transition of cluster dynamics. Based on property of KL-KSOM, we introduced KL kernel into ShSOM, and proposed a novel algorithm named sequence-based kernel SOM (Sb-KSOM). Sb-KSOM combined the advantages of kernel SOM and ShSOM, produced a cluster map reflecting the distribution and changing of sleep-related events during the whole sleep.

For evaluation of clustering performance, we labeled the extracted events based on the synchronous PSG data scored by a medical specialist, and calculated the weighted pairwise F-measure (wPF) (Fukui and Numao 2012) as the validity measure of each cluster map.

We performed a comparative interpretation between the obtained cluster maps generated by Sb-KSOM and sleep stage sequences scored by medical specialists based on PSG data. The interpretation revealed that cluster distribution changes synchronously with the transferring of sleep stages. Hence, similar to sleep stage sequences, it is feasible to discover the sleep patterns through the cluster maps generated by Sb-KSOM.

Methodology

Overview

In this section, we introduce the key methodologies applied in this study. Our method process as shown in the Fig. 1, includes following steps:

Sound recording: Recording sound data by recording device.

Data preprocessing: Converting sound format data to text format data through sound processing software.

Events extraction: Extracting sound clips of events from the recorded sound data, burst extraction algorithm is applied.

Extracted data cleansing: Setting a time threshold to filter very short sound clips that unable to properly manually annotate. In this study, the threshold was set to 0.3 second.

Spectrum data obtaining: Applying FFT to obtain the frequency power spectrum of sound clip.

Clustering: Applying standard, kernelized or Sequence-based SOM clustering algorithms on the processed spectrum data to get cluster maps.

Evaluation: Comparing cluster maps visualization; making quantitative comparison between cluster results between from different algorithms through wPF; making a comparative analysis between cluster maps generated sequence-based SOM and sleep stage sequences.

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 (Fukui et al. 2011) to differentiate the steady noise from other types of sound events including sleep disorder symptoms, such as snoring, tooth 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 (Kleinberg 2003).

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.

We assumed the background noise during the whole night to be steady and to be generated according to a constant Gaussian distribution, and other sound events, such as snoring, to be generated by Gaussian distribution with different parameters. The burst extraction method estimates the maximum-likelihood state sequence for the sound event, and is able to extract a variable-length sound event based on the amount of activity of the signals.

MFCC

In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency.

MFCCs are coefficients that collectively make up an MFC (Davis and Mermelstein 1980). They are derived from a type of cepstral representation of the audio clip (a nonlinear “spectrum-of-a-spectrum”). The difference between the cepstrum and the mel-frequency cepstrum is that in the MFC,
the frequency bands are equally spaced on the mel scale, which approximates the human auditory system’s response more closely than the linearly-spaced frequency bands used in the normal cepstrum. This frequency warping can allow for better representation of sound, for example, in audio compression. In our experiment, 12 MFCC coefficients were extracted for each sound clip as a MFCC vector, and the Euclidean distance between the MFCC vectors were applied as the similarity between sound events.

SOM

The SOM 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 (Kohonen et al. 2000), for the monitoring of machinery (Simula and Kangas 1995), and for application to medical care or economics. In this study, we generated cluster maps by clustering algorithms based on SOM, including standard SOM, kernel KSOM, and Sb-KSOM that is proposed in this work. The generation of such a map has the following advantages:

**Comprehensive evaluation:** Similar sound events are assumed to have similar frequency characteristics. Therefore, a cluster of sound events corresponds to a sleep-related event type, for example, snoring. Moreover, by introducing time dimension into the clustering, the distribution of sleep disorder events transition with the sleep time elapsing can also be displayed in the cluster map.

**Exploratory analysis:** The user can intuitively understand the entire picture of several sleep disorder events and explore particular events or high-frequency events.

Let $v$-dimensional $N$ inputs be $x_n = (x_{n,1}, \cdots, x_{n,v}), (n = 1, \cdots, N)$, the position of $M$ neurons in the visualization layer be $r_j = (\xi_j, \cdots, \eta_j), (j = 1, \cdots, M)$, and the reference vector corresponding to the $j^{th}$ neuron be $m_j$ ($v$-dimension).

The following describes the algorithm of the batch type SOM:

**Step 1:** Initialize the reference vectors $\{m_1, \cdots, m_M\}$ randomly and set the iteration step as $t = 1$.

**Step 2:** Search winner neurons $\{c(x_1), \cdots, c(x_N)\}$ for all inputs by the nearest neuron:

$$c(x_n) = \arg\min_j \|x_n - m_j\|, \quad (1)$$

**Step 3:** Exit if the best matching units $\{c(x_1), \cdots, c(x_N)\}$ were not changed or the iteration reached $t = t_{max}$.

**Step 4:** Update each reference vector by the following equation:

$$m_j^{new} = m_j + h_{c(x_n),j}[x_n - m_j], \quad (2)$$

where $h_{c(x_n),j}$ is a neighborhood function that defines the effect of neighborhood of the winner. Typically, Gaussian function is used:

$$h_{c(x_n),j} = \alpha \exp \left(-\frac{\|r_j - r_{c(x_n)}\|^2}{2\sigma^2}\right), \quad (3)$$

**Step 5:** Decrease the the learning parameters $\alpha$ and $\sigma$ , and increase the iteration counter $t \rightarrow t + 1$. Then, return to Step 2.

**Kernel SOM**

We used the frequency spectrum as input vector. 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 (Fukui et al. 2011), 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 kernel, RBF kernel and polynomial kernel were introduced to SOM through kernel SOM(Andras 2002) (Boulet et al. 2008) to cluster the sleep-related sound events.

The RBF kernel function is defined as

$$K_{RBF}(x_i, x_j) = \exp \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \quad (4)$$
where $x_i$ and $x_j$ are vectors in the input space, and $\sigma$ is a free parameter. For degree-$d$ polynomials, the polynomial kernel function is defined as:

$$K_{PL}(x_i, x_j) = (x_i^T x_j + 1)^d,$$

(5)

The KL kernel function is defined as:

$$K_{KL}(x_i, x_j) = \exp \left( -\beta JS(x_i, x_j) \right),$$

(6)

$$JS(x_i, x_j) = KL(x_i, x_j) + KL(x_j, x_i)$$

$$= \sum_{k=1}^{v} \left\{ x_{i,k} \log \frac{x_{i,k}}{x_{j,k}} + x_{j,k} \log \frac{x_{j,k}}{x_{i,k}} \right\},$$

(7)

where $KL(x_i, x_j)$ is the KL divergence, which is the distance between probability distributions, $JS(x_i, x_j)$ 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 $x$ in the updating formula of a reference vector in the standard batch type SOM by a mapped $\phi(x)$, the following updating formula can be obtained:

$$m_i(t + 1) := \gamma \sum_n h_{c(x_n),i} \phi(x_n),$$

(8)

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

$$d_{i,n}(t + 1) := ||\phi(x_n) - m_i(t + 1)||^2$$

$$= K(x_n, x_n) - 2\gamma \sum_j h_{c(x_j),i} K(x_n, x_j)$$

$$+ \gamma^2 \sum_k h_{c(x_k),i} h_{c(x_l),i} K(x_k, x_l).$$

(9)

The following describes the algorithm of the batch type kernel 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 $\{c(x_1), \cdots, c(x_N)\}$ for all inputs by the nearest neuron:

$$c(x_N) = \arg\min_{i=1,\cdots,M} d_{i,n},$$

(10)

**Step 3:** (Same as SOM) Exit if the best matching units $\{c(x_1), \cdots, c(x_N)\}$ 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. (9).

**Step 5:** Decrease the neighborhood radius $\sigma$ and increase the iteration counter $t \to t + 1$. Then, return to Step 2.

**Sequence-based SOM**

In order to clearly and easily understand the analysis report of a user’s sleep, a fine-grained map that depicts the distribution and changing of sleep-related events is necessary. Different from the normal SOM that deals with static data, Sb-SOM introduces SWF into SOM and can visualize the transition of cluster dynamics since the spatio-temporal neighborhood is converted into the topological neighborhood by the neighborhood function.

In SbSOM, the position of $M$ neurons in the visualization layer be $r_j = (\xi_j, \eta_j), (j = 1, \cdots, M)$, where $\xi$-direction indicates the temporal dimension. The $n$th input data are located at the ratio of $n/N$ within the input data sequence, and the $j$th neuron is located at the ratio of $\xi_j/\xi_M$ on the $\xi$-direction of cluster map. Let the absolute value of those differences be $\epsilon = |\xi_j/\xi_M - n/N|$. The SWF $\psi(n, \xi_j)$ is defined so as to be able to balance the spatio/temporal resolution; in case where reversal of data order is not allowed, the SWF is given as:

$$\psi(n, \xi_j) = \begin{cases} 1 & \text{if } \epsilon < \frac{1}{2K} \\ \infty & \text{otherwise} \end{cases},$$

(11)

where $K$ is the number of neurons on $\xi$-direction. The winner neuron of the input data $x_n$ is determined by spatial distance combined with SWF as follows:

$$c(x_n) = \arg\min_j \psi(n, \xi_j)||x_n - m_j||.$$  

(12)

**Proposed method: Sequenced-based kernel SOM**

In this study, the comparison of the clustering results between standard and kernel SOM demonstrated that KL divergence as kernel function exhibits better performance. Based on this premises, a novel algorithm, Sb-KSOM, is proposed, which is an extension of ShSOM. The proposed Sb-KSOM kernelized the ShSOM by replacing the Euclidean distance with KL divergence to enable it to handle the frequency spectrum data.

In the proposed Sb-KSOM, we replaced the normal Euclidean distance calculation in Eq. (12) with the KL kernel function, the following describes the algorithm of the batch type kernel SOM:

**Step 1:** (Same as kernel SOM) 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 $\{c(x_1), \cdots, c(x_N)\}$ for all inputs by spatio-temporal distance utilizing SWF $\psi(n, \xi_j)$, the dissimilarity between $j$th reference vector and $n$th data point $d_{j,n}$ is calculated by Eq. (9):

$$c(x_n) = \arg\min_j \psi(n, \xi_j) d_{j,n}.$$  

(13)

**Step 3:** (Same as SOM) Exit if the best matching units $\{c(x_1), \cdots, c(x_N)\}$ 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. (9) with KL kernel.

**Step 5:** (Same as kernel SOM) Decrease the neighborhood radius $\sigma$ and increase the iteration counter $t \to t + 1$. Then, return to Step 2.
Table 1: Subject and sound data information

<table>
<thead>
<tr>
<th>Subject id</th>
<th>Age</th>
<th>Gender</th>
<th>Recording date</th>
<th>Duration</th>
<th>Primary disorder symptoms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21</td>
<td>F</td>
<td>2014/05/13</td>
<td>08:05:22</td>
<td>tooth grinding</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>M</td>
<td>2014/05/27</td>
<td>08:16:15</td>
<td>tooth grinding, snoring</td>
</tr>
<tr>
<td>3</td>
<td>22</td>
<td>M</td>
<td>2014/06/03</td>
<td>08:01:09</td>
<td>tooth grinding, snoring</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>M</td>
<td>2014/07/29</td>
<td>08:23:01</td>
<td>tooth grinding, snoring</td>
</tr>
<tr>
<td>5</td>
<td>24</td>
<td>M</td>
<td>2015/01/20</td>
<td>08:17:34</td>
<td>tooth grinding, snoring</td>
</tr>
<tr>
<td>6</td>
<td>23</td>
<td>F</td>
<td>2015/03/03</td>
<td>08:30:30</td>
<td>tooth grinding</td>
</tr>
<tr>
<td>7</td>
<td>20</td>
<td>F</td>
<td>2015/06/02</td>
<td>07:18:30</td>
<td>tooth grinding</td>
</tr>
</tbody>
</table>

Table 2: Comparison of wPF between standard SOM and kernel SOM clustering results

<table>
<thead>
<tr>
<th>Subject id</th>
<th>SOMspectrum Mean</th>
<th>SOMspectrum SD</th>
<th>SOMMFCC Mean</th>
<th>SOMMFCC SD</th>
<th>KSOMKL Mean</th>
<th>KSOMKL SD</th>
<th>KSOMRBF Mean</th>
<th>KSOMRBF SD</th>
<th>KSOMPL Mean</th>
<th>KSOMPL SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.537</td>
<td>0.033</td>
<td>0.509</td>
<td>0.041</td>
<td>0.604</td>
<td>0.037</td>
<td>0.593</td>
<td>0.051</td>
<td>0.504</td>
<td>0.047</td>
</tr>
<tr>
<td>2</td>
<td>0.521</td>
<td>0.041</td>
<td>0.497</td>
<td>0.043</td>
<td>0.573</td>
<td>0.038</td>
<td>0.577</td>
<td>0.038</td>
<td>0.482</td>
<td>0.025</td>
</tr>
<tr>
<td>3</td>
<td>0.506</td>
<td>0.031</td>
<td>0.493</td>
<td>0.035</td>
<td>0.551</td>
<td>0.031</td>
<td>0.532</td>
<td>0.033</td>
<td>0.535</td>
<td>0.042</td>
</tr>
<tr>
<td>4</td>
<td>0.559</td>
<td>0.040</td>
<td>0.482</td>
<td>0.032</td>
<td>0.592</td>
<td>0.037</td>
<td>0.561</td>
<td>0.035</td>
<td>0.567</td>
<td>0.026</td>
</tr>
<tr>
<td>5</td>
<td>0.602</td>
<td>0.039</td>
<td>0.594</td>
<td>0.039</td>
<td>0.629</td>
<td>0.039</td>
<td>0.624</td>
<td>0.041</td>
<td>0.608</td>
<td>0.054</td>
</tr>
<tr>
<td>6</td>
<td>0.543</td>
<td>0.033</td>
<td>0.549</td>
<td>0.045</td>
<td>0.600</td>
<td>0.035</td>
<td>0.562</td>
<td>0.038</td>
<td>0.557</td>
<td>0.037</td>
</tr>
<tr>
<td>7</td>
<td>0.483</td>
<td>0.042</td>
<td>0.501</td>
<td>0.035</td>
<td>0.523</td>
<td>0.047</td>
<td>0.531</td>
<td>0.051</td>
<td>0.537</td>
<td>0.036</td>
</tr>
<tr>
<td>Mean</td>
<td>0.535</td>
<td>0.037</td>
<td>0.518</td>
<td>0.039</td>
<td>0.581</td>
<td>0.037</td>
<td>0.568</td>
<td>0.041</td>
<td>0.541</td>
<td>0.038</td>
</tr>
</tbody>
</table>

Experiment

We first applied the standard SOM and three kinds of kernel SOM to the extracted sound data, and compared the wPF. Then we used Sb-KSOM on the data to obtain the spatio-temporal dimensional cluster map, and discussed the relation between the transition of sleep stages and cluster dynamics of sound events.

Experimental setting

The data used in this study were prepared by the Graduate School of Dentistry in Osaka University. The study protocol was approved by the clinical research ethics Committee of the Osaka University Graduate School of Dentistry. Written informed consent was obtained from all subjects. All subjects were asked to sleep in a specific room (Fig. 2) from 22:30 to 8:00. The recording device included LA1250 (Ono Sokki)\(^4\) and R-4 Pro (Roland)\(^5\). 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.

All of the experimental subjects are university students from Osaka University, and hence, their age was mostly around 20-24. The male to female ratio was balanced. Table 1 shows information of the subjects and the recorded sound data that were used in the experiment.

Event extraction

We selected ten nights of sound data. Based on the burst extraction method, we obtained a total of 6775 sound events, which included sleep disorder and other sound events such as outdoor traffic noise. FFT was applied to the extracted sound data to obtain the frequency power spectrum. From 24 Hz to 20 kHz, at intervals of 4 Hz, 4995 discretized points as an input for SOM were obtained for every sound data.

Quantitative comparison between standard and kernelized clustering

In the first part of this experiment, to determine which algorithm is the most suitable for this study, we used the sound data from each subject as a respective dataset and compared the wPF values for each subject between standard and kernelized algorithms, including standard SOM based on frequency spectrum or MFCC similarity, kernel SOM with KL, RBF or polynomial kernel. In order to avoid initial value dependency, the experiments were executed 50 times and the

---

\(^4\)https://www.onosokki.co.jp/English/hp_e/products/keisoku/s_v/la1200.html

\(^5\)http://proav.roland.com/products/r-4_pro/
average values were computed. The hyper parameter of the kernel functions were tuned by a linear search. The mean wPF values and standard deviation are shown in Table 2. The average of wPF shows that MFCC feature does not perform well on this kind of sound data, and the KL kernel SOM has the best performance, which improved by about 10% from standard SOM.

**weighted Pairwise F-measure (wPF)** The original pairwise F-measure evaluates the correlation between cluster assignment and class label. However, especially in SOM visualization, neighborhood relation is also important. Then, we employed the weighted version of the pairwise F-measure to evaluate SOM visualization comprehensively. By using events scored through PSG as ground truth labels, we applied wPF (Fukui and Numao 2012) to evaluate the clustering results. wPF is an extension of pairwise-based cluster validity measures (Xu and Wunsch 2008), which introduce the likelihood function indicating a degree that a data pair belongs to the same cluster instead of the actual number of data pairs.

**Sleep pattern analysis**

In this experiment, we made a comparative analysis between cluster maps generated by Sb-KSOM and sleep stage sequences scored from PSG data to reveal the relation between them. We analyzed all the subject respectively. One of the clustering results is shown in this section. Subject 2 was chosen since the tooth grinding or snoring activities are frequent, and generated more related sound events than the others. Fig. 3 upper part shows the result when Sb-KSOM was applied to the sound data from Subject 2; the number of neurons was set to 50×10 with a two-dimensional grid. Subjects’ sleep stages were scored by a medical specialist based on PSG data from the same night, with a time window size of 30s. The sleep stage sequence of Subject 2 is shown in lower part of Fig. 3, where REM stage is shown as “R”, awake stage is shown as “W”. We defined the period that contain continuous N3 stages with intervals of other stages that less than 3 min as a deep sleep period, and periods except deep sleep periods, awakening stages and REM stages as light sleep periods. Since the REM stage is a unique phase in the sleep process, we will discuss it separately.

The sleep periods of Subject 2 were interpreted as follows:

- **Deep sleep periods**: (0:13:30 - 1:01:30), (1:39:30 - 1:52:30), (2:00:30 - 02:11:00), (2:20:30 - 02:51:00), (4:04:30 - 4:20:30), (6:00:30 - 6:18:30): There were many snoring events during these periods, few body movements, and tooth grindings. We found out that a cluster center of snoring event is usually associated with a deep sleep period.

- **REM stages**: (2:53:00 - 3:05:00), (4:42:00 - 5:39:30), (6:56:00 - 7:29:30): Compared with other stages, REM stages have a stronger association with clusters of body movement and a weaker association with those of snoring or tooth grinding.

- **Light sleep periods**: In each light sleep period, there were some clusters of tooth grinding and body movement event but only a few snoring events.

In this experiment, we found that the distribution of sound event clusters changed simultaneously with the sleep stage change, for not only Subject 2, but also the other subjects. Even though our analysis includes other subjects who have different primary sleep disorders and varying pattern of sleep stages, the finding led to similar conclusions. For example, on Subject 4, the deep sleep periods were also obviously associated with the clusters of snoring, the number of body movements was notably more in the light periods and REM stages than in the deep periods, and no snoring clusters were found in REM stages.

Similar discussions are found in other studies. According to (Fairbanks, Mickelson, and Woodson 2003), conventional snoring is most likely to occur during the deep sleep stage, as well as during the light sleep stage, but unlikely during the REM stage. The REM stage is always associated with dreaming (Hobson, Pace-Schott, and Stickgold 2000), which triggers several body movements.

From this experiment, we found that the transition of cluster dynamics and the changing of sleep stage are related. Since the sleep stage sequence is an important tool in the study of sleep pattern, its relation provides the possibility of discovering sleep patterns based on the cluster map of sleep-related sound data from Sb-KSOM.

**Discussion**

However, currently 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. Moreover, there are lots of useless noise events in the extracted sound dataset, these noise events took up many grids in the cluster maps, impacted the accuracy of our experiments. In the future, we will improve the current noise reduction algorithm, to remove more useless noise. Furthermore, although our sound-based method cannot evaluate quiet subjects, those quiet subjects might slept better than the others, and we will confirm this assumption in the future work.

**Related work**

In the academic field of sleep analysis, various studies using other methods besides PSG trying to reduce the cost and simplify the operation, such as infrared thermography (Fukumura, Okada, and Makikawa 2012), water filled mat (Noh et al. 2009) and Kinect (Metsis et al. 2014), have been proposed. Similar to PSG, 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 reducing the cost of data collection and making large-scale data collection possible.

There are several personal health products on the market that provide sleep analysis services, such as aforementioned ZEO, Beddit, Fitbit, and Slimsee (Toshiba) (Yamada et al. 2014). Compared with these products, the sound data-based
study that we have proposed incurs very little extra hardware cost and does not require users to have physical contact with the additional hardware.

With the popularity of smartphones, analyzing the sleep quality via smartphone applications has gradually grown. 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 (Gu et al. 2014), and Hao et al. developed an application called iSleep (Hao, Xing, and Zhou 2013). 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 Android6 and Sleep Cycle alarm clock7; however, no academic proof or accuracy evaluation for their outputs exists. Since our research is consistent with PSG, the application that integrates our method will be more reliable than these applications.

Conclusion

This paper proposes a novel approach to discover the sleep pattern through analyzing the sleep-related sound events based on extended self-organizing map algorithms. This method combined the advantages of kernelization and sequence-based technologies, and obtained a fine-grained map that depicts the distribution and changes of sleep-related events.

According to the experiment results, we can draw the following conclusion: the transition of cluster dynamics obtained by Sb-KSOM and the changing of sleep stage are related, in particular, cluster centers of snoring event are related to deep sleep stages, REM stages have a stronger association with clusters of body movement and a weaker association with those of snoring or tooth grinding, and clusters of tooth grinding and body movement event are more related to light sleep stage than snoring events; on the other side, conditional probabilities on different subjects revealed that sleep patterns of individuals are not the same, therefore, in the future, respectively studying and training a personalized model for each individual to evaluate the sleep quality is necessary.

Since the final objective of our research is to make the assessment of personal sleep quality more economical, more practical and more reliable, the correlation between sound data and sleep stages provides a new train of thought for studying the sleep pattern. In the future work, we will proceed to develop a predictive model for personal sleep quality scoring, the sleep-related sound data will be a main part of input data, and the relationship between sleep stages and sound events will play a key role in the algorithm development.

Acknowledgments

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.

References


Mantua, J.; Gravel, N.; and Spencer, R. 2016. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors 16(5):646.


