Learning Sleep Pattern based on Audio Data

Submitted to

Graduate School of Information Science and Technology
Osaka University

January 2018

Hongle WU

Research output

Journal publications

1. Hongle Wu, Takafumi Kato, Tomomi Yamada, Masayuki Numao, Ken-ichi Fukui. Personal sleep pattern visualization using sequence-based kernel self-organizing map on sound data, Artificial Intelligence in Medicine, Vol. 80, pp. 1-10, ISSN 0933-3657, 2017.

2. Hongle Wu, Takafumi Kato, Masayuki Numao, Ken-ichi Fukui. Statistical Sleep Pattern Modelling for Sleep Quality Assessment based on Sound Events, Health Information Science and Systems, Springer. Vol. 5(1), 2017.

International conference papers

1. Hongle Wu, Takafumi Kato, Tomomi Yamada, Masayuki Numao, Ken-ichi Fukui. Personal Sleep Pattern Visualization via Clustering on Sound Data. In Proc. AAAI 2017 Joint Workshop on Health Intelligence, pp. 592-599, Feb. 2017.

2. Hongle Wu, Takafumi Kato, Tomomi Yamada, Masayuki Numao, Ken-ichi Fukui. Sleep Pattern Discovery via Visualizing Cluster Dynamics of Sound Data. In Proc. The 29th International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems (IEA/AIE 2016) (LNAI 9799), pp. 460-471, Aug. 2016.

Other presentations

1. Hongle Wu, Takafumi Kato, Masayuki Numao, Ken-ichi Fukui. Sleep Pattern Modelling for Quality Prediction based on Sound Data, In Proc. Technical Committee on Artificial Intelligence and Knowledge-Based Processing (AI), IEICE, Nov. 2017.

2. Hongle Wu, Takafumi Kato, Tomomi Yamada, Masayuki Numao, Ken-ichi Fukui. Sleep Pattern Visualization via Clustering on Sound Data, In Proc. The 31st Annual Conference of Artificial Intelligence Society of Japan, May 2017.

3. Hongle Wu, Takafumi Kato, Tomomi Yamada, Masayuki Numao, Ken-ichi Fukui. Sleep Pattern Discovery and Visualization based on Clustering of Sound Events,

In Proc. The 30st Annual Conference of Artificial Intelligence Society of Japan, June 2016.

4. Hongle Wu, Takafumi Kato, Tomomi Yamada, Masayuki Numao, Ken-ichi Fukui. Sleep Pattern Characterization via Cluster Analysis of Audio Data, In Proc. The Society of Artificial Intelligence 106th Knowledge Base System (SIG-KBS), pp. 42-48, Nov. 2015.

5. Hongle Wu, Ken-ichi Fukui, Takafumi Kato, Masayuki Numao. Individual Sleep Pattern Characterization via Cluster Analysis of Audio Data, In Proc. Workshop on Computation: Theory and Practice (WCTP-2015), Sep. 2015. (abstract only)

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 research, we propose a method to discover sleep patterns via clustering of audio events recorded during sleep. The proposed method extends the conventional self-organizing map algorithm by kernelization and sequence-based technologies to obtain a fine-grained map that visualizes the distribution and changes of sleep-related events. We introduced features widely applied in audio processing and popular kernel functions to the proposed method to evaluate and compare performance. The proposed method provides a new aspect of sleep monitoring because the results demonstrate that audio events can be directly correlated to an individual's sleep patterns. In addition, by visualizing the transition of cluster dynamics, sleep-related audio events were found to relate to the various stages of sleep. Therefore, these results empirically warrant future study into the assessment of personal sleep quality using audio data. Then, based on this discovery, we assess the sleep quality through audio events. 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.

Acknowledgement

This dissertation would not have been possible without the help and support of my professors, family and friends. I would like to thank first and foremost Professor Kenichi Fukui who has always been very supportive of me. His guidance, encouragement and support has allowed me to learn so many new things and meet so many people in the research community allowing me to grow as a researcher and progress in my academic career. I also thank Professor Masayuki Numao, Professor Takafumi Kato and Professor Tomomi Yamada who gave me a lot of guidance and insight in my research.

I would also like to thank the members of my review committee, Professor Jun Tanida and Professor Hideyuki Suzuki who have shared their valuable thoughts and comments to further refine and improve my research.

I would like to thank the students of Numao laboratory who have also shared their thoughts on my research, their time in helping me develop and test my software and the fun times that we have had together. I thank Ms. Mayo Kamimura for helping me with my data collection. I thank the laboratory secretaries Ms. Megumi Tanabe, Ms. Akiko Yamamoto, Ms. Mitsuyo Ohtsuka, Ms. Azusa Hirabayashi and Ms. Mika Kusakabe for helping me out with all my university concerns and conference arrangements.

I would not have had the honor of being a student of Osaka University without the aid of the Management Expenses Grants for National Universities Corporations through the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan. I thank the support of my research from 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.

Most importantly, I thank to my family who has always been there for me and pushing me forward. I thank my wife Dandan who has always supported me, even during her pregnancy, she still tried her best to share the housework. I would like to thank my son Genki, his arrival gave me more motivation to achieve more than I could.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background of this research

Sleep is an important physiological state. People spend approximately one-third of their life sleeping, and the quality of sleep is very important to a person's health. Therefore, sleep monitoring technology has become essential in modern personal sleep management [12].

Polysomnography (PSG) [14] is the primary tool used in clinical treatment and sleep research [32] [38]. PSG employs electroencephalography, electrooculography, electromyography, and electrocardiography. However, PSG is complex and expensive. Therefore, a more practical and economical approach that provides acceptable accuracy is desired. To the best of our knowledge, sleep disorders are commonly associated with distinctive sounds, such as snoring, teeth grinding, body movements, and sleep talking. In addition, environmental sounds inside and outside the bedroom also directly impact sleep. Thus, we proposes a method to model personal sleep patterns based on clustering sleep related audio data. In our research, a **Personal Sleep Pattern** is defined as the unique characteristic of the clustering of sleep-related audio events; in other words, the direct correlation between audio events and sleep. Sleep-related audio events include sounds commonly associated with sleep disorders and environmental sounds. We assumed that changes in sleep environments will result in different sleep patterns, even for the same subject.

1.2 Research topics

The objective of our research is to develop a more practical and economical approach that provides acceptable accuracy of sleep study, it includes two major topics.

Firstly, we propose a method to discover sleep patterns via clustering of audio events recorded during sleep. The proposed method extends the conventional self-organizing map (SOM) algorithm by kernelization and sequence-based technologies to obtain a fine-grained map that visualizes the distribution and changes of sleep-related events. We introduced features widely applied in audio processing and popular kernel functions to the proposed method to evaluate and compare performance. By visualizing the transition of cluster dynamics, sleep-related audio events were found to relate to the various stages of sleep. In addition, we calculated the conditional probabilities of an audio event given the current sleep stage, quantified the correlation between sleep-related audio events and sleep stages. The conditional probabilities demonstrate that snore events have the strongest relationship with deep sleep and body movement is more related to rapid eye movement (REM) and light sleep than deep sleep. Teeth grinding most frequently occurs during Non-REM sleep. The proposed method provides a new aspect of sleep monitoring because the results demonstrate that audio events can be directly correlated to an individual's sleep patterns, and empirically warrant future study into the assessment of personal sleep quality using audio data.

Our second research topic is assessing the sleep quality through audio events. We used subjective sleep quality as training label, combined several machine learning approaches including kernelized self-organizing map, hierarchical clustering (HC) and hidden Markov model (HMM), obtained the models to indicate the sleep pattern of specific quality level. We found there is no significant difference on sleep stage sequence HMMs between different sleep quality, on the contrary, the HMMs of audio events from different sleep quality level have obvious difference. This evidence is interesting that sleep stage sequence is useless for assessing sleep quality. Therefore, the HMMs of audio events from good and poor sleep quality were used to model the sleep quality. The likelihoods

between an input audio 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. 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 audio data.

1.2.1 Sleep pattern visualization

We extracted audio clips of events from the recorded audio data, applied Fast Fourier Transform (FFT) to obtain the frequency spectrum as input vectors, and applied various SOM [36] algorithms to obtain cluster maps. In our previous work [61], we calculated the Euclidean distance between the frequency spectra as the only similarity measure between audio events in standard SOMs. In this study, for comparison, we applied the Mel Frequency Cepstral Coefficient (MFCC) [18] which is widely used in automatic speech recognition as an additional metric. As well as the conventional SOM, a Kullback-Leibler (KL) kernel SOM (KL-KSOM) [22] was also used. Euclidean distance applied to a conventional SOM treats each discrete point as an independent variable; thus, we introduced the KL kernel as a similarity measure to capture the distribution structure of a frequency spectrum. KL divergence has been proposed previously to introduce a distribution structure into a similarity measure of the frequency spectrum of acoustic emission events and obtained good results [22]. For comparison, we also used a radial basis function (RBF) kernel and a polynomial kernel. The experiment results show that the KL kernel outperformed the RBF and polynomial kernels. In addition, we found that the KL-SOM outperformed the standard SOM.

To visualize the transition of cluster dynamics, we introduced a sequence-based SOM (Sb-SOM) [24], which introduces a sequencing weight function (SWF). By converting the spatiotemporal neighborhood into the topological neighborhood using a neighborhood function, the Sb-SOM can visualize the transition of cluster dynamics. Based on the property of kernel SOM, we introduced the KL kernel into the Sb-SOM and propose a sequence-based kernel SOM (Sb-KSOM). The Sb-KSOM algorithm, which combines

the advantages of a kernel SOM and an Sb-SOM, produces a cluster map that reflects the distribution and change of sleep-related events during the sleep period. To evaluate clustering performance, we calculated the weighted pairwise F-measure (wPF) [23] as the validity measure of each cluster map.

Since a personal sleep pattern is directly modeled via sleep-related audio events, the proposed method does not require sleep stage estimation; however, the most accepted clinical research methods involve sleep stage. Therefore, to validate the proposed method, we performed a comparative interpretation between the obtained cluster maps generated by the Sb-KSOM and sleep stage sequences scored by medical specialists based on PSG data. The interpretation revealed that cluster distribution changes synchronously as sleep stages transition. Thus, similar to sleep stage sequences, discovering sleep patterns using cluster maps generated by an Sb-KSOM is feasible. To reveal the correlations between audio events and sleep stages, we calculated the conditional probabilities of audio events for a given sleep stage.

1.2.2 Sleep quality prediction

In the previous topic, we applied various kinds of SOM algorithms to the extracted audio clips of sleep-related events to obtain cluster maps, and proved the reliability and feasibility of kernelized self-organizing map for sleep-related audio 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 research, we applied HC on the cluster map from KL-KSOM, 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 audio 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 audio recordings with 18 in good quality and 18 in poor quality, and we classified audio events that extracted from these recordings by the aforementioned virtual classifier. After we got these categorized audio 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 [45], protein [54]. Also, there are works for classification with HMM [39]. In this study, the HMMs of good and poor sleep quality were used to predict the sleep quality. The likelihoods between an input audio 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.

1.3 Significance of this research

To date, technology has not enabled personalized, in-place sleep quality monitoring and analysis. PSG is the primary clinical tool for sleep monitoring. It can provide a quantitative pro

ling 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.

On the other hand, 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

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

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 [13]. Even if users accept to wear the device, it is not easy to properly place the sensors in the correct position. Also, according to [40], 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 [4], 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 audio, such as snoring, tooth grinding, limb movement and sleep talking, we proposes a method for sleep analysis based on clustering of audio data. The main features of our method are as follows:

Fine-grained sleep process visualization: We propose an algorithm to cluster sleep-related events relative to spatiotemporal dimensions. Sleep state transitions are visualized on a cluster map, which provides a clear and easy way to understand the analysis.

Non-invasive: The audio data can be recorded by any recording device placed near the user's bed during sleep.

No additional cost: Off-the-shelf equipment with a microphone, including smartphones, recording pens, and personal computers, can be used as the recording device.

Scientifically validated: We collaborated with medical experts in this study, a questionnaire was designed by the experts to evaluate the subjective sleep quality of

[2] `http://www.beddit.com/`
[3] `https://www.fitbit.com/`

experiment subjects, which made our training data with sufficient reliability, and the results of the proposed method are consistent with medical evidence obtained using PSG.

1.4 Structure of this thesis

In chapter 2, we introduced the related research of sleep study and the algorithms and techniques that applied in this research.

In chapter 3, we introduced the research of sleep pattern visualization, explained the system work flow, and the experiment setting and analysis the results.

In chapter 4, the framework of sleep quality prediction was introduced, including the clusters detection, event sequence generating and classifier training, also we demonstrate the experiment results and made a comparison between different classifier.

In the last chapter, we made a conclusion and talked about the contributions of this work, also we showed a application scenario imaginary of this method and made some recommendations for future work.

Chapter 2

Literature review

In this chapter, we discuss some relevant works of sleep study, and the methods we applied in our research.

2.1 Related works of sleep study

This section introduced the detail of sleep disorder, PSG method, the sleep stages that scored based on PSG data, and other attempts in sleep study besides PSG.

2.1.1 Sleep disorder

A sleep disorder, or somnipathy, is a medical disorder of the sleep patterns of a person or animal. Some sleep disorders are serious enough to interfere with normal physical, mental, social and emotional functioning. Polysomnography and actigraphy are tests commonly ordered for some sleep disorders. Disruptions in sleep can be caused by a variety of issues, from teeth grinding (bruxism) to night terrors. When a person suffers from difficulty falling asleep and/or staying asleep with no obvious cause, it is referred to as insomnia [63].

Sleep disorders are broadly classified into dyssomnias, parasomnias, circadian rhythm sleep disorders involving the timing of sleep, and other disorders including ones caused by medical or psychological conditions and sleeping sickness. Some common sleep disorders include sleep apnea (stops in breathing during sleep), narcolepsy and hypersomnia (excessive sleepiness at inappropriate times), cataplexy (sudden and transient loss of

muscle tone while awake), and sleeping sickness (disruption of sleep cycle due to infection). Other disorders include sleepwalking, night terrors and bed wetting. Management of sleep disturbances that are secondary to mental, medical, or substance abuse disorders should focus on the underlying conditions.

2.1.2 Polysomnography

Polysomnography (PSG), a type of sleep study, is a multi-parametric test used in the study of sleep and as a diagnostic tool in sleep medicine. The test result is called a polysomnogram. PSG is a comprehensive recording of the biophysiological changes that occur during sleep. It is usually performed at night, when most people sleep, though some labs can accommodate shift workers and people with circadian rhythm sleep disorders and do the test at other times of day. The PSG monitors many body functions including brain (EEG), eye movements (EOG), muscle activity or skeletal muscle activation (EMG) and heart rhythm (ECG) during sleep. After the identification of the sleep disorder sleep apnea in the 1970s, the breathing functions respiratory airflow and respiratory effort indicators were added along with peripheral pulse oximetry [14].

Polysomnography is used to diagnose, or rule out, many types of sleep disorders including narcolepsy, idiopathic hypersomnia, periodic limb movement disorder, REM behavior disorder, parasomnias, and sleep apnea. Although it is not directly useful in diagnosing circadian rhythm sleep disorders, it may be used to rule out other sleep disorders.

2.1.3 Sleep stage

According to American Academy of Sleep Medicine (AASM), the sleep stage scoring based on PSG has long been considered as the "gold standard" of sleep study [7]. 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 [30] [46] [9] [1].

Sleep is divided into two broad types: non-rapid eye movement (non-REM or NREM

sleep) and rapid eye movement (REM sleep). non-REM and REM sleep are so different that physiologists identify them as distinct behavioral states. Non-REM sleep occurs first and after a transitional period is called slow wave sleep or deep sleep. During this phase, body temperature and heart rate fall, and the brain uses less energy [14]. REM sleep (also known as paradoxical sleep), a smaller portion of total sleep time and the main occasion for dreams (or nightmares), is associated with desynchronized and fast brain waves, eye movements, loss of muscle tone, and suspension of homeostasis.

The sleep cycle of alternate NREM and REM sleep takes an average of 90 minutes, occurring 4 to 6 times in a good night's sleep. The AASM divides NREM into three stages: N1, N2, and N3, the last of which is also called delta sleep or slow-wave sleep [50]. The whole period normally proceeds in the order: N1, N2, N3, N2, REM. REM sleep occurs as a person returns to stage 2 or 1 from a deep sleep. There is a greater amount of deep sleep (stage N3) earlier in the night, while the proportion of REM sleep increases in the two cycles just before natural awakening [20] [55].

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.

2.1.4 PSG alternatives for sleep study

Besides PSG, in the academic field of sleep analysis, various studies using other methods trying to simplify the operation, such as single-channel EEG [58], radio signals [65], infrared thermography [25], water filled mat [43] and Kinect [41] 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 audio recording device including a smartphone or a personal computer, therefore greatly reduces the cost of data collection and making large-scale data collection possible. In additional, most of the researches are still focusing on sleep stage prediction

[65] [58], however according to our experiments, the sleep stage sequence doesn't work very well on sleep quality prediction.

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]. The accuracy of these devices is still controversial, according to [40], 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 [49] 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 [19]. 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 [13]. 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 [26], and Hao et al. developed an application called iSleep [27]. Gu used not only audio data but also data from the accelerometer and light sensor, which limited the range of the available equipment. Hao used only audio data; however their ground truth is another high-quality audio 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[4] and Sleep

[1] https://en.wikipedia.org/wiki/Zeo,_Inc.
[2] http://www.beddit.com/
[3] https://www.fitbit.com/
[4] http://sleep.urbandroid.org/

Cycle alarm clock[5]; however, no academic proof or accuracy evaluation for their outputs exists, which is consistent with [4], that the authors mentioned very few of the apps are based on published scientific evidence.

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 [11]. 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 [42].

2.2 Methods applied in this research

This section introduced the algorithms and techniques those were applied or provided inspiration in our research.

2.2.1 Burst extraction algorithm

The audio events were extracted by the statistical burst extraction method [35]. 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 $(t = 1, \cdots, t_{end})$ be the amplitude at time t, and audio signals are assumed to be generated from a Gaussian probability density function:

$$f_j(z_t) = \frac{1}{\sqrt{2\pi}\sigma_j} \exp\Big\{-\frac{(z_t - \mu)^2}{2\sigma_j^2}\Big\} \quad (j = 0, \cdots, L), \tag{2.1}$$

where $\mu = \sum_t z_t / t_{end}$ is the mean for all audio signals, σ_0(steady state) is the variance of all audio 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 audio signals are generated by the different variances of the Gaussian functions.

[5] `http://www.sleepcycle.com/`

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 $j = 0, \cdots, L$ by the following equation:

$$\mathrm{Cost}_j(t) = -\ln f_j(z_t) + \min_{0 \leq l \leq v} \Big\{ \mathrm{Cost}_l(t-1) + \tau(l, j) \Big\}, \tag{2.2}$$

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

$$\tau(l, j) = \begin{cases} (j - l)\gamma \ln t_{end} & \text{if } j > l \\ 0 & \text{otherwise,} \end{cases} \tag{2.3}$$

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, audio events are obtained by extracting areas where the burst level is greater than 1 $(j \geq 1)$.

2.2.2 Self-organizing map

The self-organizing map (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 [37], for the monitoring of machinery [51], 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 our research. The generation of such a map has the following advantages:

Comprehensive evaluation: Similar audio events are assumed to have similar frequency characteristics. Therefore, a cluster of audio 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 $\mathbf{x}_n = (x_{n,1}, \cdots, x_{n,v}), (n = 1, \cdots, N)$, the position of M neurons in the visualization layer be $\mathbf{r}_j = (\xi_j, \cdots, \eta_j), (j = 1, \cdots, M)$, and the reference vector corresponding to the jth neuron be $\mathbf{m}_j$ (v-dimension).

The following describes the algorithm of the batch type SOM:

Step 1: Initialize the reference vectors $\{\mathbf{m}_1, \cdots, \mathbf{m}_M\}$ randomly and set the iteration step as $t = 1$.

Step 2: Search winner neurons $\{c(\mathbf{x}_1), \cdots, c(\mathbf{x}_N)\}$ for all inputs by the nearest neuron:

$$c(\mathbf{x}_n) = arg \min_j \|\mathbf{x}_n - \mathbf{m}_j\|, \tag{2.4}$$

Step 3: Exit if the best matching units $\{c(\mathbf{x}_1), \cdots, c(\mathbf{x}_N)\}$ were not changed or the iteration reached $t = tmax$.

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

$$\mathbf{m}_j^{new} = \mathbf{m}_j + h_{c(\mathbf{x}_n),j}[\mathbf{x}_n - \mathbf{m}_j], \tag{2.5}$$

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

$$h_{c(\mathbf{x}_n),j} = \alpha \exp\big(-\frac{\|\mathbf{r}_j - \mathbf{r}_{c(\mathbf{x}_n)}\|^2}{2\sigma^2}\big), \tag{2.6}$$

Step 5: Decrease the the learning parameters α and σ ,and increase the iteration counter

$t \to t+1$. Then, return to Step 2.

2.2.3 Kernel method

In general, a kernel method extends a linear classification or clustering method to obtain a non-linear classification or clustering method or introduces an appropriate similarity measure to existing methods using a kernel function[8]. The representative classifier using a kernel method is a SVM, which has been used successfully in various domains, such as speech recognition, document classification, and bioinfomatics[16]. Here, it is important to select or construct a kernel function that adapts to the objective domain. The typical kernel functions are RBF (Gaussian), Polynomial, and Sigmoid kernels, whereas the KL kernel is used in the present study. Ishigaki and Higuchi[31] revealed that the KL kernel is robust for shifting of the spectrum distribution and applied an SVM with a KL kernel for failure diagnosis of a pressure adjuster providing accurate diagnosis, as compared to using general kernel functions.

Formally, given N inputs $\mathbf{x}_1, \cdots, \mathbf{x}_N \in \mathbb{R}^v$, the kernel function $K : \mathbb{R}^v \times \mathbb{R}^v \to \mathbb{R}$ is a function that satisfies the following conditions:

- **Symmetry** $K(\mathbf{x}_i, \mathbf{x}_j) = K(\mathbf{x}_j, \mathbf{x}_i)$.
- **Positive definite** For all inputs $\mathbf{x}_1, \cdots, \mathbf{x}_N$ and arbitrary real numbers $\alpha_1, \cdots, \alpha_N$, $K(\mathbf{x}_i, \mathbf{x}_j)$ satisfies $\sum_i \sum_j \alpha_i \alpha_j K(\mathbf{x}_i, \mathbf{x}_j) > 0$.

Under these conditions, the existence of a mapping function $\phi : \mathbb{R}^v \to \mathbb{H}$, which maps the data to a different space from the original space, is proven (Mercer's theorem). The kernel function is then defined by:

$$K(\mathbf{x}_i, \mathbf{x}_j) =< \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) >, \tag{2.7}$$

where $<,>$ represents an inner product. Here, since only the inner products of data points appear within algorithms such as SVM and SOM, the algorithms require an inner product of data points. This means that a mapped vector $\phi(\mathbf{x}_i)$ is not necessary to calculate, if the inner product is directly calculated. Using this mathematical property,

the kernel method computes within the mapped higher-dimensional space.

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 [22], proposed the use of 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[2] [10] to cluster the sleep-related audio events.

RBF kernel

The RBF kernel function is defined as

$$K_{RBF}(\mathbf{x}_i, \mathbf{x}_j) = \exp\big(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\big), \tag{2.8}$$

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are vectors in the input space, and σ is a free parameter.

Polynomials kernel

For degree-d polynomials, the polynomial kernel function is defined as:

$$K_{PL}(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j + 1)^d, \tag{2.9}$$

KL kernel

Let $v-$discrete points of a frequency spectrum be $\mathbf{x}_i = (x_{i,1}, \cdots, x_{i,v})$. Then, the KL kernel function is defined as:

$$K_{KL}(\mathbf{x}_i, \mathbf{x}_j) = \exp\big(-\beta JS(\mathbf{x}_i, \mathbf{x}_j)\big), \tag{2.10}$$

$$\begin{aligned} JS(\mathbf{x}_i, \mathbf{x}_j) &= KL(\mathbf{x}_i, \mathbf{x}_j) + KL(\mathbf{x}_j, \mathbf{x}_i) \\ &= \sum_{k=1}^{v} \big\{x_{i,k} \log \frac{x_{i,k}}{x_{j,k}} + x_{j,k} \log \frac{x_{j,k}}{x_{i,k}}\big\}, \end{aligned} \tag{2.11}$$

where $KL(\mathbf{x}_i, \mathbf{x}_j)$ is the Kullback-Leibler divergence, which is the distance between probability distributions, $JS(\mathbf{x}_i, \mathbf{x}_j)$ denotes the Jensen-Shannon divergence, which sym-

metrizes the KL divergence, and $\beta > 0$ is a scaling parameter. Note that the spectra must be normalized as $\sum_k x_{i,k} = 1$, because the KL divergence was originally developed for a probability distribution. When applied to AE data, the KL kernel measures depending not on the power of the spectrum, but rather on the relative distribution of the spectrum.

Difference between the L2-norm and the KL kernel

This section describes the difference between the L2-norm, on which the standard SOM is based, and the KL kernel. Here, L2-norm D between two frequency spectra $\mathbf{x}_i$ and $\mathbf{x}_j = \mathbf{x}_i + \Delta\mathbf{x}$ is given by:

$$D(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{v} (x_{i,k} - x_{j,k})^2 = \sum_{k=1}^{v} \Big(x_{i,k} - (x_{i,k} + \Delta x_k) \Big)^2 = \sum_{k=1}^{v} \Delta x_k^2. \tag{2.12}$$

Whereas the component of KL kernel JS is given by:

$$\begin{aligned} JS(\mathbf{x}_i, \mathbf{x}_j) &= \sum_{k=1}^{v} (x_{j,k} - x_{i,k}) \log \frac{x_{j,k}}{x_{i,k}} \\ &= \sum_{k=1}^{v} \Delta x_k \log \Big(\frac{x_{i,k} + \Delta x_k}{x_{i,k}} \Big) \\ &= \sum_{k=1}^{v} \Delta x_k \log \Big(1 + \frac{\Delta x_k}{x_{i,k}} \Big). \end{aligned} \tag{2.13}$$

The L2-norm uniformly measures the difference $\Delta\mathbf{x}$ over the spectra, whereas the KL kernel takes into account the ratio of $x_{i,k}$ to Δx_k. Fig. 1 shows the function of Jensen-Shannon divergence where the difference of two frequency spectra are evaluated. When the power of a certain frequency $x_{i,k}$ is larger, the KL kernel underestimates Δx_k. In contrast, the smaller $x_{i,k}$, the greater the effect of Δx_k. That is, the KL kernel evaluates the difference around a low power frequency rather than that near a spectrum peak. Since a frequency spectrum has sharp peaks, the KL kernel evaluates the spectrum distribution in the sense that the differences near the spectrum peaks are underestimated to some degree.

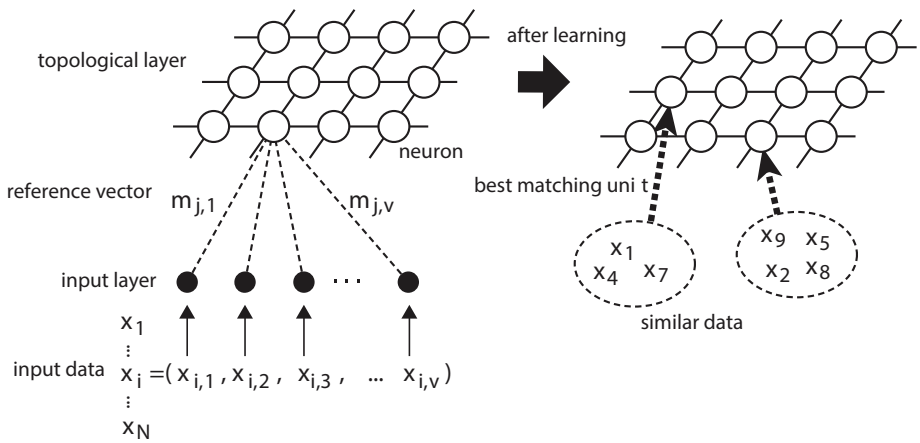

Figure 2.1: Concept of the SOM. The SOM learns similar data so as to correspond to neighbors in the neuron topology [22].

2.2.4 Kernel SOM

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 mapped $\phi(\mathbf{x})$, the following updating formula can be obtained:

$$\mathbf{m}_i(t+1) := \gamma \sum_n h_{c(\mathbf{x}_n),i} \phi(\mathbf{x}_n), \tag{2.14}$$

where t is an iteration step, and γ is a regularization term $\gamma = 1/\sum_n h_{c(\mathbf{x}_n),i}$. 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:

$$\begin{aligned} d_{i,n}(t+1) &\equiv ||\phi(\mathbf{x}_n) - \mathbf{m}_i(t+1)||^2 \\ &= K(\mathbf{x}_n, \mathbf{x}_n) - 2\gamma \sum_j h_{c(\mathbf{x}_j),i} K(\mathbf{x}_n, \mathbf{x}_j) \\ &\quad + \gamma^2 \sum_k \sum_l h_{c(\mathbf{x}_k),i} h_{c(\mathbf{x}_l),i} K(\mathbf{x}_k, \mathbf{x}_l). \end{aligned} \tag{2.15}$$

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 $\{c(\mathbf{x}_1), \cdots, c(\mathbf{x}_N)\}$ for all inputs by the nearest neuron:

$$c(\mathbf{x}_N) = \arg \min_{i=1,\cdots,M} d_{i,n}, \tag{2.16}$$

Step 3: (Same as SOM) Exit if the best matching units $\{c(\mathbf{x}_1), \cdots, c(\mathbf{x}_N)\}$ 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. (2.15).

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

2.2.5 Sequence-based SOM

The Sequence-based SOM (SbSOM) [24] is also based on SOM. Let v-dimensional N inputs be $\mathbf{x}_n = (x_{n,1}, \cdots, x_{n,v})$, $(n = 1, \cdots, N)$, the position of M neurons in the visualization layer be $\mathbf{r}_j = (\xi_j, \eta_j)$, $(j = 1, \cdots, M)$, and the reference vector corresponding to the jth neuron be $\mathbf{m}_j$ (v-dimension).

The following shows the learning algorithm that employs a batch process and decreasing strategy of the learning parameter.

Step 1. Initialize the reference vectors $\{\mathbf{m}_1, \cdots, \mathbf{m}_M\}$ randomly.

Step 2. Search winner neurons $\{c(\mathbf{x}_1), \cdots, c(\mathbf{x}_N)\}$ for all inputs.(This step is described in the next subsection.)

Step 3. Exit if the winner neurons $\{c(\mathbf{x}_1), \cdots, c(\mathbf{x}_N)\}$ were not changed.

Step 4. Update the reference vectors $\{\mathbf{m}_1, \cdots, \mathbf{m}_M\}$ by the following equation.

$$\mathbf{m}_j^{new} = \mathbf{m}_j + h_{c(\mathbf{x}),j}[\mathbf{x}_n - \mathbf{m}_j],$$

where $h_{c(\mathbf{x}),j}$ is a neighborhood function that defines the effect of neighborhood of

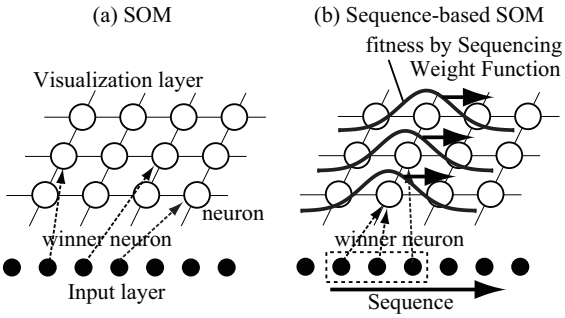

Figure 2.2: Difference in the winner neuron selection. (a) Determined only by spatio-distance in SOM. (b) Determined by spatio-distance under the sequencing weight function in SbSOM [24].

the winner. Typically, Gaussian function is used:

$$h_{c(\mathbf{x}),j} = \alpha \exp\Big(-\frac{\|\mathbf{r}_j - \mathbf{r}_{c(\mathbf{x})}\|^2}{2\sigma^2}\Big).$$

Step 5. Decrease the learning parameters α and σ every several iterations. Return to Step 2.

Sequencing weight function

In the normal SOM, the winner neuron is determined only by spatio-distance. Meanwhile in SbSOM, by introducing Sequencing Weight Function (SWF) weights are given onto the neuron topology according to the sequence of the data (Fig.2.2). The SWF introduces the concept of time onto the topology[6]. The winner is determined by spatio-temporal distance utilizing SWF $\psi(n, \xi_j)$ as follows:

$$c(\mathbf{x}_n) = \arg\min_j \psi(n, \xi_j)||\mathbf{x}_n - \mathbf{m}_j||. \tag{2.17}$$

The nth data is located at ratio of n/N within the data sequence, and the jth neuron is located at ratio of ξ_j/ξ_M to certain direction on the topology (in this case, ξ-direction). Let the absolute value of those difference be $\epsilon = |\xi_j/\xi_M\text{-}n/N|$. The SWF is defined so as to be able to take a balance between spatio/temporal resolution, by allowing reversal

[6]It is loose time concept since it allows reversal of data order which appears onto the neuron topology by the winner.

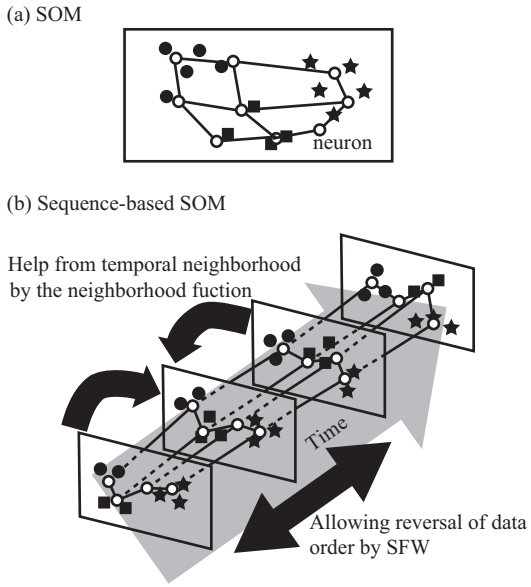

Figure 2.3: (a) Spatio-neighborhood becomes topological neighborhood in SOM. (b) Spatio-temporal neighborhood becomes topological neighborhood in SbSOM. Note that this figure shows a restricted version of SbSOM [24].

of data order:

$$\psi_{exp}(n, \xi_j) = e^{w\epsilon}, \tag{2.18}$$

where $w \geq 0$ is a parameter that controls influence degree of data order, i.e. temporal distance. The larger w is, the more reversal of data order may be restricted. Instinctively, the SWF travels to a certain direction on the neuron topology (Fig.2.2). Note that when $w = 0$ (i.e., $\psi(n, \xi_j) = 1$) SbSOM will be exactly the same as the standard SOM.

In case of NOT allowing reversal of data order, the SWF is given as:

$$\psi_{strict}(n, \xi_j) = \begin{cases} 1 & if \quad \epsilon < \frac{1}{2K}, \\ \infty & otherwise, \end{cases} \tag{2.19}$$

where K is the number of neurons to ξ-direction. When w in eq.(2.18) is taken sufficiently large, it is equivalent to eq.(2.19). However, it is beneficial for computational cost to be given by window function as eq.(2.19), because it needs only comparison with reference vectors within a window.

Neighborhood function in SbSOM

The SOM has the neighborhood function so that the reference vectors are updated affected by the spatio-neighborhood. The neighborhood function in SbSOM, on the other hand, has the meaning of spatio-temporal neighborhood owing to SWF (Fig.2.3). Therefore, SbSOM can perform clustering taking the help of temporal neighborhood data. This property mitigates the problems of an appropriate window size and decrease of sample data in window-based clustering. Moreover, cluster correspondence can be self-organized since spatio-temporal neighborhood becomes topological neighborhood in SbSOM.

2.2.6 Hierarchical clustering

In data mining and statistics, hierarchical clustering (HC, also called hierarchical cluster analysis or HCA) is a method of cluster analysis which seeks to build a hierarchy of clusters. Strategies for hierarchical clustering generally fall into two types [47]:

Agglomerative: This is a "bottom up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.

Divisive: This is a "top down" approach: all observations start in one cluster, and splits are performed recursively as one moves down the hierarchy.

In general, the merges and splits are determined in a greedy manner. The results of hierarchical clustering are usually presented in a dendrogram. In the general case, the complexity of agglomerative clustering is $\mathcal{O}(n^2 \log(n))$ [47], which makes it too slow for large data sets. Divisive clustering with an exhaustive search is $\mathcal{O}(2^n)$, which is even worse.

2.2.7 Hidden Markov model

The Hidden Markov model (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

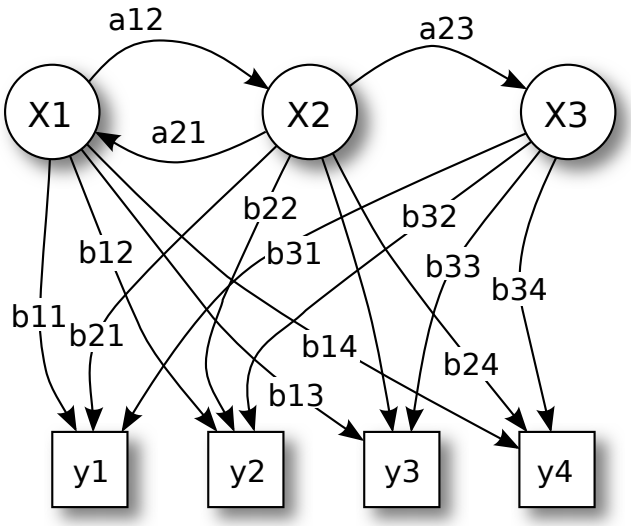

Figure 2.4: Probabilistic parameters of a HMM (example) [59].

the start probability vector $\mathbf{\Pi}$ 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 $\mathbf{\Pi}, \mathbf{A}$ and $\mathbf{\Theta_i}$ [45].

In its discrete form, a hidden Markov process can be visualized as a generalization of the Urn problem with replacement (where each item from the urn is returned to the original urn before the next step) [45]. Consider this example: in a room that is not visible to an observer there is a genie. The room contains urns $X1, X2, X3$, each of which contains a known mix of balls, each ball labeled $y1, y2, y3,$. The genie chooses an urn in that room and randomly draws a ball from that urn. It then puts the ball onto a conveyor belt, where the observer can observe the sequence of the balls but not the sequence of urns from which they were drawn. The genie has some procedure to choose urns; the choice of the urn for the n-th ball depends only upon a random number and the choice of the urn for the $(n1)$-th ball. The choice of urn does not directly depend on the urns chosen before this single previous urn; therefore, this is called a Markov process. It can be described by the upper part of Fig. 2.4.

The Markov process itself cannot be observed, only the sequence of labeled balls, thus this arrangement is called a "hidden Markov process". This is illustrated by the

lower part of the diagram shown in Fig. 2.4, where one can see that balls $y1, y2, y3, y4$ can be drawn at each state. Even if the observer knows the composition of the urns and has just observed a sequence of three balls, e.g. $y1, y2andy3$ on the conveyor belt, the observer still cannot be sure which urn (i.e., at which state) the genie has drawn the third ball from. However, the observer can work out other information, such as the likelihood that the third ball came from each of the urns.

2.2.8 Support vector machine

In machine learning, support vector machines (SVMs, also support vector networks) [15] are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. When data are not labeled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups. The clustering algorithm which provides an improvement to the support vector machines is called support vector clustering [5].

More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks like outliers detection [53].

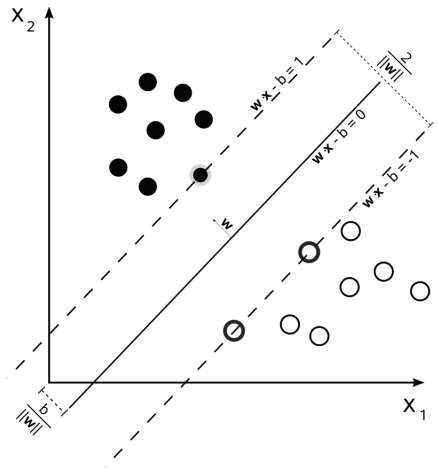

Figure 2.5: Maximum-margin hyperplane and margins for an SVM trained with samples from two classes. Samples on the margin are called the support vectors [60].

SVMs can be used to solve various real world problems:

SVMs are helpful in text and hypertext categorization as their application can significantly reduce the need for labeled training instances in both the standard inductive and transductive settings.

Classification of images can also be performed using SVMs. Experimental results show that SVMs achieve significantly higher search accuracy than traditional query refinement schemes after just three to four rounds of relevance feedback. This is also true of image segmentation systems, including those using a modified version SVM that uses the privileged approach as suggested by Vapnik [3].

The SVM algorithm has been widely applied in the biological and other sciences. They have been used to classify proteins with up to 90% of the compounds classified correctly. Permutation tests based on SVM weights have been suggested as a mechanism for interpretation of SVM models [17].

2.2.9 Mel Frequency Cepstral Coefficient

In sound processing, an Mel Frequency Cepstral (MFC) represents the short-term power spectrum of a sound based on the linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency.

Mel Frequency Cepstral Coefficient (MFCC) collectively make up an MFC [18].

MFCCs are derived from a type of cepstral representation of the audio clip (a non-linear "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 sound representation, e.g., in audio compression. In our experiment, 12 MFCCs were extracted for each sound clip as an MFCC vector, and the Euclidean distance between the MFCC vectors was considered to represent the similarity between sound events.

2.2.10 Weighted pairwise F-measure

The original F-measure evaluates the correlation between a cluster assignment and a class label. However, particularly in an SOM visualization, the neighborhood relation is also important. We employed the weighted version of the pairwise F-measure to evaluate the SOM visualization comprehensively. Using events scored through PSG as ground truth labels, we applied weighted pairwise F-measure (wPF) [23] to evaluate the clustering results. wPF is an extension of pairwise-based cluster validity measures [62], that introduces a likelihood function that indicates the degree to which a pair of data elements belongs to the same cluster rather than the actual number of data pairs.

Table 2.1 shows a class and cluster confusion matrix of data pairs where a, b, c, and d represent the number of data pairs wherein $\mathbf{x}_i$ and $\mathbf{x}_j$ do or do not belong to the same class/cluster.

Table 2.1: Class and cluster confusion matrix of data pairs

	$t(i) = t(j)$	$t(i) \neq t(j)$
$c(i) = c(j)$	a	b
$c(i) \neq c(j)$	c	d

The $likelihood(c(i) = c(j))$ is introduced to indicate the degree to which that a data pair $\mathbf{x}_i$ and $\mathbf{x}_j$ belongs to the same cluster rather than the actual number of data pairs. The likelihood is given by the inter-cluster distance of the data pair. In

this study, the following Gaussian function is used: $likelihood(c(i) = c(j)) = h_{c(i),c(j)}$: $h_{i,j} = \exp(-d_{i,j}/\sigma^2)$, where $d_{i,j}$ denotes inter-cluster distance and $\sigma(> 0)$ is a smoothing (neighborhood) radius. Then, a, b, and c are replaced by a summation of the likelihoods as follows:

$$a' = \sum_{\{i,j|t(i)=t(j)\}} h_{c(i),c(j)}, \tag{2.20}$$

$$b' = \sum_{\{i,j|t(i)\neq t(j)\}} h_{c(i),c(j)}, \tag{2.21}$$

$$c' = \sum_{\{i,j|t(i)=t(j)\}} \left(1 - h_{c(i),c(j)}\right) = a + c - a'. \tag{2.22}$$

Using the extended a', b', and c', the weighted pairwise accuracy and pairwise F-measure can be defined as follows:

$$\mathrm{wPF}(\mathbf{C}) = \frac{2 \cdot P \cdot R}{P + R}, \tag{2.23}$$

where $P = a'/(a' + b')$ is precision, which is a measure of the same class among each cluster, and $R = a'/(a' + c')$ is recall, which is a measure of the same cluster among each class. The original pairwise F-measure is the harmonic average of precision and recall. The wPF is based on the degree to which the data pairs belong to the same cluster.

Chapter 3

Sleep pattern visualization

This chapter introduces a method to identify sleep patterns by analyzing sleep-related audio events based on extended SOM algorithms. This proposed method combines the advantages of kernelization and sequence-based technologies, to obtain a fine-grained map that visualizes the distribution and changes of sleep-related audio events.

The experimental results indicate that sleep-related audio events are related to sleep stages. These results empirically warrant future study of personal sleep quality using audio data.

3.1 Methodology

In this section, we introduced the work flow of the system and the proposed method on sleep pattern visualization.

3.1.1 Overview

The work flow of this study (Figure 3.1) included the following steps.

Audio recording: Audio data were captured and converted from audio to text format using audio processing software.

Event extraction: Audio clips of events were extracted from the recorded audio data and a burst extraction algorithm (Section 2.2.1) was applied.

Data preprocessing: A time threshold (100 ms) was set to filter out very short audio

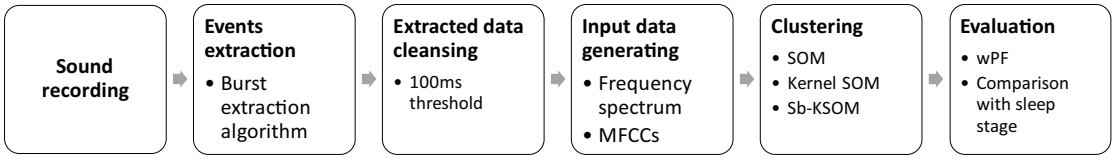

Figure 3.1: System work flow

clips that cannot be annotated manually.

Generating input data: FFT was applied to obtain the frequency power spectrum of audio clips as an input vector. The MFCCs (Section 2.2.9) of the audio clips were derived as another kind of feature.

Clustering: Using the obtained frequency power spectrum, which is a vector of discretized frequencies, or MFCCs as input vectors respectively, we applied a standard SOM (Section 2.2.2), kernel SOM (Section 2.2.4) and the proposed Sb-KSOM (Section 3.1.2) to the processed data to obtain cluster maps.

Evaluation: We performed a quantitative comparison of the cluster results from different algorithms. The wPF was used to assess the validity of each cluster (Section 2.2.10). 3.9). We also compared cluster maps relative to visualization and performed a comparative analysis of cluster maps generated by the Sb-SOM and sleep stage sequences.

3.1.2 Proposed method: Sb-KSOM

A comparison of clustering results demonstrates that kernel SOM exhibits better performance than standard SOM in audio data clustering, and KL kernel works better than RBF and polynomial kernels. Based on these results, we propose the Sb-KSOM algorithm, which is an extension of the Sb-SOM. The proposed Sb-KSOM kernelizes the Sb-SOM by replacing the Euclidean distance with the KL divergence, which enables the proposed method to handle frequency spectrum data.

In the proposed Sb-KSOM, we replaced the normal Euclidean distance calculation in Eq. (2.17) with the kernel function. The batch type Sb-KSOM algorithm is described as follows.

Step 1: Initialize all dissimilarities between the reference vectors and data points

$\forall i, n$ $d_{i,n}$ randomly and set the iteration step to $t = 1$.

Step 2: Search the best matching units $\{c(\mathbf{x}_1), \cdots, c(\mathbf{x}_N)\}$ for all inputs relative to spatiotemporal distance by utilizing the SWF $\psi(n, \xi_j)$ as follows (dissimilarity between the jth reference vector and n^{th} data point : $d_{j,n}$ is calculated using Eq. (2.15)):

$$c(\mathbf{x}_n) = \arg\min_j \psi(n, \xi_j) d_{j,n}. \tag{3.1}$$

Step 3: Exit if the best matching units $\{c(\mathbf{x}_1), \cdots, c(\mathbf{x}_N)\}$ were not changed or the iteration reached $t = t_{max}$.

Step 4: Update the dissimilarity of each reference vector to all inputs $\forall n$ $d_{i,n}$ using Eq. (2.15).

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

3.2 Experiment

We first applied the standard SOM with two types of similarity measures and three types of kernel SOM to the extracted audio data to compare the wPF. We then determined the property of KL kernel and evaluated the performance of the Sb-SOM and the proposed Sb-KSOM relative to the wPF. Then, we compared the visualization between the standard SOM and the KL-KSOM.

We also applied the proposed Sb-KSOM to the data to obtain a spatiotemporal dimensional cluster map. Here, we examined the relationship between the transition of sleep stages and the cluster dynamics of audio events.

In the final part of this experiment, we calculated the conditional probabilities of the primary sleep disorder events in different sleep stages. The disorder events include snoring, teeth grinding and body movement.

Table 3.1: Subject and audio data information

Subject id	Age	Gender	Recording date	Duration	Primary disorder symptoms
1	21	F	2014/05/13	08:05:22	snoring
2	22	M	2014/05/27	08:16:15	teeth grinding
3	22	M	2014/06/03	08:01:09	snoring
4	23	M	2014/07/29	08:23:01	teeth grinding, snoring
5	24	M	2015/01/20	08:17:34	teeth grinding, snoring
6	23	F	2015/03/03	08:30:30	snoring
7	20	F	2015/06/02	07:18:30	teeth grinding
8	23	M	2015/08/03	08:23:30	snoring
9	22	M	2015/09/29	08:01:22	snoring
10	23	M	2015/10/14	08:21:53	snoring

3.2.1 Experimental setting

The data used in this study were prepared by the Graduate School of Dentistry, Osaka University. The study protocol was approved by the research ethics committee of the Osaka University Graduate School of Dentistry and the Osaka University Dental Hospital. Written informed consent was obtained from all subjects. All subjects were asked to sleep in a specific room (Figure 3.2) from 22:30 to 8:00. The experiment devices included a sound level meter: LA1250 (Ono Sokki)[6] and a recorder: R-4 Pro (Roland)[7]. The detail specification of recorder R-4 Pro is listed in Table 3.2. A microphone was placed 50 cm from the subjects' heads. The audio data were recorded using a single channel (mono) at a sampling rate of 48 kHz. In addition, all subjects were measured by PSG simultaneously.

Table 3.2: R-4 Pro specification

Channels	4
Signal Processing	AD/DA Conversion: 24 bits Sampling Frequency: 44.1/48/88.2/96/192 kHz
Data Type	Format: BWF, WAV Bit Depth: 16/24 bits Sampling Frequency: 44.1/48/88.2/96/192 kHz
Recording Media	Internal Hard Disk Drive (80 GB)

All subjects were university students from Osaka University (20-24 years), and the male to female ratio was balanced. Table 3.1 shows information about the subjects and

[6] https://www.onosokki.co.jp/English/hp_e/products/keisoku/s_v/la1200.html
[7] http://proav.roland.com/products/r-4_pro/

Figure 3.2: Experimental environment

the recorded audio data used in the experiment.

The audio data were stored and processed on a Linux server (two Intel Xeon 12-Core 2.7GHz CPUs with 128 GB of memory). Note that processing (including burst extraction, spectrum generation, and clustering) a dataset of 10000 audio events required five hours.

3.2.2 Event extraction

We selected audio data from ten nights. Based on the burst extraction method, we obtained 14635 audio events (hyperparameters $L = 6$, $s = 1.5$, and $\gamma = 100$). The hyperparameters were tuned manually in a certain range to extract as many useful events as possible, and keep extracted useless noise at an acceptable level. However, the noise data were also clustered and could be separated from sleep-related events. Note that including some noise does not pose any problems for the clustering results. To validate the clustering results, we annotated all events by listening to the audio data.

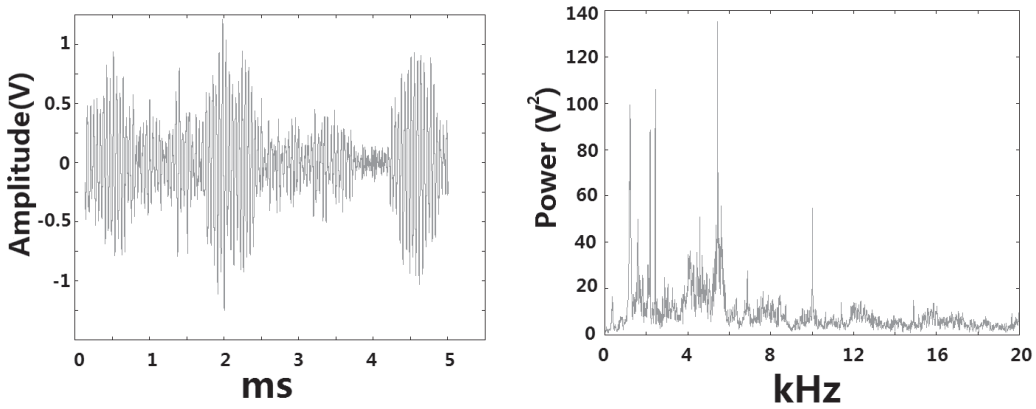

Figure 3.3: An example of an extracted audio wave and frequency spectrum

The events were categorized as snoring, teeth grinding, body movement, human voice, and environmental sounds, including air conditioner and outdoor traffic sounds. FFT was applied to the extracted audio data to obtain the frequency power spectrum. From 24 Hz to 20 kHz (at intervals of 4 Hz), 4995 discretized points as input for the SOM were obtained from a audio event. Fig. 3.3 shows an example of an extracted audio event and its preprocessed frequency spectrum.

In consideration of annotation deviation, we evaluated anotation accuracy using Cohen's kappa coefficient [52]. We asked four people to annotate the same 50 events. The average kappa value in this experiment was 0.801; thus we consider the annotation reliable.

3.2.3 Quantitative comparison between standard and kernelized clustering

In the first part of this experiment, we used the audio data from each subject as a respective dataset and compared the wPF values for each subject between the standard and kernelized algorithms, including standard SOMs based on the frequency spectrum or MFCC similarity and kernel SOMs with KL, RBF, and polynomial kernels. The number of neurons was set to 15×15 with a two-dimensional regular grid. Generally, the number of neurons is not sensitive to these results in that an SOM captures the data distribution in the feature space. Here, a Gaussian function was used as the neighborhood

function. To avoid initial value dependency, the experiments were executed 50 times and the average values were computed. The hyperparameters of the kernel functions were tuned by linear search. The mean wPF values and standard deviations are shown in Table 3.3. The average wPF value shows that the MFCC feature did not perform well with the acquired audio data, similar results can also be found in [64]. The KL-KSOM demonstrated the best performance, with an approximate 12% improvement compared to the standard SOM.

Once the KL kernel property was determined, we compared the Sb-SOM and Sb-KSOM relative to wPF. For both sequence-based algorithms, the cluster distributions in the temporal direction were the same; therefore, we modified the wPF slightly, i.e., the inter-cluster distance $d_{i,j}$ was calculated only in the spatial direction. The mean wPF values and standard deviations are shown in Table 3.4. As can be seen, the average wPF of the proposed Sb-KSOM improved by approximately 6% compared to Sb-SOM.

Table 3.3: Comparison of wPF between standard SOM and kernel SOM clustering results

Subject id	$SOM_{spectrum}$		SOM_{MFCC}		$KSOM_{KL}$		$KSOM_{RBF}$		$KSOM_{PL}$	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	0.537	0.033	0.509	0.041	**0.604**	0.037	0.593	0.051	0.504	0.047
2	0.521	0.041	0.497	0.043	0.573	0.038	**0.577**	0.038	0.482	0.025
3	0.506	0.031	0.493	0.035	**0.551**	0.031	0.532	0.033	0.535	0.042
4	0.559	0.040	0.482	0.032	**0.592**	0.037	0.561	0.035	0.567	0.026
5	0.602	0.039	0.594	0.039	**0.629**	0.039	0.624	0.041	0.608	0.054
6	0.543	0.033	0.549	0.045	**0.600**	0.035	0.562	0.038	0.557	0.037
7	0.483	0.042	0.501	0.035	0.523	0.047	0.531	0.051	**0.537**	0.036
8	0.582	0.030	0.552	0.032	**0.711**	0.056	0.654	0.049	0.664	0.045
9	0.503	0.032	0.464	0.031	**0.644**	0.031	0.612	0.035	0.623	0.048
10	0.505	0.046	0.433	0.055	**0.574**	0.045	0.548	0.044	0.519	0.037
Mean	0.534	0.037	0.507	0.039	**0.600**	0.040	0.580	0.042	0.560	0.040

3.2.4 Comparison between standard SOM and KL-KSOM visualization

In the second part of this experiment, all extracted data were combined into a single dataset to demonstrate the differences among various disorders. In Figure 3.4, a cell corresponds to a neuron and similar disorder events are clustered into the same or neighboring neuron. Note that the coordinates in the cluster map do not express any physical

Table 3.4: Comparison of wPF between Sb-SOM and Sb-KSOM clustering results

Subject id	Sb-SOM		Sb-KSOM	
	Mean	SD	Mean	SD
1	0.531	0.035	0.571	0.043
2	0.506	0.032	0.511	0.036
3	0.488	0.059	0.496	0.035
4	0.518	0.042	0.569	0.049
5	0.571	0.039	0.568	0.039
6	0.490	0.056	0.532	0.054
7	0.475	0.031	0.514	0.042
8	0.553	0.044	0.608	0.041
9	0.492	0.036	0.559	0.044
10	0.533	0.044	0.546	0.052
Mean	0.516	0.042	0.547	0.044

quantity other than the relative distances to the neighboring data.

We marked each neuron by majority decision, and the different types of neurons are shown by different colors on the map. The teeth grinding events are dispersed into five groups in Figure 3.4(a) and two groups in Figure 3.4(b). The snoring events are dispersed into four groups in Figure 3.4(a) and two groups in Figure 3.4(b). By comparing the results of the SOM and KL-KSOM, we can intuitively realize that the same type of events was better concentrated by the KL-KSOM.

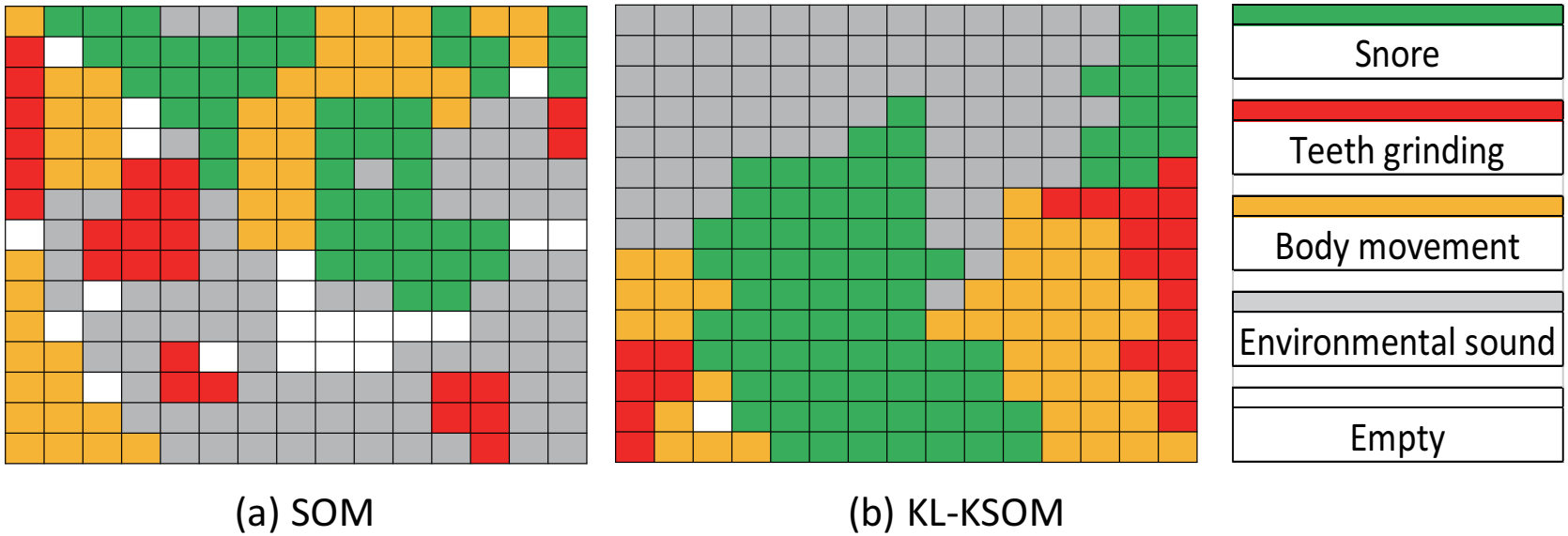

Figure 3.4: Cluster maps generated by standard SOM and KL-KSOM

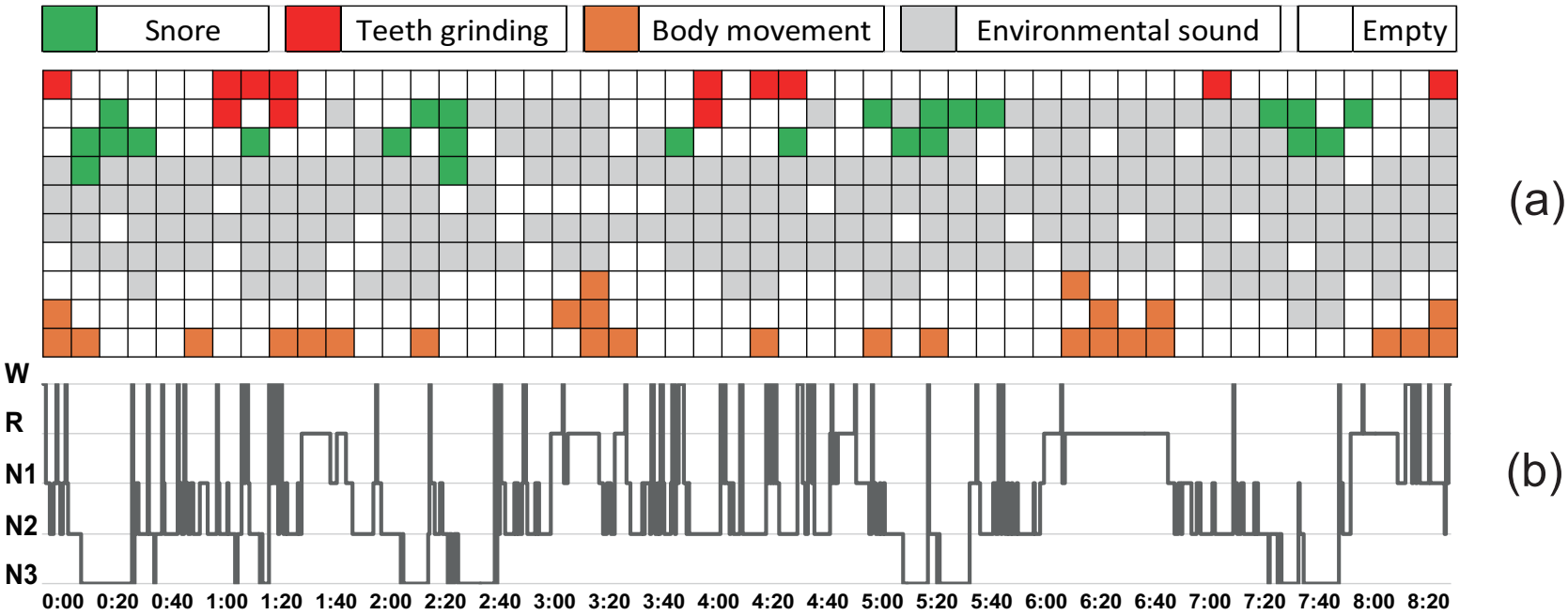

Figure 3.5: Cluster map generated by proposed Sb-KSOM for Subject 4

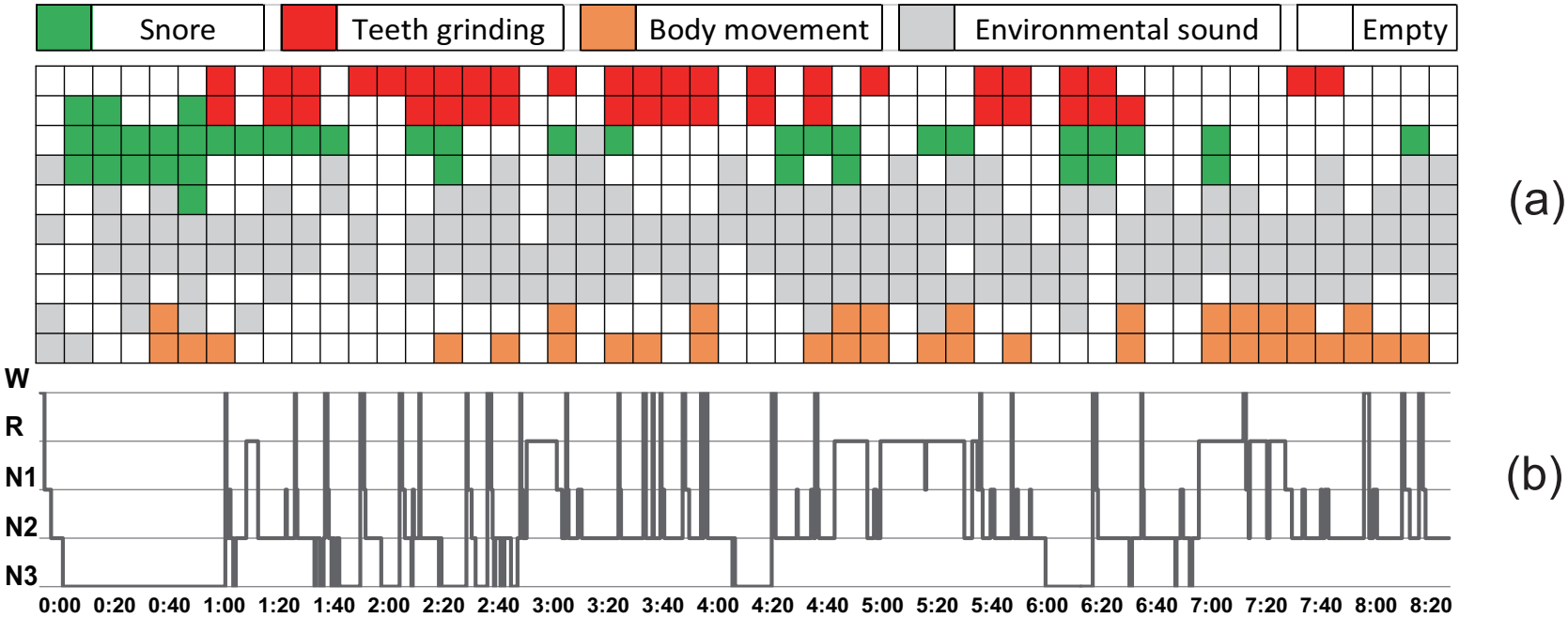

Figure 3.6: Cluster map generated by proposed Sb-KSOM for Subject 2

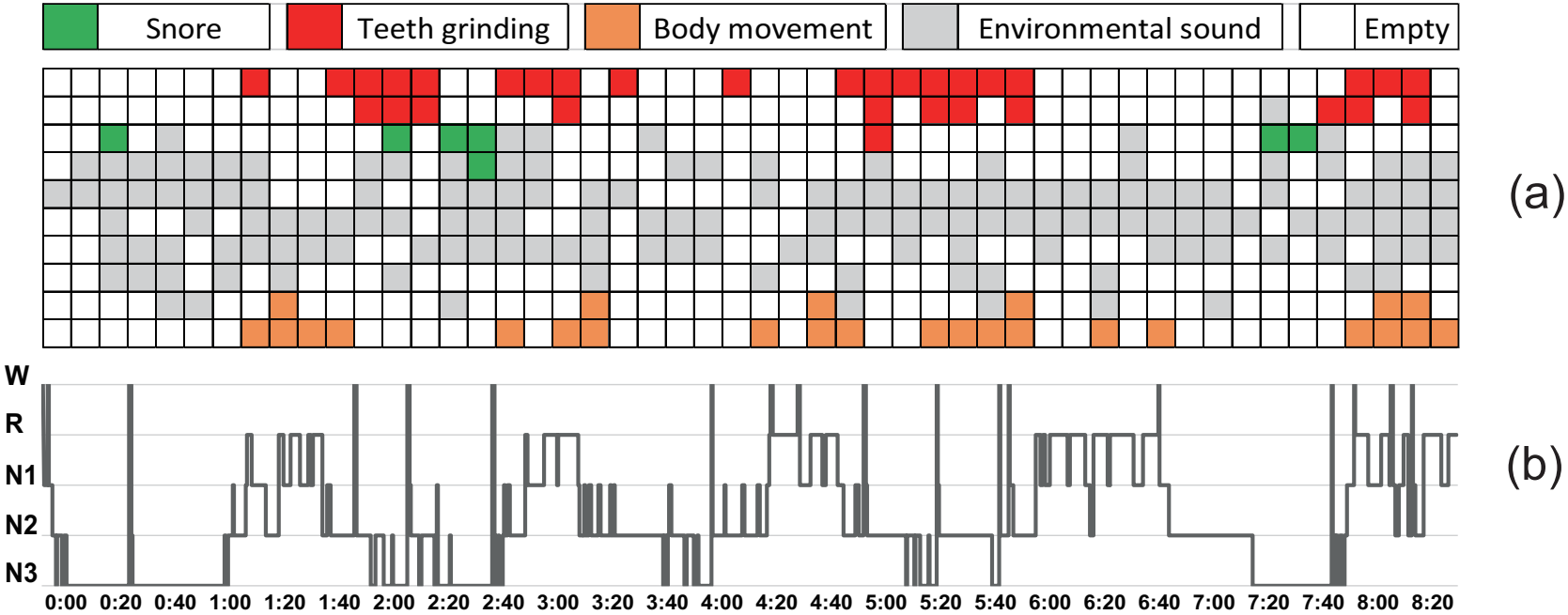

Figure 3.7: Cluster map generated by proposed Sb-KSOM for Subject 7

3.2.5 Sleep pattern analysis

Sleep pattern visualization by Sb-KSOM

In this experiment, we performed a comparative analysis to reveal the relationship between the cluster maps generated by the proposed Sb-KSOM and the sleep stage sequences. Here, all subjects were analyzed respectively. Three subjects' clustering results are discussed in this section. These subjects were chosen because their teeth grinding or snoring activities were frequent and the generated more related audio events. Figure 3.5(a) shows the results obtained when the proposed Sb-KSOM was applied to the audio data from Subject 4. Here, the number of neurons was set 50×10 with a two-dimensional grid. The subjects' sleep stages were scored by a medical specialist based on PSG data from the same night using a 30-s time window. The sleep stage sequence of Subject 4 is shown in Figure 3.5(b), where the REM stage is shown as "R" and the awake stage is shown as "W". We defined a period that contain continuous N3 stages with intervals between other stages of less than three minutes as a deep sleep period, and periods without deep sleep, awakening stages and REM stages as light sleep periods. Since the REM stage is a unique phase in the sleep process, we discuss it separately.

The sleep periods of Subject 4 were interpreted as follows:

Deep sleep periods (0:13:30 - 0:31:30), (2:09:30 - 2:42:00), (5:09:30 - 05:33:00), (7:25:30 - 7:45:30): There were many snoring events during these periods, very few body movements, and no teeth grinding. We found that the cluster center of the snoring events was mostly associated with a deep sleep period.

REM stages (1:33:00 - 1:43:00), (3:02:00 - 3:19:30), (4:43:00 - 4:51:30), (6:00:00 - 6:05:30), (6:07:30 - 6:44:00), (7:50:30 - 8:07:00): Compared to other stages, REM stages have a stronger association with body movement clusters and a weaker association with snoring and teeth grinding clusters.

Light sleep periods: In each light sleep period, there were some teeth grinding and body movement event clusters and a few snoring events.

We found that the distribution of audio event clusters changed simultaneously with changes to the sleep stage, for all subjects. Even though our analysis included other

subjects who have different primary sleep disorders and varying sleep stages patterns, the findings led to similar conclusions. For example, with Subjects 2 (Figure 3.6) and 7 (Figure 3.7), the deep sleep periods were strongly associated with snoring clusters, the number of body movements was notably greater in light periods and REM stages than in the deep periods, and very few snoring clusters were found in REM stages.

Similar discussions can be found in the literature. For example, teeth grinding most frequently occurs during NREM sleep [32], and the REM stage is always associated with dreaming [29], which triggers several body movements.

We found that the transition of cluster dynamics and changing sleep stages are related. The sleep stage sequence is an important tool in the study of sleep patterns, and its relations provide the possibility of discovering sleep patterns based on the cluster map of sleep-related audio data from an Sb-KSOM.

Quantifying the relationship between sleep stages and audio events

To quantify the correlation between sleep-related audio events and sleep stages, we calculated the conditional probabilities of a audio event given the current sleep stage. Sleep stages were scored every 30 s, and each column of the Sb-KSOM results indicates a 10-min time window; therefore, the proportion of each sleep stage in each time window was calculated and applied to weight the amount of audio events. Here, N1 and N2 stages are combined as "LIGHT", and the N3 stage is showed as "DEEP". We calculated the amount of each event in each sleep stage as follows:

$$A_{e,s} = \sum_{w=0}^{N} C_{e,w} P_{s,w}, \tag{3.2}$$

where $e \in E, (E = \{body\_movement, snore, teeth\_grinding\})$, $s \in S, (S = \{WAKE, REM, LIGHT, DEEP\})$, $A_{e,s}$ is the weighted occurrence count of audio event type e in sleep stage type s, w is the time window index, $C_{e,w}$ is audio event type e's occurrence count in the wth time window, and $P_{s,w}$ is the proportion of sleep stage type s in the wth time windows.

The data from all subjects were computed together. Table 3.5 shows the weighted

Table 3.5: Weighted occurrence count of audio events

	Snore	Body Movement	Teeth Grinding
WAKE	384.8	359.6	23.6
LIGHT	1982.6	591.9	173.5
DEEP	3767.4	140	76.6
REM	439.2	939.5	46.3
TOTAL	6574	2031	320

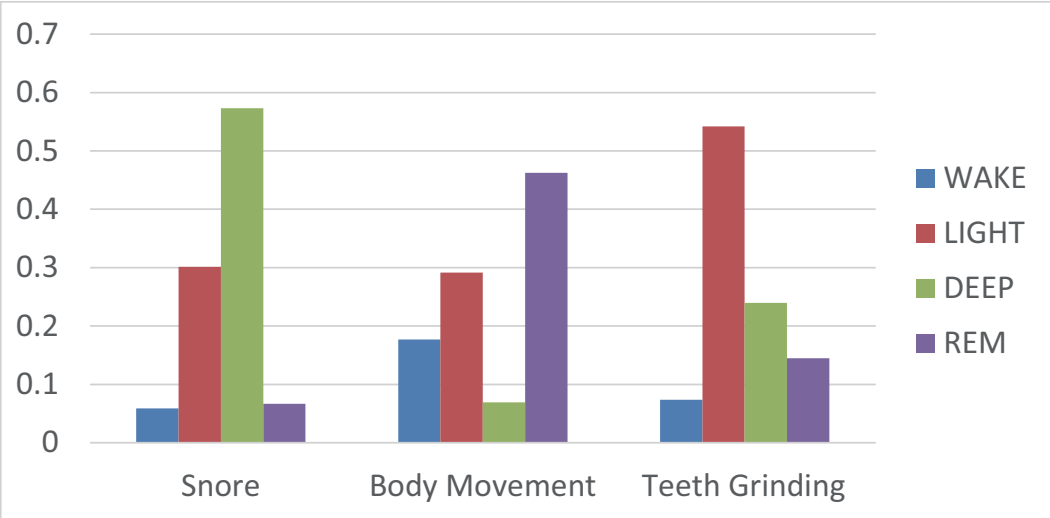

Figure 3.8: Conditional probabilities of audio event given sleep stage

occurrence count $A_{e,s}$ of audio events for each sleep stage. The conditional probabilities of audio events given a sleep stage are shown in Figure 3.8, which was calculated as follows:

$$Pr(e|s) = A_{e,s} / \sum_{k} A_{e,k}. \tag{3.3}$$

In addition, we applied Pearson's chi-squared test and addressed the fact that inconsistent conditional probabilities were not caused by sampling variation. The conditional probabilities demonstrate that snore events have the strongest relationship with deep sleep and body movement is more related to REM and light sleep than deep sleep. Teeth grinding most frequently occurs during NREM sleep. These results are consistent with our previous observations.

Chapter 4

Sleep quality prediction

In this chapter we proposed a novel approach for assess the sleep quality through audio data. We 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 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 audio 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.

4.1 Methodology

4.1.1 Overview

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

Audio recording: Recording audio data by recording device, and converting audio format data to text format data through audio 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 audio 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: Audio clips of events were extracted from the audio recordings, burst extraction algorithm is applied.

Input data preprocessing: Applying FFT to obtain the frequency power spectrum of each audio 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 to get the cluster map. Then agglomerative HC (sec. 4.1.2) 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.

Audio events categorizing: By selecting an appropriate stop-criteria for agglomerative HC through silhouette [48], the cells was divided into several major clusters, thus we obtained a virtual classifier as aforementioned. Classification was performed on all extracted audio events, then a sequence with categorized data points was obtained for each audio recording.

Modelling by HMM: According to the sleep quality labels on audio recordings, data sequences obtained from last step were divided into good and poor data sequence sets. The multinomial HMMs (sec. 4.1.3) for good or poor sleep quality were trained respectively by corresponding data sequence set.

Classification based on HMMs: The likelihoods between an input audio event sequence and obtained HMMs are used as input data for sleep quality level classification. (sec.4.1.4)

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

4.1.2 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 [47]. HC is applied in this study instead of other methods like K-Means because it is typically used to obtain major clusters in SOM [57].

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 [6].

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:

$$\begin{aligned} d_{a,b} &\equiv ||\mathbf{m}_a - \mathbf{m}_b||^2 \\ &= \gamma_a^2 \sum_i^n \sum_j^n h_{c(\mathbf{x}_i),a} h_{c(\mathbf{x}_j),a} K(\mathbf{x}_i, \mathbf{x}_j) \\ &\quad - 2\gamma_a \gamma_b \sum_k^n \sum_l^n h_{c(\mathbf{x}_k),a} h_{c(\mathbf{x}_l),b} K(\mathbf{x}_k, \mathbf{x}_l) \\ &\quad + \gamma_b^2 \sum_u^n \sum_v^n h_{c(\mathbf{x}_u),b} h_{c(\mathbf{x}_v),b} K(\mathbf{x}_u, \mathbf{x}_v), \end{aligned} \tag{4.1}$$

where $\mathbf{x}_i (i = 1, \cdots, n)$ 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. (2.15).

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 [48], 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 $\mathbf{x}_i (i = 1, \cdots, n)$ 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:

$$\begin{aligned} d_{i,n+1} &\equiv ||\phi(\mathbf{x}_{n+1}) - \mathbf{m}_i||^2 \\ &= K(\mathbf{x}_{n+1}, \mathbf{x}_{n+1}) - 2\gamma \sum_{j}^{n+1} h_{c(\mathbf{x}_j),i} K(\mathbf{x}_{n+1}, \mathbf{x}_j) \\ &\quad + \gamma^2 \sum_{k}^{n+1} \sum_{l}^{n+1} h_{c(\mathbf{x}_k),i} h_{c(\mathbf{x}_l),i} K(\mathbf{x}_k, \mathbf{x}_l), \end{aligned} \tag{4.2}$$

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. (2.15).

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:

$$c(\mathbf{x}_{n+1}) = \arg \min_{i=1,\cdots,M} d_{i,n+1}. \tag{4.3}$$

Therefore the virtual classifier for input audio event was created. Classification was performed on all extracted audio events, then a sequence with categorized data points was obtained for each audio 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.

4.1.3 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 $\mathbf{\Pi}$ 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 $\mathbf{\Pi}, \mathbf{A}$ and $\mathbf{\Theta_i}$ [45].

According to the sleep quality labels on audio 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 [44], 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 [50]. 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.

4.1.4 Classification based on HMMs

To classify the sleep quality level of a new obtained audio recording, we firstly extract the clips of sleep-related audio 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 [15] with 2 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 3 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 [21] 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.

4.2 Experiment

4.2.1 Overview

We first applied the KL-KSOM to the extracted audio data, obtained the cluster map as result. Then HC was 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 audio events. After getting this classifier, classification was performed on all extracted audio events, then for every night's audio 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.

Table 4.1: 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

4.2.2 Experimental setting

The audio recording setting in this experiment is same as previous chapter. Subjects were asked to fulfill a self-rating questionnaire after waking up during the experiment. The questions regarding sleep quality are showed in Table 4.1. Question 1 to 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 [56]. Based on this founding, the audio 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 audio recordings from 36 different subjects with 18 in good quality and 18 in poor quality. The age range of these subjects is 20 to 29, and gender distribution was balanced.

4.2.3 Event extraction

Based on the burst extraction method, from 36 recordings, we obtained a total of 39105 audio 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 audio 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 audio data.

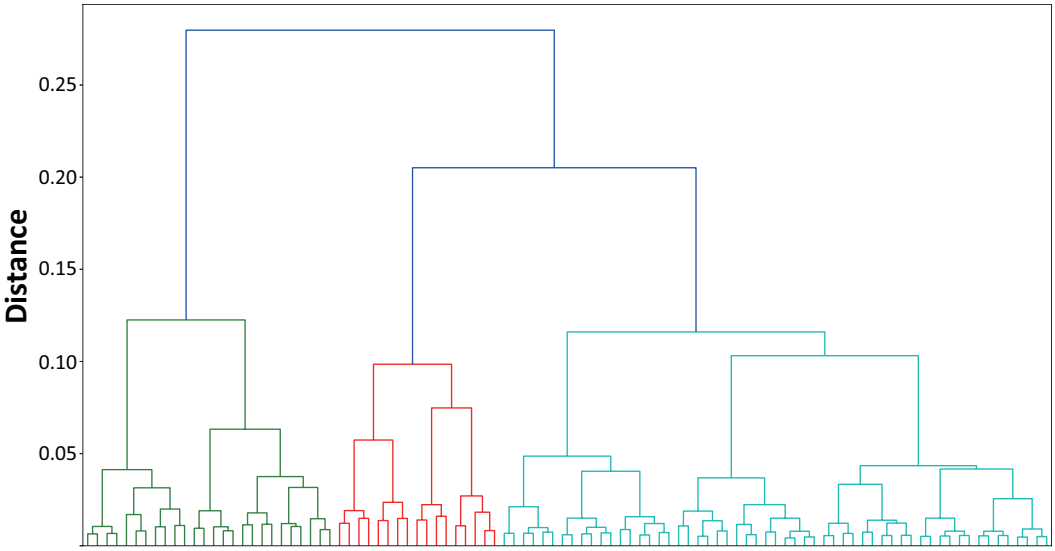

Figure 4.1: Dendrogram by HC

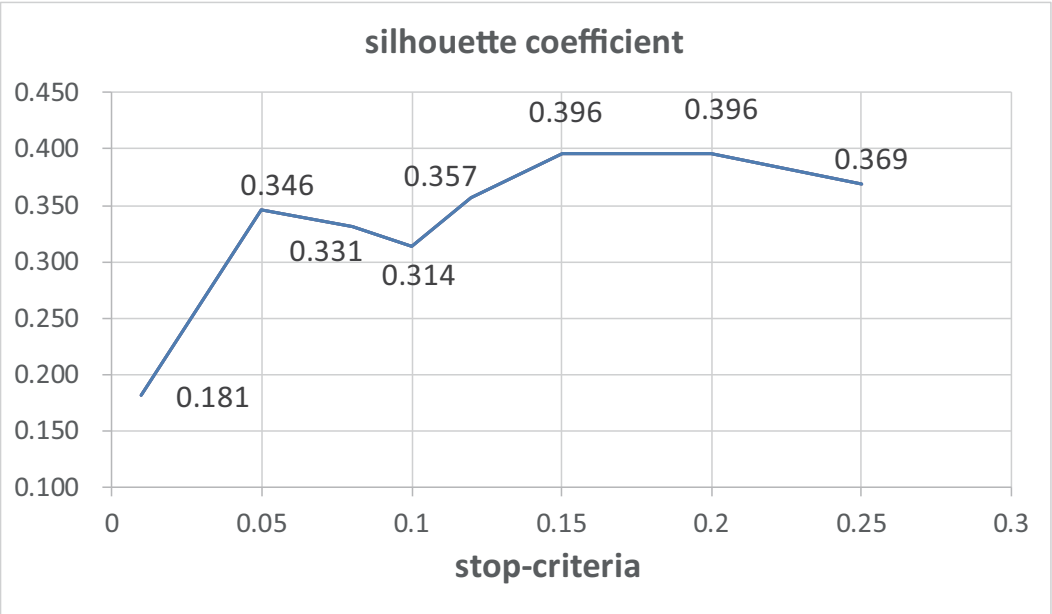

Figure 4.2: Silhouette coefficient on different stop-criteria

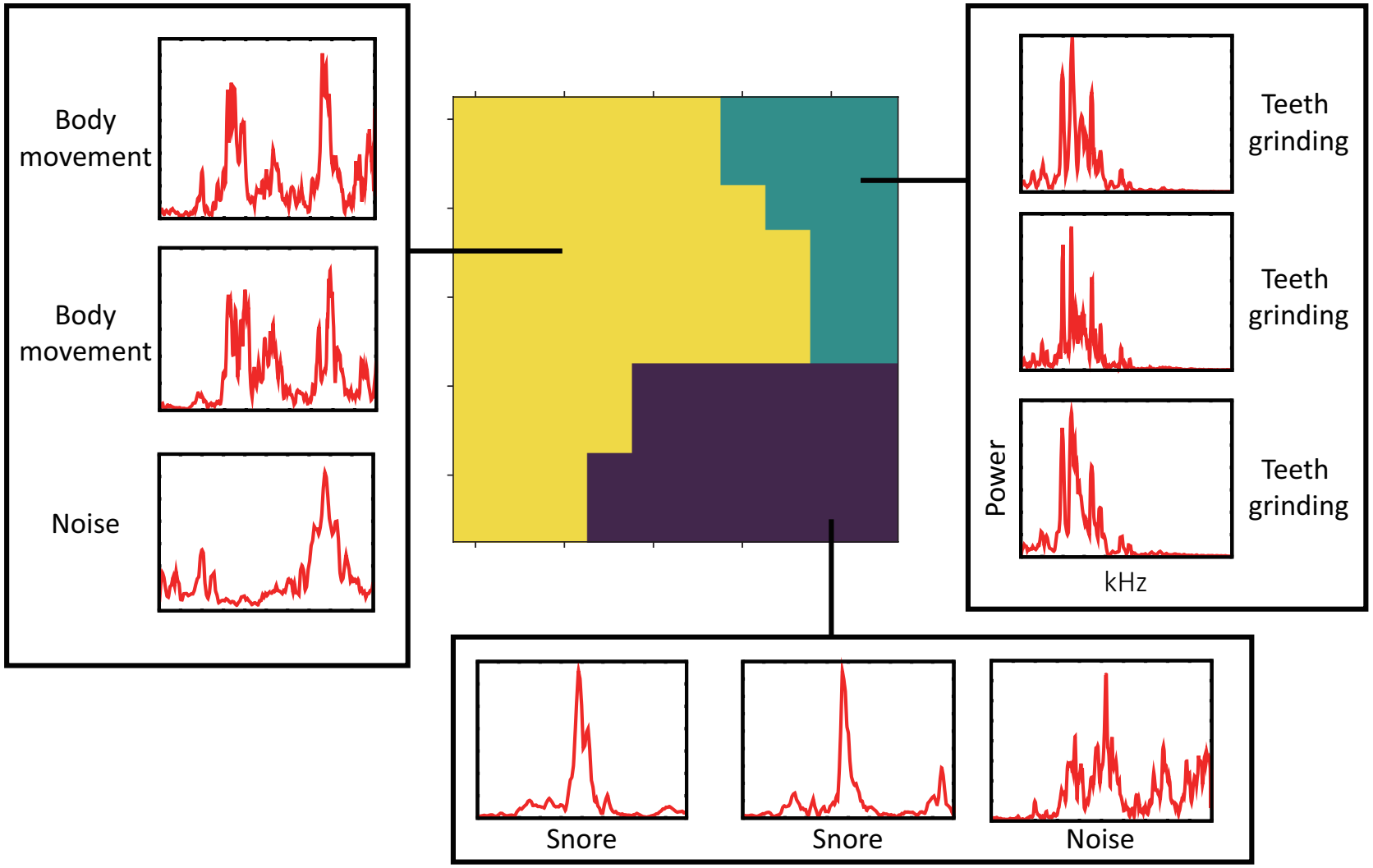

Figure 4.3: Major clusters on KL-KSOM cluster map with frequency spectrum of event examples from each cluster

5 hidden states models | 3 hidden states models

Good

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

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

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

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

(a) HMMs trained on sound events

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

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

(b) HMMs trained on sleep stage sequences

Figure 4.4: Transition probability matrices of HMMs

4.2.4 Audio 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. (4.1). Then, we applied agglomerative HC algorithm with Ward's criterion, the dendrogram is shown as Fig. 4.1. The stop-criteria for agglomerative HC was determined through silhouette, the silhouette coefficient on different stop-criteria is showed in Fig. 4.2. In this experiment, 0.15 was selected as the stop-criteria, thus cells was divided into 3 clusters as shown in Fig. 4.3.

We determined BMU for extracted audio events through Eq. (4.3), each audio event was then assigned to one of the 3 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 audio recordings, and found similar events were mainly categorized into same cluster. Therefore, all the events with time stamps extracted from one same audio recording formed a categorized data sequence to indicate that subject's sleep pattern.

Table 4.2: 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

4.2.5 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 con-

taining 18 sequences. We trained multinomial HMMs with $2, 3, 4$, 5 and 6 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 4.2 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 $2, 3, 4$ and 5 hidden states HMMs respectively, and as shown in Table 4.3, we found out data with good sleep quality got best accuracy on 5 hidden states models and poor ones on 3 hidden states models. 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 audio event are shown as Fig. 4.4(a) and sleep stage as Fig. 4.4(b). 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 audio events. However, we found there is no significant difference on sleep stage sequence HMMs between good and poor sleep quality (Fig. 4.4(b)). On the contrary, the HMMs of audio events from different sleep quality level have obvious difference (Fig. 4.4(a)). This evidence is interesting that sleep stage sequence is useless for assessing sleep quality.

About the relationship between sleep stage transitions and sleep quality, we found some discussion in [34] and [33]. In [33], the author compared the transition probabilities between healthy subjects and others with chronic fatigue syndrome (CFS), both the global and normed relative frequencies in transitions between REM and non-REM sleep are greater in healthy controls than in patients with CFS. Similar discussion can be

Table 4.3: 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

found in [34], too. However considering the probabilities that stage remain unchanged is mostly 90% in sleep stage sequence, the transition probability difference between REM and non-REM will be unconspicuous in HMM transition probability matrices. But these works provided a new idea, although the overall HMM model of sleep stage sequence is ineffective in sleep quality prediction, applying the transition probabilities between different sleep stages may get a different result.

About the difference between sound event and sleep stage sequence, we found that variety of sleep stages among subjects are larger than that of good/poor sleep, therefore if we learn the model on whole subjects it is quite hard to distinguish good/poor sleep. Also sound event has variety on subjects, even having less sound does not mean more deep sleep, but there is clear tendency that good sleep has less sound in every subject. Note that our experiment revealed that just event count features cannot classify good/poor sleep correctly, therefore transition probability on sound event is an importance factor.

By comparing good and poor models of 5 hidden state HMM on audio events (Fig. 4.4(a)), 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. 4.4(b)). In contrast, transition probabilities in the poor model on audio event are varied, which implies poor sleep do not have specific sleep pattern related to sounds.

Chapter 5

Conclusion

5.1 Summary

In this research, to visualize the sleep pattern, we applied FFT on extracted audio clips of events to obtain the frequency spectrum as input vectors, and applied various SOM algorithms to obtain cluster maps. We calculated the Euclidean distance between the frequency spectra and MFCC as the similarity measure between audio events in standard SOMs. Euclidean distance applied to a conventional SOM treats each discrete point as an independent variable; thus, we introduced the KL kernel as a similarity measure to capture the distribution structure of a frequency spectrum. For comparison, we also used a RBF kernel and a polynomial kernel. The experiment results show that the KL kernel outperformed the RBF and polynomial kernels. In addition, we found that the KL-SOM outperformed the standard SOM. To visualize the transition of cluster dynamics, we introduced the Sb-SOM, which introduces a SWF. By converting the spatiotemporal neighborhood into the topological neighborhood using a neighborhood function, the Sb-SOM can visualize the transition of cluster dynamics. Based on the property of kernel SOM, we introduced the KL kernel into the Sb-SOM and proposed the Sb-KSOM. The Sb-KSOM algorithm, which combines the advantages of a kernel SOM and an Sb-SOM, produces a cluster map that reflects the distribution and change of sleep-related events during the sleep period. To evaluate clustering performance, we calculated the wPF as the validity measure of each cluster map.

Since a personal sleep pattern is directly modeled via sleep-related audio events, the

proposed method does not require sleep stage estimation; however, the most accepted clinical research methods involve sleep stage. Therefore, to validate the proposed method, we performed a comparative interpretation between the obtained cluster maps generated by the Sb-KSOM and sleep stage sequences scored by medical specialists based on PSG data. The interpretation revealed that cluster distribution changes synchronously as sleep stages transition. Thus, similar to sleep stage sequences, discovering sleep patterns using cluster maps generated by an Sb-KSOM is feasible. To reveal the correlations between audio events and sleep stages, we calculated the conditional probabilities of audio events for a given sleep stage. The experimental results indicate that sleep-related audio events are related to sleep stages. These results empirically warrant the next topic of this research which is prediting personal sleep quality using audio data.

To build a model to predit sleep quality, we applied HC on the cluster map from KL-KSOM, 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 BMU on the cluster map. Therefore this divided cluster map can be used as a virtual classifier for input audio event, which we called virtual classifier. The output from this virtual classifier is necessary and sufficient to form a categorized data sequence for the following HMM modelling. After we got these categorized audio events sequences which represents sleep pattern, the HMMs of good and poor sleep quality were trained respectively. The likelihoods between an input audio event sequence and HMMs are calculated as input vectors, then several classification methods are applied. The results revealed this novel approach of sleep quality assessment is feasible.

5.2 Contributions

The major contribution of this work is the design and implementation of a methodology that visualizes the personal sleep pattern and predicts sleep quality based on audio data.

Current sleep study systems and devices are either too complex and expensive or lack of medical reliability, our method solved these problems simultaneously.

First, since we applied only audio data, any device with a microphone, including smartphones, recording pens, and personal computers can be used as the recording device, user doesn't need to purchase any additional device.

Second, not only PSG, but also almost all the sleep study methods all relied on devices that invasive to users, which means that users have to wear an additional device or place a device on their bed during sleep, on the contrary, our method is completely non-invasive.

Last but not least, our method is scientifically validated. We collaborated with medical experts in this study, the visualization results are consistent with medical evidence obtained using PSG, and in the sleep quality study, a questionnaire was designed by medical experts to evaluate the subjective sleep quality of experiment subjects, which made our training data with sufficient reliability.

5.3 Future issues

More work needs to be done in this study. First, the age range of the subjects is not general since all of subjects are university students. With the age increasing, the sleep state may change [9], more snore and body movement will happen on middle-aged than on young people, hence the sound event related model that generated from young subjects' data will possibly not work well on other age group. However, with the scope of data collection enlarging, this problem will be solved.

Also, in the events sequence generation, sometimes there is a long time interval between two audio events, it usually happened on quiet subjects, in the future work, researcher can 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.

Furthermore, currently we are still focus on single user application, for multi-user scenarios, also known as "Cocktail Party Problem" [28] for audio 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 audio sources based on aspect and phase difference.

Finally, based on current achievement, more practical research can be done, for example: monitoring the transition of sleep stage based on audio data, making a linkage between sleep quality predition and sleep aid devices such as smart air conditioner. We believe future researchers can find more interesting directions in it.

Bibliography

[1] T. Åkerstedt, M. Billiard, M. Bonnet, G. Ficca, L. Garma, M. Mariotti, P. Salzarulo, and H. Schulz. Awakening from sleep. *Sleep medicine reviews*, 6(4):267–286, 2002.

[2] P. Andras. Kernel-kohonen networks. *International journal of neural systems*, 12(02):117–135, 2002.

[3] L. Barghout. Spatial-taxon information granules as used in iterative fuzzy-decision-making for image segmentation. In *Granular Computing and Decision-Making*, pages 285–318. Springer, 2015.

[4] J. Behar, A. Roebuck, J. S. Domingos, E. Gederi, and G. D. Clifford. A review of current sleep screening applications for smartphones. *Physiological measurement*, 34(7):R29–R46, 2013.

[5] A. Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik. Support vector clustering. *Journal of machine learning research*, 2(Dec):125–137, 2001.

[6] P. Berkhin. A survey of clustering data mining techniques. In *Grouping multidimensional data*, pages 25–71. Springer, 2006.

[7] R. B. Berry, R. Brooks, C. E. Gamaldo, S. M. Harding, C. Marcus, and B. Vaughn. The aasm manual for the scoring of sleep and associated events. *Rules, terminology and technical specifications, darien, illinois, American Academy of Sleep Medicine*, 2012.

[8] C. M. Bishop. *Pattern recognition and machine learning.* springer, 2006.

[9] M. H. Bonnet and L. C. Johnson. Relationship of arousal threshold to sleep stage distribution and subjective estimates of depth and quality of sleep. *Sleep*, 1(2):161–168, 1978.

[10] R. Boulet, B. Jouve, F. Rossi, and N. Villa. Batch kernel som and related laplacian methods for social network analysis. *Neurocomputing*, 71(7):1257–1273, 2008.

[11] D. J. Buysse, C. F. Reynolds, T. H. Monk, S. R. Berman, and D. J. Kupfer. The pittsburgh sleep quality index: a new instrument for psychiatric practice and research. *Psychiatry research*, 28(2):193–213, 1989.

[12] Z. Chen, M. Lin, F. Chen, N. D. Lane, G. Cardone, R. Wang, T. Li, Y. Chen, T. Choudhury, and A. T. Campbell. Unobtrusive sleep monitoring using smartphones. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*, pages 145–152. IEEE, 2013.

[13] E. K. Choe, J. A. Kientz, S. Halko, A. Fonville, D. Sakaguchi, and N. F. Watson. Opportunities for computing to support healthy sleep behavior. In *CHI'10 extended abstracts on human factors in computing systems*, pages 3661–3666. ACM, 2010.

[14] S. Chokroverty. *Sleep disorders medicine: basic science, technical considerations, and clinical aspects*. Butterworth-Heinemann, 2013.

[15] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[16] N. Cristianini and J. Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[17] R. Cuingnet, C. Rosso, M. Chupin, S. Lehéricy, D. Dormont, H. Benali, Y. Samson, and O. Colliot. Spatial regularization of svm for the detection of diffusion alterations associated with stroke outcome. *Medical image analysis*, 15(5):729–737, 2011.

[18] S. Davis and P. Mermelstein. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE transactions on acoustics, speech, and signal processing*, 28(4):357–366, 1980.

[19] D. L. Dickinson, J. Cazier, and T. Cech. A practical validation study of a commercial accelerometer using good and poor sleepers. *Health psychology open*, 3(2), 2016.

[20] I. Feinberg. Changes in sleep cycle patterns with age. *Journal of psychiatric research*, 10(3):283–306, 1974.

[21] Y. Freund and R. E. Schapire. A desicion-theoretic generalization of on-line learning and an application to boosting. In *European conference on computational learning theory*, pages 23–37. Springer, 1995.

[22] K. Fukui, S. Akasaki, K. Sato, J. Mizusaki, K. Moriyama, S. Kurihara, and M. Numao. Visualization of damage progress in solid oxide fuel cells. *Journal of environment and engineering*, 6(3):499–511, 2011.

[23] K. Fukui and M. Numao. Neighborhood-based smoothing of external cluster validity measures. In *Advances in knowledge discovery and data mining*, pages 354–365. Springer, 2012.

[24] K. Fukui, K. Saito, M. Kimura, and M. Numao. Sequence-based som: Visualizing transition of dynamic clusters. In *Proceedings of IEEE 8th international conference on computer and information technology (CIT 2008)*, pages 47–52. IEEE, 2008.

[25] H. Fukumura, S. Okada, and M. Makikawa. Estimation of sleep stage using svm from noncontact measurement of forehead and nasal skin temperature. *BME*, 50(1):131–137, 2012.

[26] W. Gu, Z. Yang, L. Shangguan, W. Sun, K. Jin, and Y. Liu. Intelligent sleep stage mining service with smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, pages 649–660. ACM, 2014.

[27] T. Hao, G. Xing, and G. Zhou. isleep: Unobtrusive sleep quality monitoring using smartphones. In *Proceedings of the 11th ACM conference on embedded networked sensor systems*, SenSys '13, pages 4:1–4:14. ACM, 2013.

[28] S. Haykin and Z. Chen. The cocktail party problem. *Neural computation*, 17(9):1875–1902, 2005.

[29] J. A. Hobson, E. F. Pace-Schott, and R. Stickgold. Dreaming and the brain: toward a cognitive neuroscience of conscious states. *Behavioral and brain sciences*, 23(06):793–842, 2000.

[30] C. C. Hoch, C. F. Reynolds, D. J. Kupfer, S. R. Berman, P. R. Houck, and J. A. Stack. Empirical note: self-report versus recorded sleep in healthy seniors. *Psychophysiology*, 24(3):293–299, 1987.

[31] T. Ishigaki and T. Higuchi. Dynamic spectrum classification by kernel classifiers with divergence-based kernels and its applications to acoustic signals. *International journal of knowledge engineering and soft data paradigms*, 1(2):173–192, 2009.

[32] T. Kato, Y. Masuda, A. Yoshida, and T. Morimoto. Masseter emg activity during sleep and sleep bruxism. *Archives italiennes de biologie*, 149(4):478–491, 2011.

[33] A. Kishi, Z. R. Struzik, B. H. Natelson, F. Togo, and Y. Yamamoto. Dynamics of sleep stage transitions in healthy humans and patients with chronic fatigue syndrome. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, 294(6):R1980–R1987, 2008.

[34] A. Kishi, H. Yasuda, T. Matsumoto, Y. Inami, J. Horiguchi, M. Tamaki, Z. R. Struzik, and Y. Yamamoto. Nrem sleep stage transitions control ultradian rem sleep rhythm. *Sleep*, 34(10):1423–1432, 2011.

[35] J. Kleinberg. Bursty and hierarchical structure in streams. *Data mining and knowledge discovery*, 7(4):373–397, 2003.

[36] T. Kohonen. The self-organizing map. *Neurocomputing*, 21(1):1–6, 1998.

[37] T. Kohonen, S. Kaski, K. Lagus, J. Salojärvi, J. Honkela, V. Paatero, and A. Saarela. Self organization of a massive document collection. *IEEE transactions on neural networks*, 11(3):574–585, 2000.

[38] G. Lavigne, P. Rompre, and J. Montplaisir. Sleep bruxism: validity of clinical research diagnostic criteria in a controlled polysomnographic study. *Journal of dental research*, 75(1):546–552, 1996.

[39] J. Li, A. Najmi, and R. M. Gray. Image classification by a two-dimensional hidden markov model. *IEEE transactions on signal processing*, 48(2):517–533, Feb 2000.

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

[41] V. Metsis, D. Kosmopoulos, V. Athitsos, and F. Makedon. Non-invasive analysis of sleep patterns via multimodal sensor input. *Personal and ubiquitous computing*, 18(1):19–26, 2014.

[42] T. Mollayeva, P. Thurairajah, K. Burton, S. Mollayeva, C. M. Shapiro, and A. Colantonio. 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 medicine reviews*, 25:52–73, 2016.

[43] T. Noh, Y. Serizawa, T. Kimura, K. Yamazaki, Y. Hayasaka, T. Itoh, S. Izumi, and T. Sasaki. The assessment of sleep stage utilizing body pressure fluctuation measured by water mat sensors. *Journal of advanced science*, 21(1&2):27–30, 2009.

[44] P. L. Parmeggiani. *Systemic Homeostasis and Poikilostasis in sleep: Is REM sleep a physiological paradox?* World Scientific, 2011.

[45] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[46] B. Riedel and K. Lichstein. Objective sleep measures and subjective sleep satisfaction: how do older adults with insomnia define a good night's sleep? *Psychology and aging*, 13(1):159–163, March 1998.

[47] L. Rokach and O. Maimon. Clustering methods. In *Data mining and knowledge discovery handbook*, pages 321–352. Springer, 2005.

[48] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20:53–65, 1987.

[49] K. Russo, B. Goparaju, and M. T. Bianchi. Consumer sleep monitors: is there a baby in the bathwater? *Nature and science of sleep*, 7:147–157, 2015.

[50] M. H. Silber, S. Ancoli-Israel, M. H. Bonnet, S. Chokroverty, M. M. Grigg-Damberger, M. Hirshkowitz, S. Kapen, S. A. Keenan, M. H. Kryger, T. Penzel, et al. The visual scoring of sleep in adults. *Journal of clinical sleep medicine*, 3(2):121–131, 2007.

[51] O. Simula and J. Kangas. Process monitoring and visualization using self-organizing maps. *Neural networks for chemical engineers*, 6:371–384, 1995.

[52] N. C. Smeeton. Early history of the kappa statistic. *Biometrics*, 41(3):795–795, 1985.

[53] A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.

[54] E. L. Sonnhammer, G. Von Heijne, A. Krogh, et al. A hidden markov model for predicting transmembrane helices in protein sequences. In *Ismb*, volume 6, pages 175–182, 1998.

[55] K. Spruyt, D. L. Molfese, and D. Gozal. Sleep duration, sleep regularity, body weight, and metabolic homeostasis in school-aged children. *Pediatrics*, 127(2):e345–e352, 2011.

[56] H. Tanaka and S. Shirakawa. Sleep health, lifestyle and mental health in the japanese elderly: ensuring sleep to promote a healthy brain and mind. *Journal of psychosomatic research*, 56(5):465–477, 2004.

[57] J. Vesanto and E. Alhoniemi. Clustering of the self-organizing map. *IEEE transactions on neural networks*, 11(3):586–600, 2000.

[58] L. Wei, Y. Lin, J. Wang, and Y. Ma. Time-frequency convolutional neural network for automatic sleep stage classification based on single-channel eeg. In *Proceedings of 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 88–95. IEEE, 2017.

[59] Wikipedia, the free encyclopedia. Hidden markov model, 2017. `https://commons.wikimedia.org/wiki/File:HiddenMarkovModel.svg` [Online; accessed Nov. 29, 2017].

[60] Wikipedia, the free encyclopedia. Support vector machine, 2017. `https://commons.wikimedia.org/wiki/File:Svm_max_sep_hyperplane_with_margin.png` [Online; accessed Nov. 29, 2017].

[61] H. Wu, T. Kato, T. Yamada, M. Numao, and K. Fukui. Sleep pattern discovery via visualizing cluster dynamics of sound data. In *International conference on industrial, engineering and other applications of applied intelligent systems*, pages 460–471. Springer, 2016.

[62] R. Xu and D. C. Wunsch. Cluster validity. *Clustering*, pages 263–278, 2008.

[63] S. C. Yudofsky and R. E. Hales. *Essentials of neuropsychiatry and clinical neurosciences.* American Psychiatric Publishing, Inc., 2004.

[64] Y. Zhang, Y. Chen, L. Hu, X. Jiang, and J. Shen. An effective deep learning approach for unobtrusive sleep stage detection using microphone sensor. In *Proceedings of 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, pages 37–44. IEEE, 2017.

[65] M. Zhao, S. Yue, D. Katabi, T. S. Jaakkola, and M. T. Bianchi. Learning sleep stages from radio signals: A conditional adversarial architecture. In *Proceedings of the 34th International Conference on Machine Learning*, pages 4100–4109, 2017.