Diffuse to fuse EEG spectra – intrinsic geometry of sleep dynamics for classification

Gi-Ren Liu

joint work with
Hautieng Wu (Duke University)
Yu-Lun Lo (Chang Gung Memorial Hospital)
Yuan-Chung Sheu (National Chiao Tung University)

Department of Mathematics
National Cheng Kung University

June 27, 2018

Outline

- Background
- Related Works
- Short-Time Fourier Transform
- Synchrosqueezed Transform
- Diffusion Map
- Inter-Individual Sleep Assessment
- Performance Evaluation

Background and Motivation

Each sleep cycle broadly consists of sleep stages: **Awake**, **Rapid Eye Movement (REM)**, and **Non-REM**.
The non-REM stage can be further classified into **N1**, **N2** (shallow sleep) and **N3** (deep sleep).

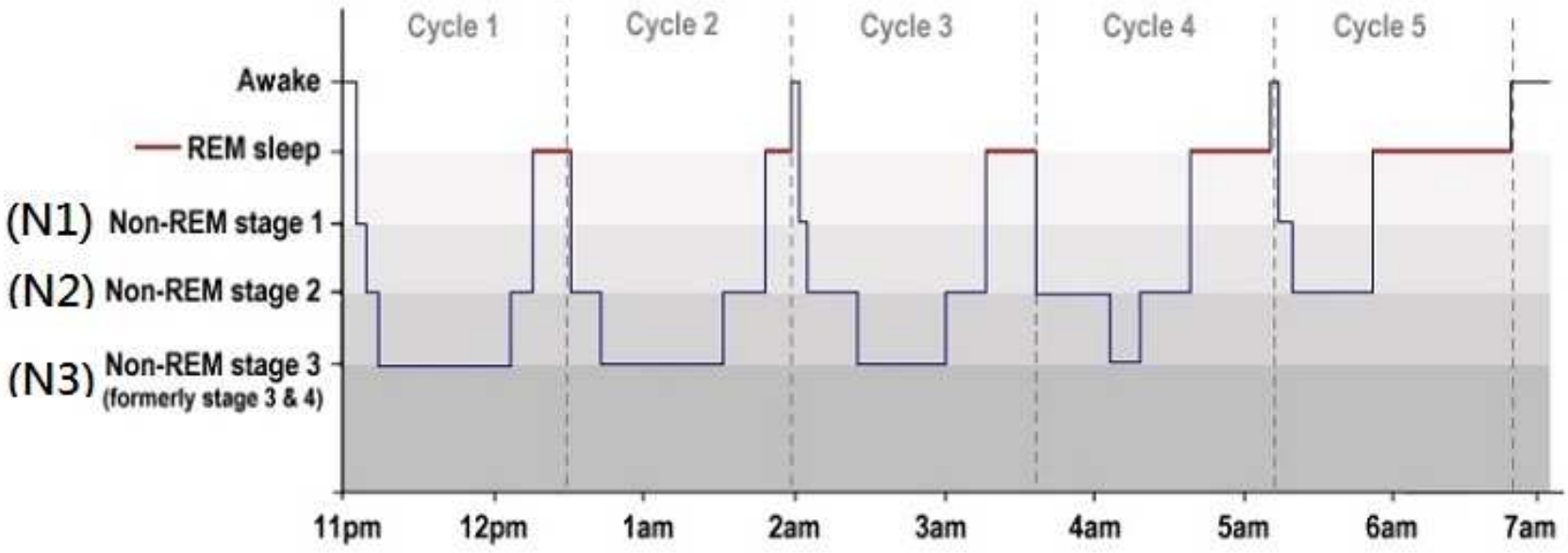


Figure: An 8-hour sleep of a healthy subject

Background

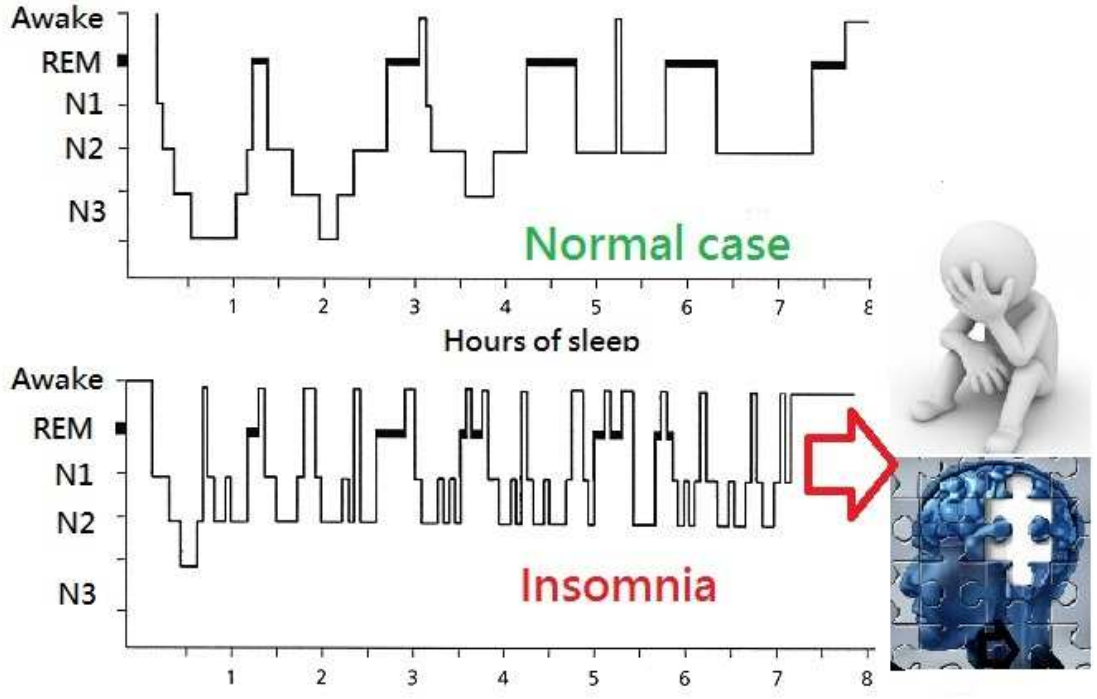
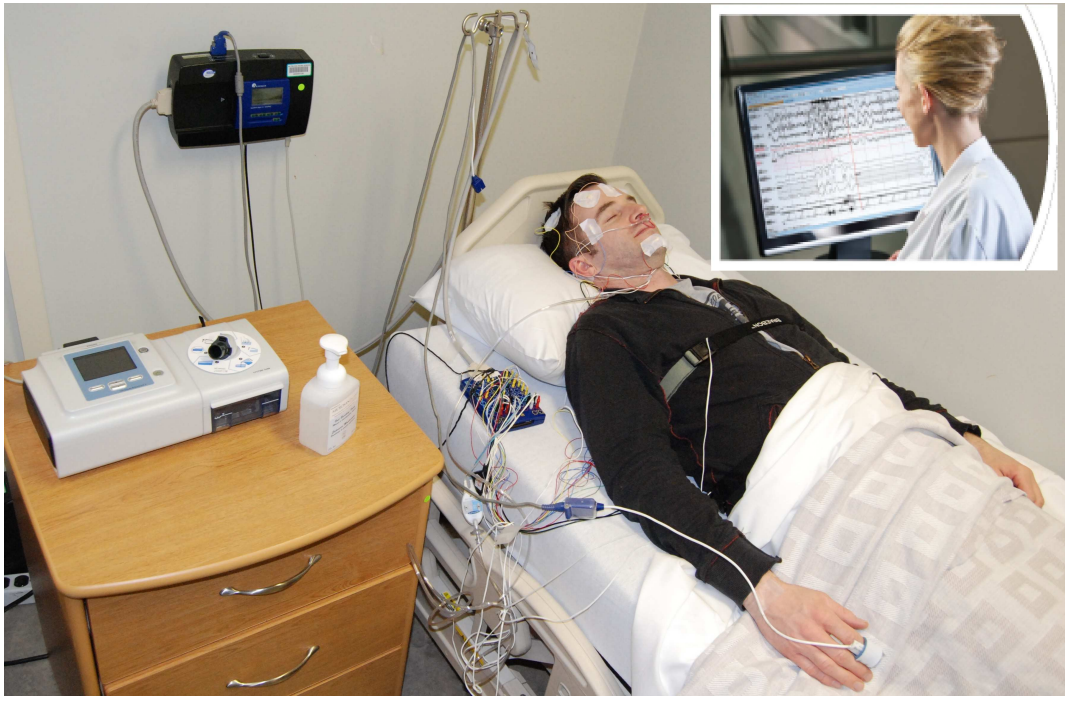


Figure: Difference between normal sleep and insomnia

It is hard for patients to be aware of the shortage of stages REM and N3, which may be linked to depression, memory loss, and apnea.

Mahowald, M. W. and Schenck, C. H. (2005). Insights from studying human sleep disorders. *Nature*, **437**, 1279-1285.

Visual EEG scoring



Clinically, the sleep stage within a 30s epoch is mainly determined by the pattern of EEG, EOG and EMG signals based on a standard named AASM.

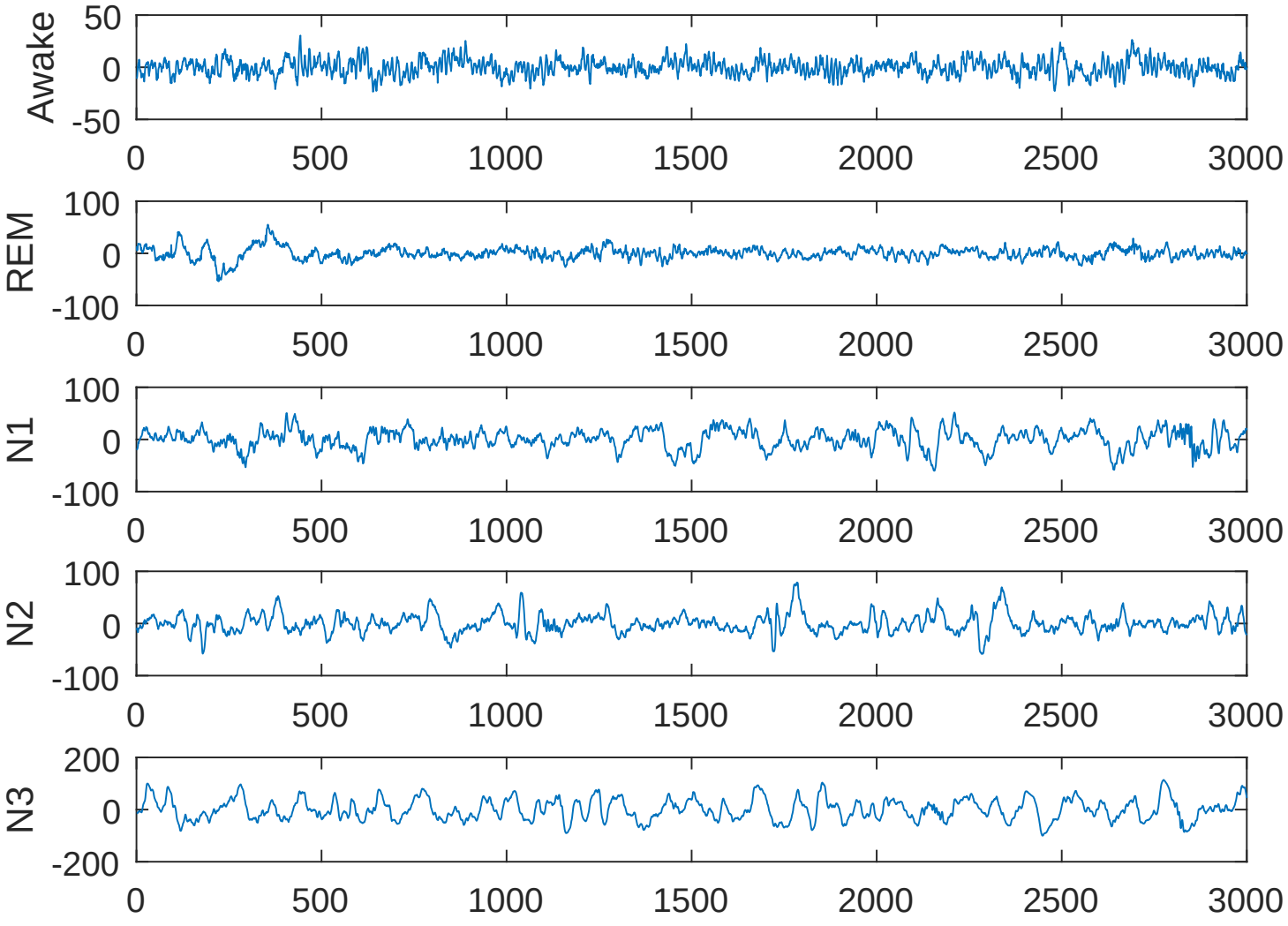


Figure: 30s raw EEG data for each stage

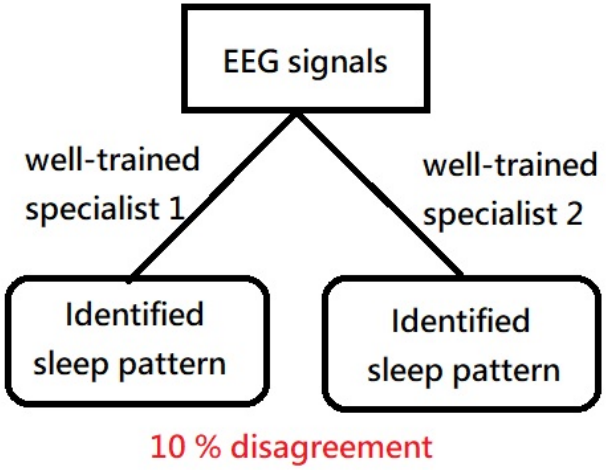


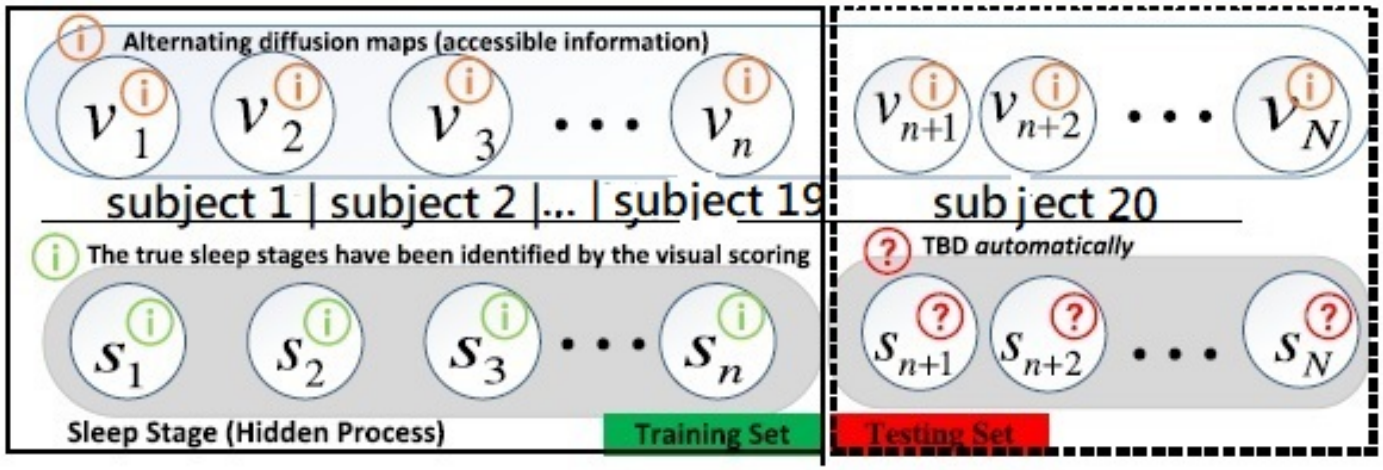
Figure: Drawback of the visual scoring method (There exists some disagreement on the sleep patterns scored by different specialists.)

Goal: designing an automatic sleep scoring algorithm

Our algorithm consists of the following three steps:

- Feature extraction by the synchroqueezed short time Fourier transform
- Feature clustering by diffusion maps
- Classification by hidden Markov model

Question: How to recover the sleep-stage pattern for the 20th subject?



THE SLEEP-EDF DATABASE [EXPANDED]

This database is described in

B Kemp, AH Zwinderman, B Tuk, HAC Kamphuisen, JJL Oberyé. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. *IEEE-BME* **47**(9):1185-1194 (2000).

Please cite this publication when referencing this material, and also include the standard citation for PhysioNet:

Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* **101**(23):e215-e220 [Circulation Electronic Pages; http://circ.ahajournals.org/cgi/content/full/101/23/e215]; 2000 (June 13).

On this p
- Sleep
and Hy
Europe
Forma
- Data
stud
effe
in h
subj
- Data
stud
tem
effe
- Soft
view
EDF
- Refe

SLEEP RECORDINGS AND HYPNOGRAMS IN EUROPEAN DATA FORMAT (EDF)

Bob Kemp, Sleep Centre, MCH-Westeinde Hospital, Den Haag, The Netherlands
bk at hsr dot nl

This collection of 61 polysomnograms (PSGs) with accompanying hypnograms (expert annotations of sleep stages) comes from two studies (briefly described below, and in detail in [1,2]). A small subset of this dataset was previously

Figure: Baseline Database

Feature extraction by STFT

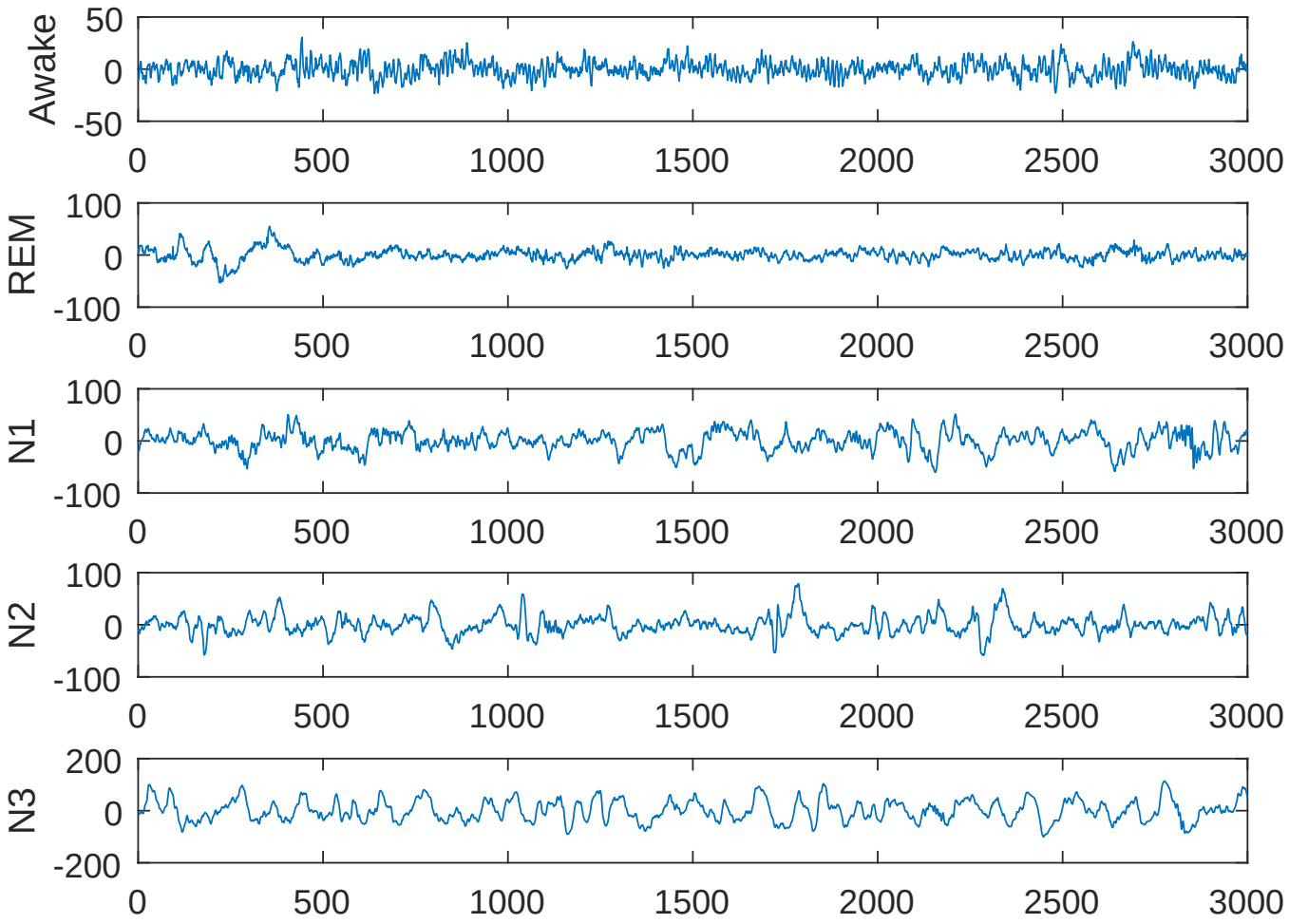


Figure: 30s raw EEG data for each stage

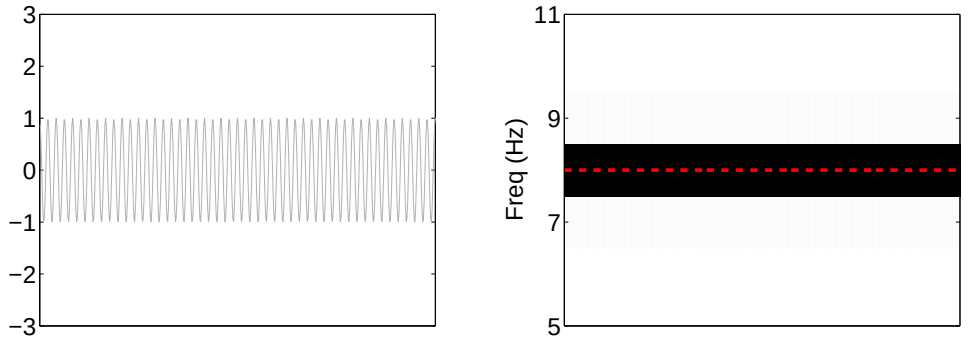
- $\mathbf{x} \in \mathbb{R}^N$: an EEG signal with sampling rate S Hz.
  If the database consists of 20 subjects and the length of sleep for each subject is 6 hours, then $N = 20 \times S \times 3600 \times 6$.
- Since we would like to quantize the EEG patterns per 30 seconds, $\mathbf{x}$ is split into $J := N/W$ ($W = 30S$) almost disjoint frames
$$\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(J)}.$$
- Each frame $\mathbf{x}^{(j)}$ has $30S + 1$ data points, which are expressed as
$$\mathbf{x}^{(j)} = (x^{(j)}_{-15S}, ..., x^{(j)}_0, ..., x^{(j)}_{15S}).$$
- The STFT of $\mathbf{x}$ is defined by
$$\mathbf{X}_g(k,j) := \sum_{m=-15S}^{15S} e^{-i2\pi \frac{k}{K} m} x^{(j)}_m \frac{1}{H} g(\frac{m}{H}),\ k \in \{0, , 1, ..., K-1\},$$
  where $K \in \mathbb{N}$ is a constant relevant to the frequency resolution and the window function $g(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$.

For the special case $\mathbf{x}(t) = \exp(i2\pi\omega_0 t)$, the continuous version of STFT

$$\mathcal{X}_g(\omega, t) := \int_{\mathbb{R}} e^{-i2\pi\omega(s-t)} \mathbf{x}(s) \frac{1}{H} g\left(\frac{s-t}{H}\right) ds$$

can be rewritten as

$$\mathcal{X}_g(\omega, t) = e^{i2\pi\omega_0 t} \exp\left(-2\pi^2(\omega - \omega_0)^2 H^2\right).$$



(a) test signal (b) STFT

Figure: Spectrogram of $\mathbf{x}(t) = e^{i2\pi\omega_0 t}$ with $\omega_0 = 8$. The dashed line in (b) is the instantaneous frequency curve of $\mathbf{x}$. The instantaneous frequency curve $\omega = 8$ Hz is blurred by the Gaussian kernel.

- For the special case,

$$\omega_0 = \frac{1}{2\pi}\mathrm{Im}\left(\frac{\partial}{\partial t}\ln \mathcal{X}_g(\omega,t)\right).$$

-

$$\begin{aligned}\mathcal{X}_g(\omega,t)\frac{\partial \ln \mathcal{X}_g}{\partial t}(\omega,t) =& i2\pi\omega\int_{\mathbb{R}} e^{-i2\pi\omega(s-t)}\mathbf{x}(s)\frac{1}{H}g(\frac{s-t}{H})ds \\ &- \int_{\mathbb{R}} e^{-i2\pi\omega(s-t)}\mathbf{x}(s)\frac{1}{H^2}g'(\frac{s-t}{H})ds \\ =& i2\pi\omega\mathcal{X}_g(\omega,t) - H^{-1}\mathcal{X}_{g'}(\omega,t). \qquad (1)\end{aligned}$$

Substituting

$$\frac{\partial \ln \mathcal{X}_g}{\partial t}(\omega, t) = i2\pi\omega - H^{-1} \frac{\mathcal{X}_{g'}(\omega, t)}{\mathcal{X}_g(\omega, t)}$$

into

$$\omega_0 = \frac{1}{2\pi} \mathrm{Im} \left(\frac{\partial}{\partial t} \ln \mathcal{X}_g(\omega, t) \right)$$

yields

$$\omega_0 = \omega - \mathrm{Im} \left(\frac{1}{2\pi H} \frac{\mathcal{X}_{g'}(\omega, t)}{\mathcal{X}_g(\omega, t)} \right),$$

which means that the energy near ω should be shifted to $\omega - \mathrm{Im} \left(\frac{1}{2\pi H} \frac{\mathcal{X}_{g'}(\omega,t)}{\mathcal{X}_g(\omega,t)} \right)$ for each $t > 0$.

Synchrosqueezed transform (SST)

◇ For the continuous-time case, the new spectral energy distribution is

$$E_{\text{new}}(\hat{\omega}, t) = \int_{\Lambda(\hat{\omega})} |\mathcal{X}_g(\omega, t)|^2 d\omega, \ \hat{\omega} \in [0, 1],$$

where

$$\Lambda(\hat{\omega}) = \left\{ \omega \in [0, 1] \mid \omega - \text{Im} \left(\frac{1}{2\pi H} \frac{\mathcal{X}_{g'}(\omega, t)}{\mathcal{X}_g(\omega, t)} \right) = \hat{\omega} \right\}.$$

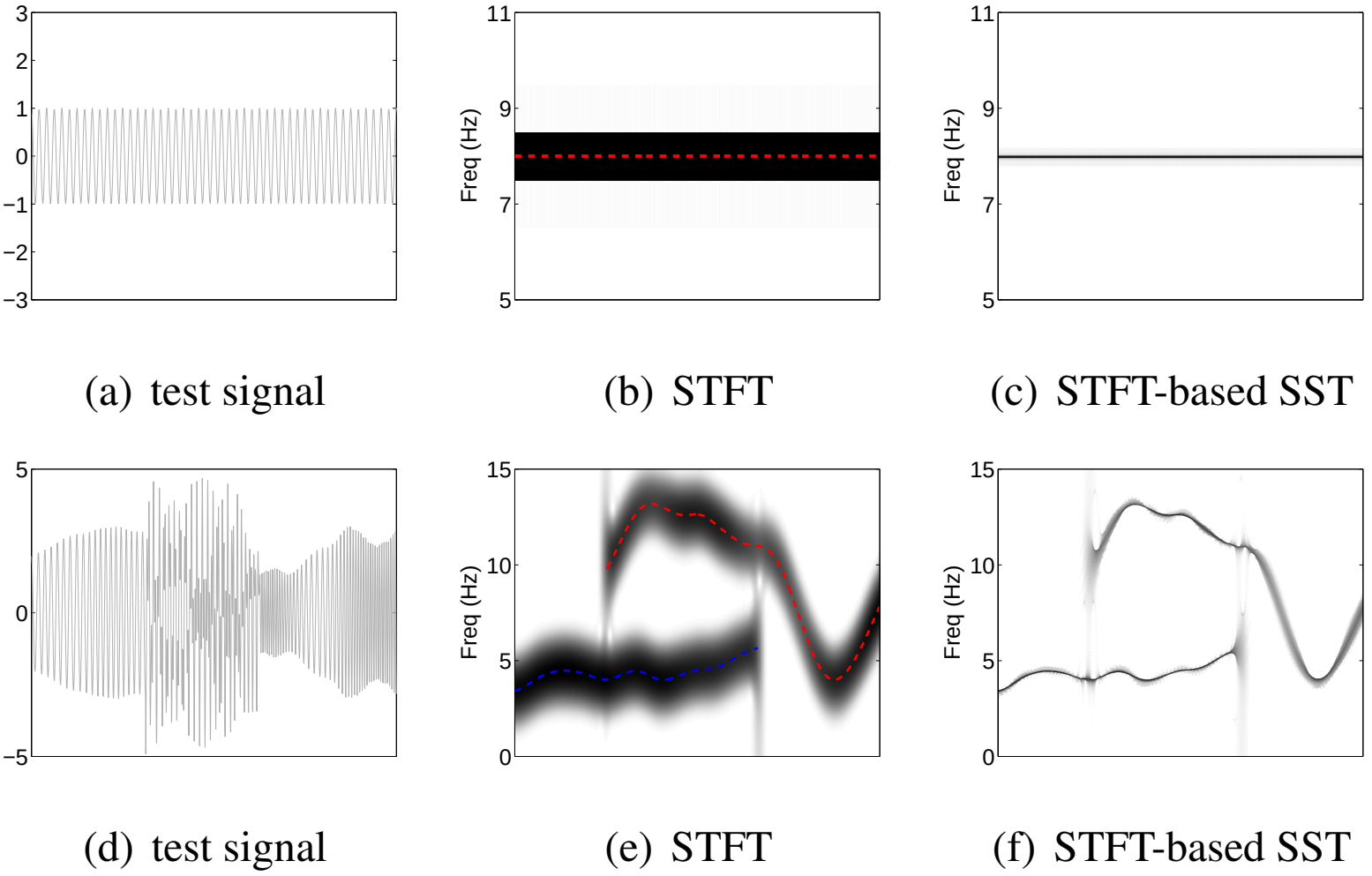
◇ For the discrete-time case, the new energy distribution for the jth frame $\mathbf{x}^{(j)}$ is

$$E_{\text{new}}(\hat{k}, j) = \sum_{k \in \Lambda(\hat{k})} |\mathbf{X}_g(k, j)|^2$$

where

$$\Lambda(\hat{k}) = \left\{ k \in \{0, \ldots, K-1\} \mid k - \text{Im} \left(\frac{K}{2\pi H} \frac{\mathbf{X}_{g'}(k, j)}{\mathbf{X}_g(k, j)} \right) \in [\hat{k} - \frac{1}{2}, \hat{k} + \frac{1}{2}) \right\}.$$

Effects of energy reallocation



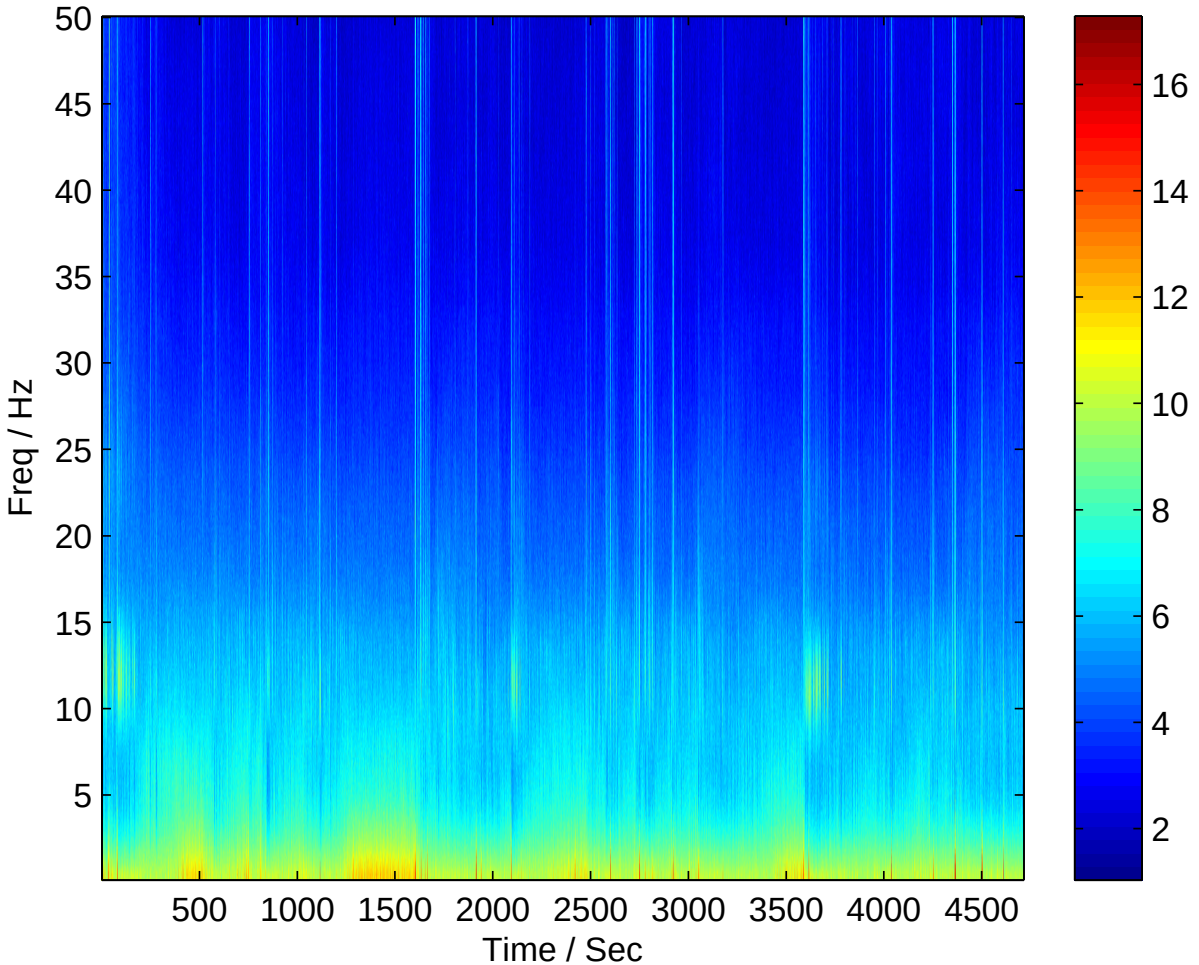
(a) test signal (b) STFT (c) STFT-based SST

(d) test signal (e) STFT (f) STFT-based SST

Figure: Spectrogram of (a) the test signal obtained by (b) STFT and (c) SST. The dashed lines in (b) are the correct instantaneous frequency curves.

Spectrogram of an EEG signal

Due to the presence of noise and the complication of physiological signals, the patterns in the spectrogram are not easy to be observed only by visual observation.



According to [2], the depth of sleep can be partially inferred from the main IF of the electrical activity of brain.

wave name	frequency band	possible occurrence time
Delta	0.5-3 Hz	Deep sleep (N3)
Theta	4-7 Hz	REM or Dreaming period
Alpha	7.5-12.5 Hz	Relaxation with closed eyes (N1 or N2)
Beta	14-31 Hz	Wakefulness

The EEG pattern for the jth epoch is quantified by $\{y^1(j)\ y^2(j)\ \cdots\}$:

- $y^1(j) = \sum_{0.5<Sk/K<3} E_{\text{new}}(k,j)$
- $y^2(j) = \sum_{4<Sk/K<7} E_{\text{new}}(k,j)$
- $y^3(j) = \sum_{7.5<Sk/K<12.5} E_{\text{new}}(k,j)$
- $y^4(j) = \sum_{14<Sk/K<31} E_{\text{new}}(k,j)$

---

[2] Buzsaki, Gyorgy (2006). Rhythms of the Brain. New York: Oxford University Press.

Feature Clustering

Given an EEG signal, let

$$\mathbf{y}(j) := [y^1(j)\ y^2(j)\ \cdots\ y^m(j)]^{\mathrm{T}},\ \ j \in \{1, 2, ..., N\},$$

be the feature corresponding to the jth time slot. We suppose that the feature series $\mathbf{y}(1), \mathbf{y}(2), ...$ is controlled by an underlying factors $\boldsymbol{\theta}(1), \boldsymbol{\theta}(2), ...$

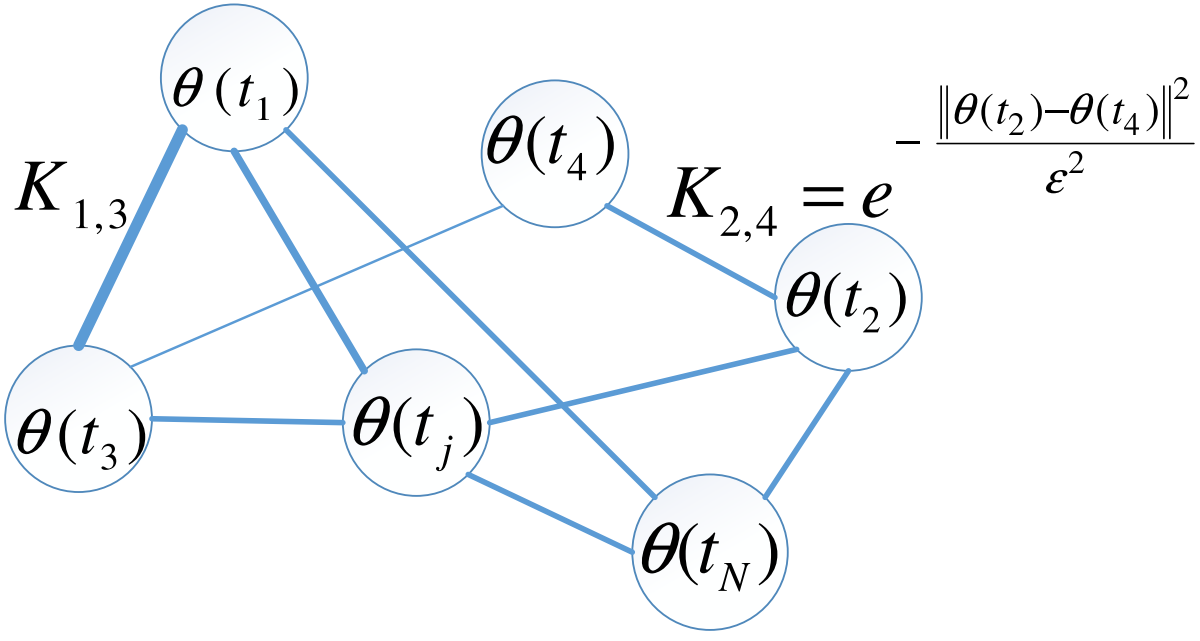
Weighted Euclidean distance

$$\mathrm{dist}(\theta(j), \theta(k)) \triangleq \sqrt{(\mathbf{y}(j) - \mathbf{y}(k))^{\mathrm{T}}(\frac{\mathbf{C}_j^{-1} + \mathbf{C}_k^{-1}}{2})(\mathbf{y}(j) - \mathbf{y}(k))}, \quad (2)$$

where $\mathbf{C}_j$ (resp. $\mathbf{C}_k$) is the sample covariance matrix of the features within the K-nearest neighbors of $\mathbf{y}(j)$ (resp. $\mathbf{y}(k)$). K is a predetermined integer.

Diffusion Distance (proposed by Coiman & Lafan in 2005)

In order to evaluate the affinity between different features (or the underlying factors $\{\boldsymbol{\theta}(j)\}_{j=1}^{N} \subset \mathbb{R}^d$), we consider an edge-weighted graph $\mathbf{G}$. (notation exchange: $\theta(j) \Leftrightarrow \theta(t_j)$)



$\mathbf{K} = [K_{i,j}]$ is called the similarity (or affinity) matrix.

Singer, A. and Coifman, R. R. Non-linear independent component analysis with diffusion maps. *Applied and Computational Harmonic Analysis*

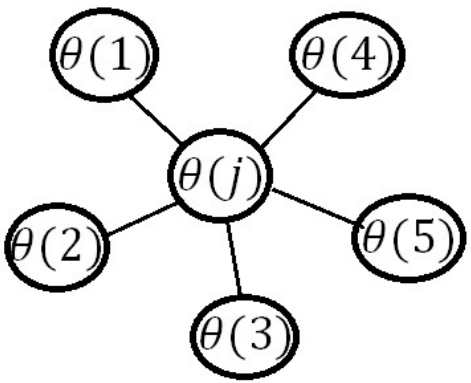
On the graph $\mathbf{G}$, we construct a Markov chain with transition matrix $\mathbf{P}$ by row normalizing the similarity matrix $\mathbf{K}$:

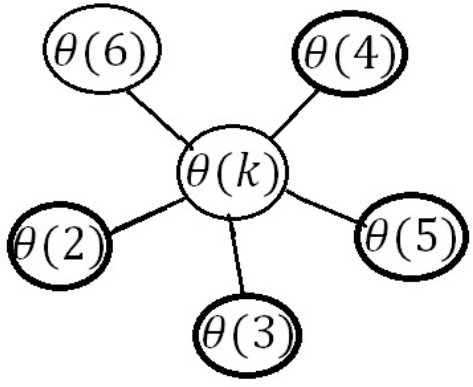
$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{K},\ \mathbf{D} = \mathrm{diag}(\sum_{j=1}^{N} K_{1,j}, \ldots, \sum_{j=1}^{N} K_{N,j}).$$

Its jth row

$$\mathbf{P}(j,:) = \left[K_{j,1}\ K_{j,2}\ K_{j,3} \cdots\right] \left[\sum_{\ell} K_{j,\ell}\right]^{-1}$$

is the probability distribution of one-step random walk starting from vertex $\theta(j)$.





- Two underlying factors $\theta(j)$ and $\theta(k)$ are expected to be very similar if their neighborhoods greatly overlap.
- The strength of overlapping can be evaluated by the difference between $P(j,:)$ and $P(k,:)$.

Definition of Diffusion Distance

The diffusion distance between $\theta(j)$ and $\theta(k)$ is defined by comparing the difference between two probability distributions:

$$D_h(\theta(j),\theta(k)) = \sum_{\ell=1}^{N} |P^h_{j,\ell} - P^h_{k,\ell}|^2 / D_\ell,$$

where $P^h_{j,\ell}$ (resp. $P^h_{k,\ell}$) is the probability that a particle, whose initial position is at vertex $\theta(j)$ (resp. $\theta(k)$), appears at vertex $\theta(\ell)$ after **P**-transitioning h steps.

Diffusion map

Consider the diagonalization of the transition matrix $\mathbf{P}$

$$\mathbf{P} = \begin{bmatrix} u_1 & u_2 & \cdots & u_N \end{bmatrix} \mathrm{Diag}(\lambda_1, \lambda_2, ..., \lambda_N) \begin{bmatrix} u_1 & u_2 & \cdots & u_N \end{bmatrix}^{-1}$$

and define a mapping $\Phi_h : \mathbb{R}^d \to \mathbb{R}^{N-1}$

$$\theta(j) \xrightarrow{\Phi_h} \begin{bmatrix} \lambda_2^h u_2(j) \\ \lambda_3^h u_3(j) \\ \lambda_4^h u_4(j) \\ \vdots \\ \lambda_N^h u_N(j) \end{bmatrix}. \tag{3}$$

$$D_h(\theta(j), \theta(k)) = \|\Phi_h(\theta(j)) - \Phi_h(\theta(k))\|_{\mathbb{R}^{N-1}}$$

In the following discussion, the subscript h of Φ is ignored.

Open access database PhysioNet

- This database contains 20 subjects, including 10 males and 10 females.
- Their ages range from 25-34 years old.
- Two channels of EEG signals for each subject are recorded. The first one is named **O1A2**, while the second one is named **O2A1**.

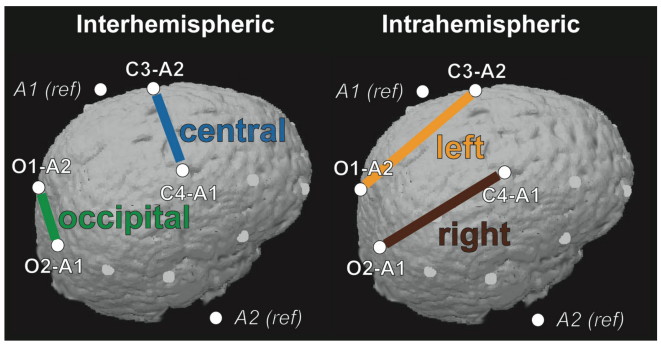


Figure: reference: [3]

- The sleep stage for each time slot has been assigned by the well trained specialists (the traditional visual scoring method).

The data are downloaded from *https* : *//www.physionet.org/pn4/sleep − edfx/*

[3] Salome Kurth, *et al.* Development of Brain EEG Connectivity across Early Childhood: Does Sleep Play a Role? *Brain Science* (2013)

Truncated diffusion map

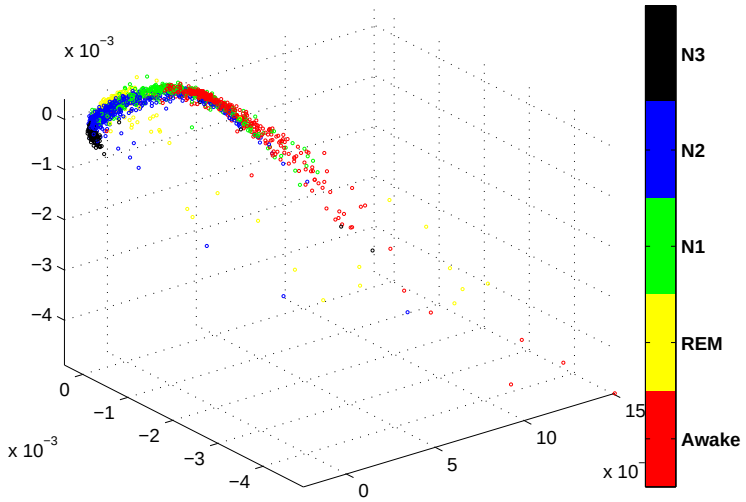
- θ_{O1A2}, θ_{O2A1}: the underlying processes controlling the observable features of O1A2 and O2A1, respectively.
- $\mathbf{P}_{\mathrm{O1A2}}$, $\mathbf{P}_{\mathrm{O2A1}}$: the transition matrices generated by the observable features of O1A2 and O2A1, respectively.
- Φ_{O1A2}, Φ_{O2A1}: the diffusion maps defined by the eigenvector decomposition of $\mathbf{P}_{\mathrm{O1A2}}$ and $\mathbf{P}_{\mathrm{O2A1}}$, respectively.

For $j = 1, 2, ..., N$,

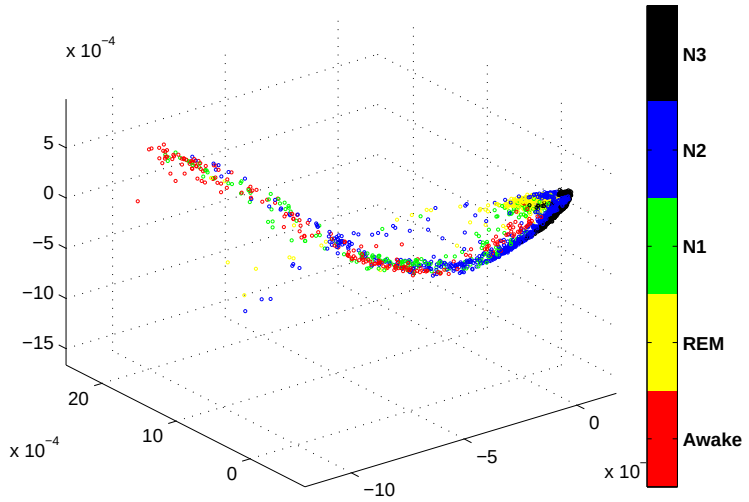
$$\theta_{\mathrm{O1A2}}(j) \xrightarrow{\Phi_{\mathrm{O1A2}}} \begin{bmatrix} \lambda_2^h u_2(j) \\ \lambda_3^h u_3(j) \\ \lambda_4^h u_4(j) \end{bmatrix}, \quad \theta_{\mathrm{O2A1}}(j) \xrightarrow{\Phi_{\mathrm{O2A1}}} \begin{bmatrix} \lambda_2^h \widetilde{u}_2(j) \\ \lambda_3^h \widetilde{u}_3(j) \\ \lambda_4^h \widetilde{u}_4(j), \end{bmatrix}$$

u_2, u_3, u_4 (resp. $\widetilde{u}_2, \widetilde{u}_3, \widetilde{u}_4$) are the first three nontrivial eigenvectors of $\mathbf{P}_{\mathrm{O1A2}}$ (resp. $\mathbf{P}_{\mathrm{O2A1}}$).

Feature clustering



(a) Φ_{O1A2}



(b) Φ_{O2Aa1}

After coloring the diffusion maps of $\theta_{\mathrm{O1A2}}(j)$ and $\theta_{\mathrm{O2A1}}(j)$ according to the sleep stage at the jth time slot obtained by the traditional visual scoring method, we observed that

At some level, Φ_{O1A2} and Φ_{O2A1} cluster underlying factors $\{\theta_{\mathrm{O1A2}}(j)|j = 1, ..., N\}$ and $\{\theta_{\mathrm{O2A1}}(j)|j = 1, ..., N\}$ according to the sleep stages.

Alternating diffusion map (ADM)

For merging information, we consider an alternative scheme:

Based on the diagonalization of $\mathbf{P}_{\mathrm{O1A2}}\mathbf{P}_{\mathrm{O2A1}}$

$$\mathbf{P}_{\mathrm{O1A2}}\mathbf{P}_{\mathrm{O2A1}} = \begin{bmatrix} u_1 & \cdots & u_N \end{bmatrix} \mathrm{Diag}(\lambda_1, \ldots, \lambda_N) \begin{bmatrix} v_1 & \cdots & v_N \end{bmatrix}^{\mathrm{T}},$$

the alternating diffusion map $\Phi_{\mathrm{alt}} : \mathbb{R}^{2d} \to \mathbb{R}^{N-1}$ is defined by

$$\big(\theta_{\mathrm{O1A2}}(j), \theta_{\mathrm{O2A1}}(j)\big) \xrightarrow{\Phi_{\mathrm{alt}}} v_j := \begin{bmatrix} \lambda_2^h u_2(j) \\ \lambda_3^h u_3(j) \\ \lambda_4^h u_4(j) \\ \vdots \\ \lambda_N^h u_N(j) \end{bmatrix}.$$

Lederman, R. R., Talmon, R., Wu, H. T., Lo, Y. L., & Coifman, R. R. Alternating diffusion for common manifold learning with application to sleep stage assessment. In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing.

Visualization of the alternating diffusion map

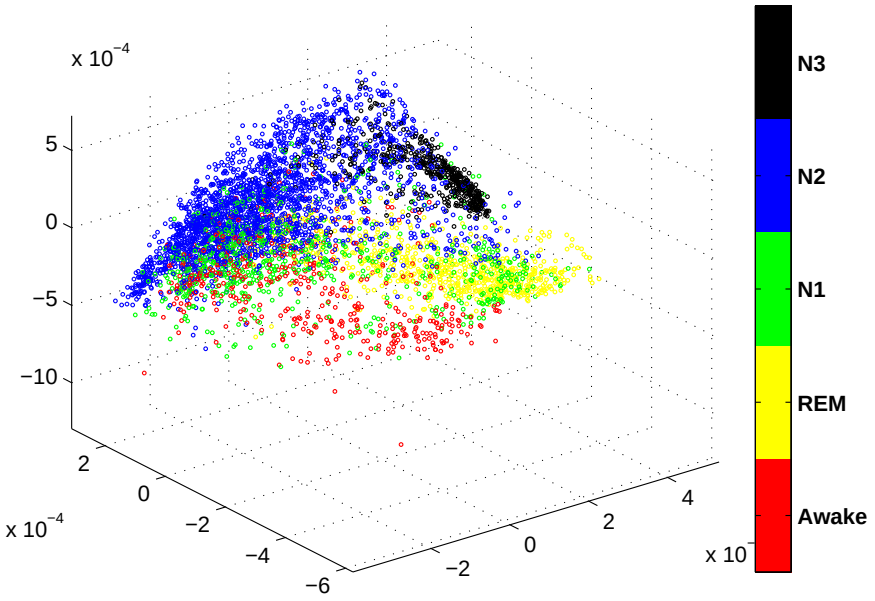
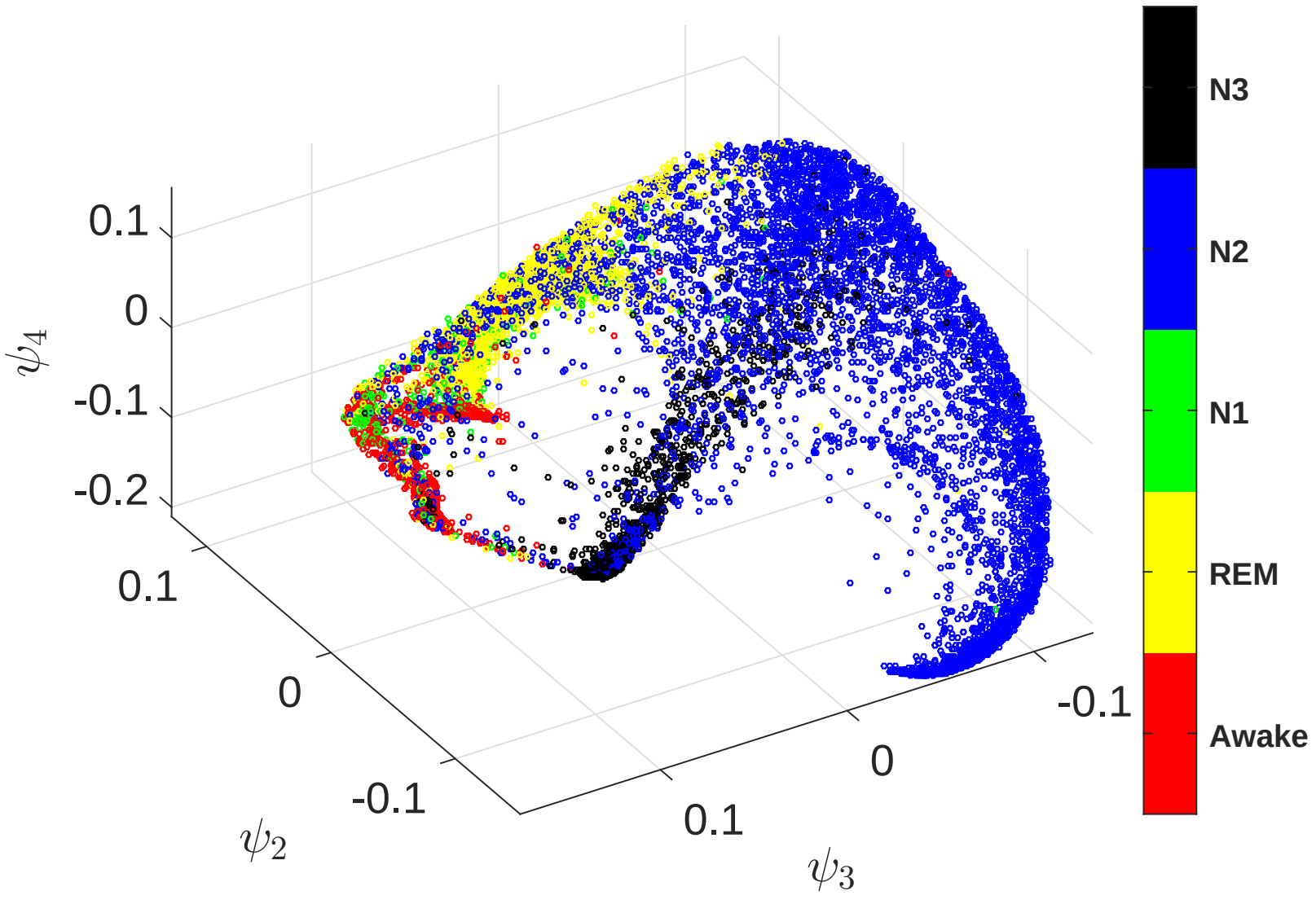


Figure: Alternating diffusion map of $\left(\theta_{\mathrm{O1A2}}(j), \theta_{\mathrm{O2A1}}(j)\right)$, $j = 1, 2, ..., N$.

The figure above shows that the points $v_j = \Phi_{\mathrm{alt}}\left(\left(\theta_{\mathrm{O1A2}}(j), \theta_{\mathrm{O2A1}}(j)\right)\right), j = 1, ..., N$, can be divided into five groups (labeled by Awake, REM, N1, N2, N3) roughly.

ADM of EEG features extracted from different subjects



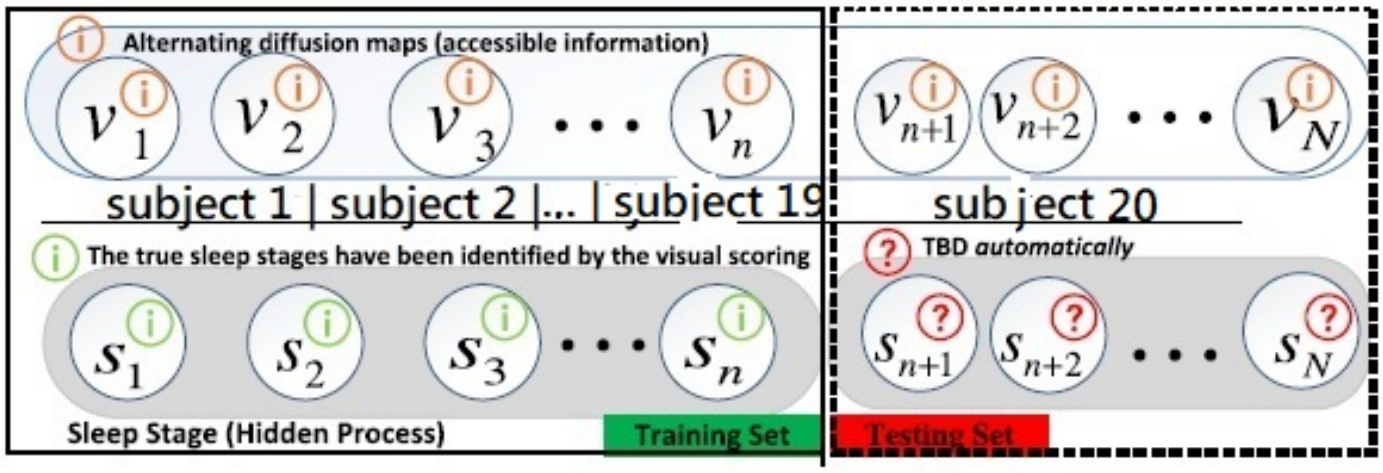
Problem formulation (Inter-individual sleep assessment)

Notation:

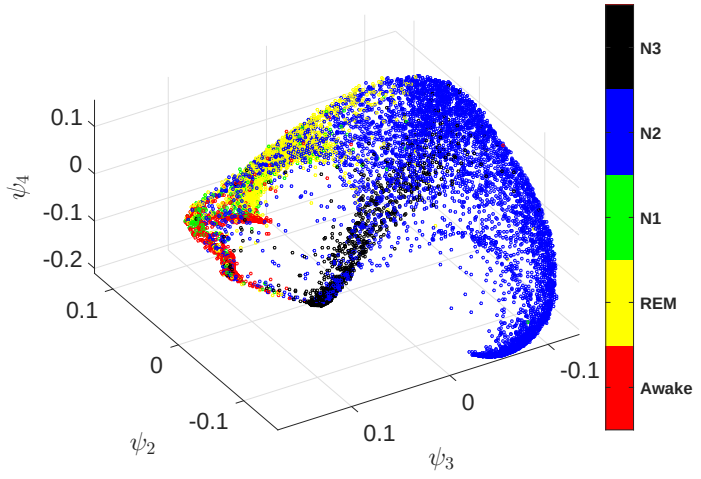
- v_j: the diffusion-mapped feature for the jth frame/epoch
- s_j: the label of v_j

$$s_j \in \{\text{Awake, REM, N1, N2, N3}\}.$$

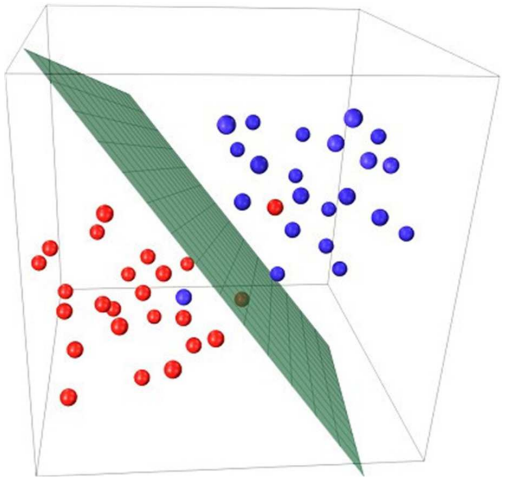
Question: When the sleep-stage pattern for the 20th subject is unknown, i.e., the labels of v_{n+1}, v_{n+2},...,v_N are unknown for some n, how to recover it?



Approach 1: Support Vector Machine (SVM)



(a) Partially colored ADM (Only $v_1, \ldots, v_n$ are colored according to their sleep stages.)



(b) Idea of SVM

The idea of SVM is

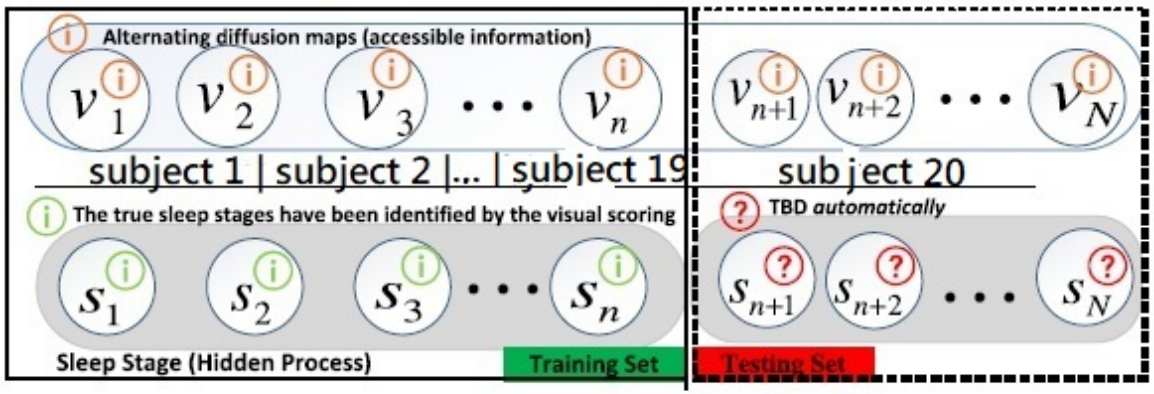
- Splitting the $\mathbb{R}^3$ space into several regions based on the colored alternating diffusion maps $v_j, j \in \{1, 2, \ldots, n\}$.
- For each $j \in \{n, n+1, \ldots, N\}$, the sleep stage s_j is guessed according to v_j belonging to which region.

Approach 2: Hidden Markov Model (HMM)

- The sleep-stage sequence $s_{n+1}, s_{n+2}, \ldots$ is modeled by a Markov chain.
  The states can not be observed directly.
  Empirical transition matrix ($\mathbf{p}$).
  $\forall\, s, s' \in \{\text{Awake,REM,N1,N2,N3}\}$,

$$p(s'|s) = \left[\sum_{j=1}^{n} 1\{s_j = s, s_{j+1} = s'\}\right] \left[\sum_{\ell=1}^{n} 1\{s_\ell = s\}\right]^{-1},$$

- Each state s_k, where $k \in \{n+1, n+2, \ldots\}$, has an emission v_k.



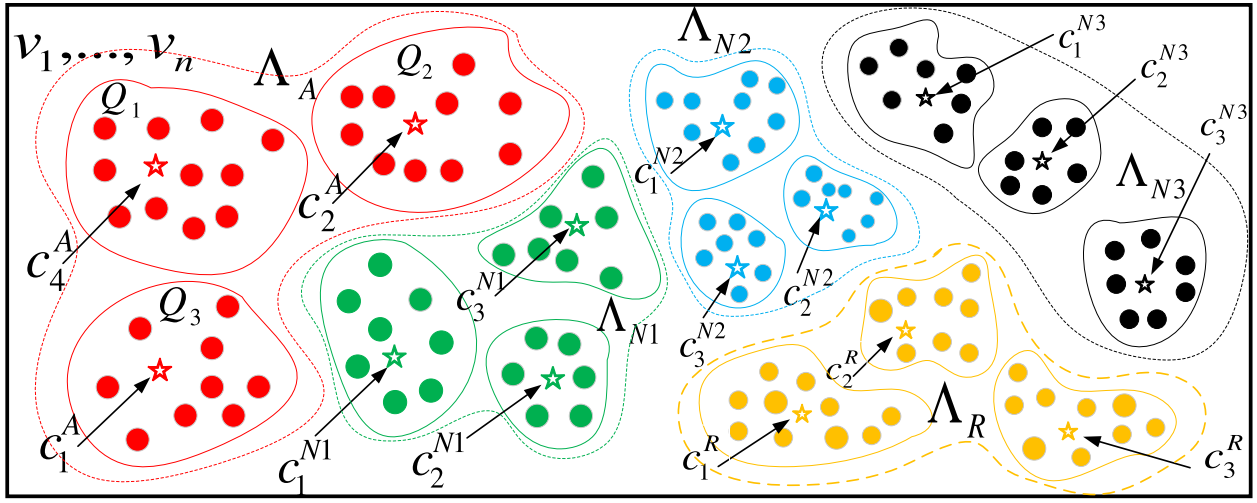
Discretization of the observation space

To introduce a HMM, we need to create a cookbook

$$\mathbf{B} = \{c_1^A, \ldots, c_K^A, c_1^R, \ldots, c_K^R, c_1^{N1}, \ldots, c_K^{N1}, c_1^{N2}, \ldots, c_K^{N2}, c_1^{N3}, \ldots, c_K^{N3}\},$$

where $c_1^A, \ldots, c_K^A$ are the centroids of the partition $\{Q_1, \ldots, Q_K\}$ of $\Lambda_A = \{v_j | s_j = \text{Awake}, j \leq n\}$ and $\{Q_1, \ldots, Q_K\}$ minimizes the value

$$\sum_{i=1}^{K} \sum_{x \in P_i} |x - \text{centroid}(P_i)|^2, \text{ where } \Lambda_A = \bigsqcup_{i=1}^{K} P_i. \tag{4}$$



◇ Vector quantization: $v_j \xrightarrow{\text{VQ}} e_j = \arg\min_{c \in \mathbf{B}} |v_j - c|, j = 1, \ldots, N.$

Based on $\{s_j, e_j\}_{j=1}^n$, we consider a hidden Markov chain with

$\diamond$ Transition matrix (**p**). $\forall\ s, s^{'} \in \{$Awake,REM,N1,N2,N3$\}$,

$$p(s^{'}|s) = \left[\sum_{j=1}^{n} 1\{s_j = s, s_{j+1} = s^{'}\}\right]\left[\sum_{\ell=1}^{n} 1\{s_\ell = s\}\right]^{-1},$$

$\diamond$ Emission matrix (**b**). $\forall s \in \{$Awake,REM,N1,N2,N3$\}$ and $c \in \mathbf{B}$,

$$b_s(c) = \left[\sum_{j=1}^{n} 1\{s_j = s, e_j = c\}\right]\left[\sum_{\ell=1}^{n} 1\{s_\ell = s\}\right]^{-1}.$$

Conditioning on $S_n = s_n$ and emissions $E_j = e_j$ for $j \geq n+1$,

$$\begin{aligned}
&\Pr\left(S_{n+1} = s_{n+1}, ..., S_N = s_N | e_{n+1}, ..., e_N, S_n = s_n\right) \\
=&p(s_{n+1}|s_n)\left[\prod_{j=n+1}^{N-1} b_{s_j}(e_j)p(s_{j+1}|s_j)\right]b_{s_N}(e_N)\Pr(E_{n+1} = e_{n+1}, \ldots\ |S_n = s_n)^{-1}.
\end{aligned}$$

Results for the inter-individual sleep assessment

◇ Using the labeled diffusion-mapped features $\{(v_j, s_j)\}_{j=1}^{n}$ to train the hidden Markov model (HMM).

◇ Revealing unknown sleep stages $\{s_j\}_{j=n+1}^{N}$ based on the accessible information $(v_{n+1}, \ldots, v_N)$.

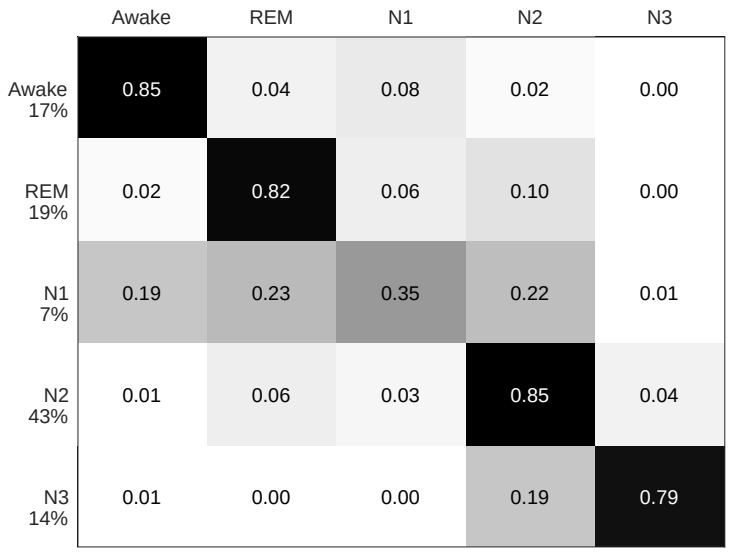
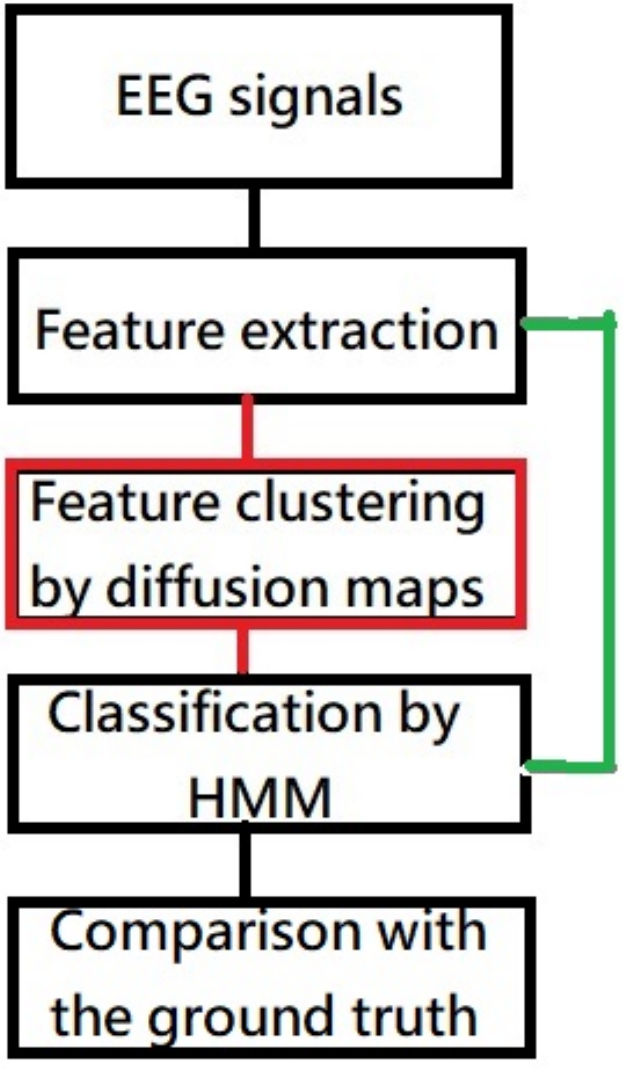


Figure: Mean confusion matrix(Explanation: 17% of the entire sleep period belongs to Awake and the HMM classifier has a 85% chance to make correct predictions if the underlying sleep stage is Awake.)

Does the ADM indeed improve the accuracy of classification?



Classification accuracy (ACC) comparison

The classification is performed in the alternating diffusion-mapped features and in the SST-extracted features, respectively.

Database	Classification is performed in ...	
	Diffusion mapped features	SST-extracted features
Sleep-EDF SC*	82%	63%
Sleep-EDF ST*	76%	64%
CGMH	68%	62%

Thank you for your attention!