

# Nonparametric Hierarchical Hidden Semi-Markov Model for Brain Fatigue Behavior Detection of Pilots During Flight

Edmond Q. Wu<sup>ID</sup>, Member, IEEE, Li-Min Zhu<sup>ID</sup>, Member, IEEE, Gui-Jiang Li, Hong-Jun Li, Zhiri Tang, Student Member, IEEE, Ruihan Hu, and Gui-Rong Zhou

**Abstract**—The evaluation of pilot brain activity is very important for flight safety. This study proposes a Hidden semi-Markov Model with Hierarchical prior to detect brain activity under different flight tasks. A dynamic student mixture model is proposed to detect the outlier of emission probability of HSMM. Instantaneous spectrum features are also extracted from EEG signals. Compared with other latent variable models, the proposed model shows excellent performance for the automatic inference of brain cognitive activity of pilots. The results indicate that the consideration of hierarchical model and the emission probability with  $t$  mixture model improves the recognition performance for Pilots' fatigue cognitive level.

**Index Terms**—Pilots' fatigue, electroencephalogram signals, hidden Markov models, hierarchical Dirichlet process,  $t$ -mixture model.

## I. INTRODUCTION

### A. Motivation

MANY flight accidents indicate that quantitative identification of pilot cognitive behavior can be used to assess the reliability of pilots in hazardous conditions such as severe weather. When pilots manipulate an airplane, the underlying state of brain activity can represent their cognitive level [14]. This problem can be further divided into three consecutive sub-problems: 1) how to extract more useful brain activity characteristics; 2) how to detect their latent state; and 3) how to improve the detection accuracy and robustness of cognitive level. Therefore, this work develops some instantaneous

Manuscript received September 26, 2019; revised September 19, 2020 and November 28, 2020; accepted January 15, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61671293 and Grant U1933125, in part by the Chinese Military Commission Equipment Development Department under Grant 6142219180403, and in part by the Jiangxi Province Key Research and Development Program under Grant 2019BBE50065. The Associate Editor for this article was F.-Y. Wang. (Corresponding authors: Edmond Q. Wu; Li-Min Zhu; Zhiri Tang; Ruihan Hu.)

Edmond Q. Wu is with the Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: edmondqwu@163.com).

Li-Min Zhu is with the Department of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhulm@sjtu.edu.cn).

Gui-Jiang Li is with The First Aircraft Institute, Xi'an 710089, China.

Hong-Jun Li is with First Army Representative Office, Xi'an 710089, China.

Zhiri Tang is with the Department of Computer Science, City University of Hong Kong, Hong Kong, SAR, China (e-mail: gerin.tang@my.cityu.edu.hk).

Ruihan Hu is with the Guangdong Key Laboratory of Modern Control Technology, Guangdong Institute of Intelligent Manufacturing, Guangzhou 510070, China (e-mail: rh.hu@giim.ac.cn).

Gui-Rong Zhou is with the COMAC Shanghai Aircraft Design and Research Institute, Shanghai 201210, China.

Digital Object Identifier 10.1109/TITS.2021.3052801

spectral characteristics via a smooth pseudo Wigner-Ville distribution. An improvement version of hidden semi-Markov model is inferred to detect latent status of EEG signals. Student mixture model is proposed to learn the emission probability, which is more robust to outlier and noises.

### B. Related Work

1) *Fatigue Evaluation*: EEG signals have been applied to evaluate cognitive level of human beings in different task learning [12], [17]. Cognitive behavior presents not only the creativity and judgment of a human [12], [16], but also the confusion status. The cognitive behavior detection has been discussed widely in many scenes [21]–[23], [25], [48], [49]. How to find the degree of a cognitive stage is its important challenge.

There have been many achievements in the field of fatigue detection. Liu *et al.* [1] propose a generalized prediction system called a recurrent self-evolving fuzzy neural network to address brain dynamics for driving fatigue. The results indicate that their model is superior to competing models regardless of the use of recurrent or nonrecurrent structures. Bose *et al.* [2] develop a regression-based continuous fatigue estimation by using power spectral features in conjunction with time as the fatigue label. Compared with other three regression models, the results demonstrate their effects on the performance of fatigue estimation. Foong *et al.* [4] investigates an EEG-based motor imagery brain-computer interface employing visual feedback for upper-limb stroke rehabilitation, and presents EEG correlates of mental fatigue during BCI usage. Beta power is more closely related to fatigue. Gao *et al.* [9] develop a novel EEG-based spatial-temporal convolutional neural network to detect driver fatigue. Their model fulfills a better classification accuracy of 97.37% than these compared methods. Sun *et al.* [10] identifies three of the most effective contextual features, i.e., continuous driving time, sleep duration time, and current time, to facilitate the online recognition of fatigue state. By applying gray relational analysis, the three contextual features, together with the most effective facial and vehicle behavior features, are introduced in a two-level fusion structure to improve fatigue driving recognition. Chaudhuri *et al.* [11] develop approximate and sample entropies from specific electrodes for different subjects and varying fatigue levels. These measures are used to train a support vector machine. They obtain a classification accuracy of 86%. Huang *et al.* [13] propose a multi-granularity deep

convolutional model for driver fatigue detection. Experimental results on the public drowsy driver dataset demonstrate significant performance improvements of their model over all published state-of-the-art methods. Mandal *et al.* [15] propose a vision-based fatigue detection system for bus driver monitoring. The system consists of modules of head-shoulder detection, face detection, eye detection, eye openness estimation, fusion, and drowsiness measure percentage of eyelid closure estimation. Zhang *et al.* [18] propose a novel framework based on spatial clustering to explore the sources of mental fatigue and functional activity changes. The results show that the extracted nodes correspond to the fMRI sources across different subjects and different tasks. Qi *et al.* [19] developed a fully cross-validated, data-driven analysis framework incorporating multivariate regression model to explore the feasibility of utilizing functional connectivity to predict the fatigue-related behavioral impairment at individual level. Their findings extended conventional brain-behavioral correlation analysis to individualized prediction of fatigue-related behavioral impairments. Lin *et al.* [20] proposed a 4-D convolutional neural-network algorithm for monitoring changes in the human brain state and behavior. It achieves superior forecasting performance over its peers. The above supervised learning models rely more on labeled fatigue features. Therefore, it is very important to learn the latent variables of fatigue characteristics. He *et al.* [47] proposes a common Bayesian network (CBN), to discriminate multiclass motor imagery EEG signals. They conduct experiments on two well-known BCI datasets and perform a numerical analysis of the propose algorithm for EEG classification in a multiclass motor imagery BCI. Experimental results show that CBN method not only has excellent classification performance, but also is highly efficient.

2) *Latent State Detection Model:* To quantitatively assess workload, latent variable models are often used to detect the cognitive levels. The typical model is Hidden Markov Model (HMM) [29]. Li and Wang [51] propose a mandarin Chinese singing voice synthesis system through a hidden Markov model. Their experimental results demonstrate that their system outperforms the baseline system in both objective and subjective evaluations. HMMs have good performance in detection of latent states, while they also have some drawbacks: 1) the number of latent states is a priori, which results in the inferred models that do not match actual data. Its parameters need to be specially set, faced with different data sets; 2) their geometric duration distribution is not suitable for many situations; 3) its emission probability is usually modeled by the Gaussian Mixture Model (GMM), which cannot handle real applications with outliers. Dong and Zhou [52] presents a GMM-based evolutionary strategy to solve multimodal optimization problems. Their results show that the improved GMM model is not only simple to program and understand, but also provides better and consistent performance.

However, an important issue in HMM is how to choose the correct number of latent states. Bayesian nonparametric (BNP) provides a good solution for this issue [29], [31], namely Infinite Hidden Markov Model (IHMM) [28]. It provides a way to allow the machine to learn the correct number of latent states. BNP [36] applied to HMM relies on the hierarchical

Dirichlet process (HDP) [27] to infer the number of hidden states of the posterior distribution [32]–[34]. Some extensions of the HDP-HMM model, such as adaptive HDP-HMM [31], left-to-right HDP-HMM [37] and online HDP-HMM [38] have been developed for more applications. The traditional HDP-HMM have no ability to simulate the temporal persistence of states, which lead to a lot of redundancy and rapid shifting of states. One better solution is to improve the model by sticky extension [35], [39]. For example, the sticky HDP-HMM [35], [39] design a self-transfer bias to stop the fast shift and does not require adjustment of the penalty parameter. Compared with the standard HDP-HMM, it has better robustness [35]. However, it limits the expressiveness of the model in terms of duration structure. At the same time, the self-transition shares among all countries, which bans the duration information of specific countries. Therefore, another method is to use the hidden semi-Markov model (HSMM) [30], [40]–[42], which is an extended version of HMM [43]. Compared to HMM, HSMM with the state resident probability is defined as an explicit type. It solves the state limitation problem in HMM by adding time components. The improvement of this model is to add the duration of a particular state to a regular HSMM. When the Markov chain enters the state, its duration is estimated from a particular state, and when the duration ends, the Markov chain transitions to the next state. Unlike standard HMMs, each state corresponds to an observation.

### C. Contribution

The emission probability of HMM is extensively modeled by a GMM, because more mixture components can approximate more latent distribution. However, GMMs have no capability for outliers and need more time to model them. This leads to an increase in the mixture components, which is not a good solution because the fitted model does not conform to the real data. It is clear that the long tail of student distribution can capture these outliers. To identify the distributions of emission, this work develops a Dynamic Student Mixture Model (DSMM) that observes the dynamic changes in the emission distribution. Therefore, it is very suitable for fatigue detection.

These shortcomings of HDP-HMM prompt us to explore HDP-HSMM with  $t$  distribution for emission probability. A hidden semi-Markov model with HDP priors constructs a general model that allows Bayesian nonparametric inference to the probability of transmission. The proposed model is used to infer cognitive level of pilot brain activity. The proposed technical flow chart is shown in Fig.1. It includes three steps: 1) rhythm extraction and fatigue indicators feature learning via a smooth pseudo Wigner-Ville distribution; 2) fatigue cognitive feature dimensionality reduction via a Treelet transformation; and 3) Hidden state learning of brain cognition via an HDP-t-HSMM.

The contribution of this work can be summarized as follows:

- (1) An HDP-HSMM is proposed by combining the HDP with HSMM for automatic detection of brain activity instead of relying on human experience.

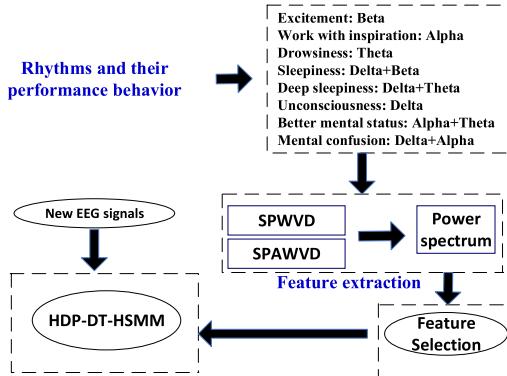


Fig. 1. The framework of brain cognitive inference.

- (2) Student mixture model is proposed to detect emission probability of HSMM instead of traditional Gaussian mixture model.
- (3) The application under wicked flight environment like drastic bumps provides new perspectives for the quantification of pilots' cognitive status.

The rest of this paper is organized as follows: Section II and III proposes HDP-t-HSMM and its solution. Experiments and discussions are arranged in Section IV. Section V draws the conclusion.

## II. HDP-T -HSMM

### A. Hidden Semi-Markov Model With Remaining Service Life

Suppose the current time is  $t$ , the process has made  $n - 1$  jumps, and the time spent since the previous jump is  $X_t = t - \sum_{l=1}^{n-1} d_l$ , the process  $\{S_{1:T}, O_{1:T}\}$  is a hidden semi-Markov process. In HSMM [30], when a state ends at time  $t$ , it cannot transit to the same state at the next time  $t + 1$  because the state durations are explicitly specified by some distributions. Then, a regular HSMM can be given as

$$\begin{cases} \alpha_{(i,h)(j,d)} \equiv P[S_{[t+1:t+d]} = j | S_{[t-h+1:t]} = i] \\ p_j(d) \equiv P[S_{t+1:t+d} = j | S_{t+1} = j] \\ \pi_{j,d} = P[S_{[t-d+1:t]} = j], t \leq 0 \\ b_{j,d}(O_{t+1:t+d}) \equiv P[O_{t+1:t+d} | S_{[t+1:t+d]} = j] \end{cases} \quad (1)$$

where  $\alpha_{(i,h)(j,d)}$  is transition probability. It is independent of time, for  $i, j \in S, h, d \in D$ , which satisfy  $\sum_{i \in S \setminus \{j\}} \sum_{d \in D} \alpha_{(i,h)(j,d)} = 1$ , for all given  $i \in S$  and  $h \in D$ , with zero self-transition probabilities  $\alpha_{(i,h)(j,d)} = 0$ , for all  $i \in S$  and  $h, d \in D$ .  $b_{j,d}$  is emission probability.  $p_j(d)$  is the probability of duration at state  $j$ .  $\pi_{j,d}$  is the probability of the initial state and its duration before time  $t = 1$  or before the first observation  $O_1$  is obtained.

The occurrence of brain fatigue cognitive state is our interest, and its development status can be monitored, which is the focus of our research. Via a review of some HSMMs [30], this work presents a definition of a remaining service life HSMM that can represent this phenomenon.

**Definition 1.** Suppose that the state transition probability is independent of the previous state and its duration. The obtained model is as the remaining service life HMM, as shown in Fig. 2.

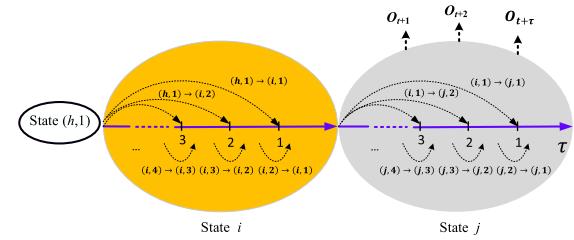


Fig. 2. Remaining service life HSMM.

(1) If the state transition is assumed to be independent of the duration of its previous state, the state transition probability becomes  $\alpha_{(i,h)(j,d)} = \alpha_{i(j,d)}$ .

(2) If it is further assumed that the state duration is independent of its previous state, the state transition probability is  $\alpha_{i(j,d)} = \alpha_{i,j} p_i(d)$ .

State  $i$  transits to state  $j$  at its remaining service life 1 in terms of the transition probability  $\alpha_{i(j,\tau)}$ . Our interest focuses on the ending time, i.e., time 1. The occurrence of state is not of our concern. Analogously, this work only cares the brain fatigue cognitive degree instead of its occurrence. The starting phenomenon of fatigue behavior is difficult to be detect. State  $j$  has remaining service life  $\tau$ , and it tails off to 1. Then, state  $j$  end at the  $(t + \tau)$ th time and transit to another state. During state  $j$ , emission probability  $b_j(o_{t+1:t+\tau})$  form  $\tau$  observations  $O_{t+1}, \dots, O_{t+\tau}$ .

Then, parameters set of the proposed HSMM is given as

$$\lambda \equiv \{\alpha_{i(j,\tau)}, b_j(o_{t+1:t+\tau}), p_j(d), \pi_{j,d}\} \quad (2)$$

The observations are able to be supposed dependent or conditionally independent for given states, that is,  $b_j(o_{t+1:t+\tau}) = \prod_{\tau=1}^d b_j(o_{t+\tau})$ . The conditional independence is propitious to the modelling of HSMMs.

### B. Dynamic Student Mixture Model

In this paper, SMM is proposed to approximate the emission distribution of HSMM.

The SMM is defined as

$$\begin{cases} p(x) = \sum_{k=1}^N c_{jk} S(x | Z = k, \xi_k), 0 \leq c_{jk} \leq 1 \\ \xi_k = \{\mu_k, \sum_k, v_k\} \end{cases} \quad (3)$$

with

$$\begin{cases} S(\xi_k) = \frac{\Gamma(\frac{v_k+D}{2})}{\Gamma(\frac{v_k}{2})} |\sum_k|^{-\frac{1}{2}} \left(1 + \frac{\Delta_x^2(\mu_k, \sum_k)}{v_k}\right)^{-\frac{1}{2}(v_k+D)} \\ \Gamma(x) = \int_0^\infty v^{x-1} e^{-v} dv \\ \Delta_x^2(\mu_k, \sum_k) = (x - \mu_k)^T \sum_k^{-1} (x - \mu_k) \end{cases} \quad (4)$$

where  $p(x)$  is likelihood function of SMM,  $K$  is the number of mixture components,  $D$  is the dimensionality of  $x$ ,  $c_{jk}$  is the mixing weight,  $\sum_k c_{jk} = 1$ ,  $S(x | Z = k, \theta_k)$  is  $t$  distribution with  $\xi_k = \{\mu_k, \sum_k, v_k\}$ .  $v_k$  is the degree-of-freedom which controls the tail length of  $t$  distribution. In particular,  $t$  distribution with  $v_k \rightarrow \infty$  becomes a Gaussian distribution with mean of  $\mu_k$  and covariance of  $\sum_k^{-1} (\mu_k, \sum_k)$  is the square Mahalanobis distance.

Now the emission distribution becomes

$$b_j(x_t) = \sum_{k=0}^{K-1} c_{jk} S\left(x_t \mid Z_{jt} = k, \mu_{jk}, \sum_{jk}, v_{jk}\right) \quad (5)$$

Note that there are no close-form solutions for the degree of freedom  $v_{jk}$ . An additional representation of the  $t$  distribution can be used in (6), as shown at the bottom of the next page, where  $(D = x_k - \mu_k).N(\cdot)$  is a Gaussian distribution, and  $g(\cdot)$  is a Gamma one. Evidently,  $t$ -distribution is an infinite weighted sum of Gaussians with different precisions.

In order to enforce consistency of component positions over a period of time, this work introduces a priori on the positional parameters that penalize large changes in continuous time steps. This prior has the following joint multivariate Gaussian distribution in (7), as shown at the bottom of the next page, where Gaussian parameters  $(\mu_k, Q_k^\pm)$  obey Gaussian-Wishart priors.  $t \in \{1, \dots, T\}$  denotes the time frame for an EEG signal.  $u$  is the mean of  $\mu_{kt}$ , and  $r$  is the relative precision of  $\mu_{ct}$ .  $v$  is the number of degrees of freedom for  $Q_k^\pm$ .  $W(Q_k^\pm | S^\mp, v)$  is the Wishart distribution.  $Q_k^\pm$  is a  $d \times d$  symmetric matrix of random positive variable.  $S^\pm$  is a positive definite matrix of size  $d \times d$ , and its inverse matrix  $S^\mp$  represents the scale matrix for  $Q_k^\pm$ . Then, if  $v \geq d$ ,  $Q_k^\pm$  has a Wishart distribution with  $v$  degrees of freedom. The dynamic behavior can be controlled by covariance matrix  $Q$ .

Next, this work obtains the maximum a posteriori (MAP) estimation of the model parameters  $\theta$  by maximizing the log-posterior, which is equivalent (up to an additive constant) to:

$$\begin{aligned} L(\theta) = & \sum_{k=1}^N \log p(x_k \mid \mu_k, \sum_k, v_k) \\ & + \sum_{c=1}^C \log p_{pri}(\mu_{ct} \mid \mu_{ct-1}, Q) \end{aligned} \quad (8)$$

### C. HDP-t-HSMM

Similar to HDP-HMM, this work presents generative process of HDP-t-HSMM, which can be defended as follows:

$$\left. \begin{aligned} \gamma &\sim \text{Gam}(c, d) \\ \beta &\sim \text{GEN}(\gamma) \\ &\text{iid} \\ \pi_j &\sim \text{DP}(\alpha, \beta) \\ \\ (\theta^{(i)}, \vartheta^{(i)}) &\sim H \times G \quad i = 1, 2, \dots, N \\ \\ z_s &\sim \bar{\pi}_{z_s} \\ d_s &\sim g(\vartheta^{(z_s)}) \\ x_{t(s):t(s+1)-1} &= z_s \\ y_t &\sim f(\theta^{(x_t)}) \\ f(\theta^{(x_t)}) &= \sum_{k=0}^{K-1} c_{jk} S(x \mid Z = k, \mu_k, \sum_k, v_k) \end{aligned} \right\} \quad (9)$$

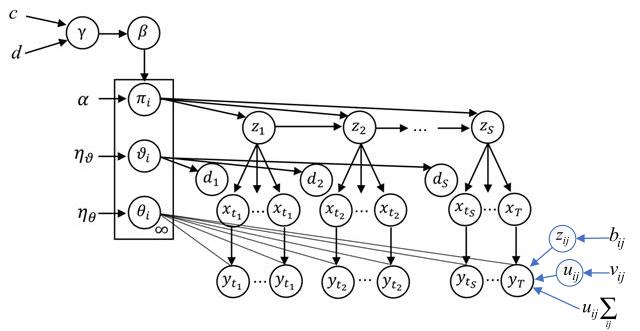


Fig. 3. A graphical model of the HDP-t-HSMM.

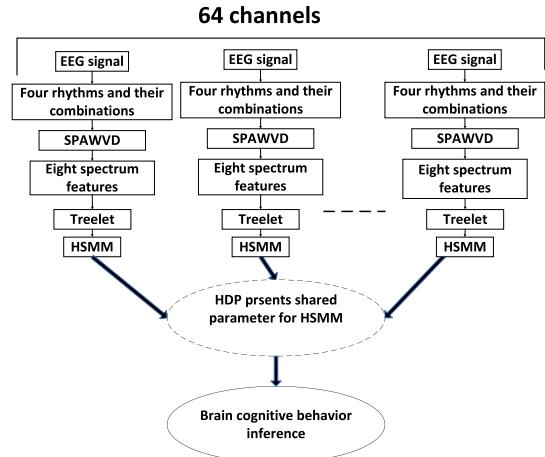


Fig. 4. The flow chart of brain cognitive inference.

where  $f$  and  $g$  denote its emission and duration distributions. They obey from  $H$  and  $G$ , respectively. GEM denotes a stick breaking process.  $\gamma$  denotes a Gamma function.

The HDP-t-HSMM shown in Fig.3 can be modeled by the set of parameters  $\Psi = \{c, d, \alpha, \lambda, C, \xi\}$  where  $\lambda$  is parameter of HSMM.  $C = (c_{ij})_{i,j=1}^{N,K}$  is the mixture coefficient matrix of hidden emission distribution, with  $c_{i,j}$  denoting the mixture proportion of the  $j$ th component at the  $i$ th state.  $\xi$  is parameter of SMM. In Fig. 3, each continuous state  $d_s$  has some observations  $y_t$  corresponding to a certain state  $z_s$ .

HDP-t-HSMM is used to evaluate pilots' cognitive states under the specific case of stall detection and alerting. The technique chart of brain cognitive behavior inference is shown in Fig. 4. This work provides a parameter shared mechanism for multi-HSMMs used in the detection of EEG states of each channel. In Fig. 4, the EEG signal changes of each channel are different. This work extracted the cognitive features of each channel through SPAWD, and then reduced the dimensionality through treelet transformation, and then used HSMM to model the dynamic relationship of cognitive features of each channel. All HSMMs in Fig. 4 share a group of parameters [24] through an HDP, which presents a behavior show of whole brain cognitive for a human.

### III. LEARNING ALGORITHMS

#### A. Forward-Backward Algorithm for HSMM With Remaining Service Life

When state  $i$  is transited to state  $j$  with remaining service life  $\tau$ , its transition probability is

$$\left\{ \begin{array}{l} \alpha_i(j, \tau) \equiv P[S_{[t:t+\tau-1]} = j | S_{t-1} = i] = \alpha_{ij} P_j(\tau) \\ \sum_{i \in S \setminus \{j\}} \sum_{\tau \in D} \alpha_{i(j, \tau)} = 1, i \neq j \end{array} \right. \quad (10)$$

The forward variable  $\check{\alpha}_t(i, \tau)$  is the joint probability that the partial observation sequence and the current state  $i$  that stays for the next  $\tau$  steps and then end at  $t + \tau - 1$ . The backward variable  $\check{\beta}_t(i, \tau)$  is the conditional probability that the future observations and the current state  $i$  with  $\tau$  steps of remaining service life. Then, the forward-backward variables of HSMM with remaining service life can be defined by

$$\left\{ \begin{array}{l} \check{\alpha}_t(i, \tau) \equiv \sum_{S_{1:T}} P[S_{t:t+\tau-1} = i, O_{1:T} | \lambda] \\ \check{\beta}_t(i, \tau) \equiv \sum_{S_{1:T}} P[O_{t+1:T} | S_{t:t+\tau-1} = i, \lambda] \end{array} \right. \quad (11)$$

Because the state transition is assumed independent of the duration of the previous state, the following relationship holds.

$$\left\{ \begin{array}{l} \alpha_{(i,d)(j,\tau)} = \alpha_{i(j,\tau)} \\ \beta_\tau(i, d) = \check{\beta}_t(i, 1) \end{array} \right. \quad (12)$$

When the conditional independence of output is assumed, backward variable of HSMM with remaining service life has the following formulation:

$$\check{\beta}_t(i, \tau) = \check{\beta}_{t+\tau-1}(i, 1) \prod_{k=t+1}^{t+\tau-1} b_i(o_k) \quad (13)$$

The following relationships holds.

$$\sum_{d \geq \tau} \alpha_\tau(i, d) = \check{\alpha}_{t-\tau+1}(i, \tau) \prod_{k=t-\tau+2}^t b_i(o_k) \quad (14)$$

And then, forward-backward formulas of remaining service life HSMM can be presented as

$$\begin{aligned} \check{\alpha}_t(i, \tau) &= \check{\alpha}_{t-1}(i, \tau + 1) b_i(o_t) \\ &+ \sum_{i \in S \setminus \{j\}} \check{\alpha}_{t-1}(j, 1) \alpha_{ij} \beta_{t+1}(j, \tau) \end{aligned} \quad (15)$$

and

$$\left\{ \begin{array}{l} \check{\beta}_t(i, \tau) = b_i(o_{t+1}) \check{\beta}_{t+1}(i, \tau - 1), \tau > 1 \\ \check{\beta}_t(i, 1) = \sum_{i \in S \setminus \{j\}} \sum_{\tau \geq 1} \alpha_{ij} b_i(o_{t+1}) \check{\beta}_{t+1}(j, \tau) \end{array} \right. \quad (16)$$

The boundary conditions are  $\check{\alpha}_0(i, 1) = \pi_i$ ,  $\check{\alpha}_0(i, \tau) = 0$  for  $\tau > 1$ , and  $\check{\beta}_T(i, 1) = 1$ ,  $\check{\beta}_T(i, \tau) = 0$  for  $\tau > 1$ .

The computational complexity involved in the remaining service life HSMM can be reduced significantly via forward-backward variables (17) of HSMM with explicit duration, if the state duration is supposed to be independent of its previous state. Its forward-backward variables can be represented as

$$\left\{ \begin{array}{l} \alpha_t^*(j) = \sum_{i \in S \setminus \{j\}} \check{\alpha}_{t-1}(j, 1) \alpha_{ij} \\ \beta_{t+1}^*(j) = b_j(o_{t+1}) \sum_{\tau \in D} P_j(\tau) \check{\beta}_{t+1}(j, \tau) \end{array} \right. \quad (17)$$

where  $\alpha_t^*(j)$  is the joint probability that state  $j$  starts at  $t$  with the partial observation sequence is  $o_{1:t-1}$ ,  $\beta_{t+1}^*(j)$  is the conditional probability that state  $j$  starts at  $t+1$  with the future observation sequence is  $o_{t+1:T}$ .

Then the forward-backward variables of remaining service life HSMM are further obtained as

$$\left\{ \begin{array}{l} \check{\alpha}_t(i, \tau) = \check{\alpha}_{t-1}(i, \tau + 1) b_i(o_t) + \alpha_t^*(i) P_j(\tau) b_i(o_t) \\ \check{\beta}_t(i, 1) = \sum_{i \in S \setminus \{j\}} \alpha_{ij} \beta_{t+1}^*(j) \end{array} \right. \quad (18)$$

Their smoothed probabilities are obtained via forward-backward variables as

$$\left\{ \begin{array}{l} \xi_t(i; j, d) \equiv P[S_t = i, S_{[t+1:t+d]} = j, O_{1:T} | \lambda] \\ = \check{\alpha}_t(i, 1) \alpha_{i(j,d)} b_j(O_{t+1}) \check{\beta}_{t+1}(j, d) \\ \eta_{t+d}(j, d) \equiv P[S_{[t+1:t+d]} = j, O_{1:T} | \lambda] = \sum_{i \in S \setminus \{j\}} \xi_t(i; j, d) \end{array} \right. \quad (19)$$

where  $\xi_t(i; j, d)$  is the smoothed probability of the transition from state  $i$  to state  $j$  with duration  $d$  and the observation sequence at time  $t$ .  $\eta_{t+d}$  is the smoothed probability of transition at state  $j$  with duration  $d$  and the observation sequence at time  $t$ .

$$\left\{ \begin{array}{l} S(x_k | Z = k, \mu_k, \sum_k, v_k) = \int_0^{+\infty} \mathcal{N}\left(x_k | \mu_k, \frac{\sum_k}{\lambda}\right) G\left(\lambda | \frac{v_k}{2}, \frac{v_k}{2}\right) d\lambda \\ \mathcal{N}\left(xx_k | \mu_k, \frac{\sum_k}{\lambda}\right) = (2\pi)^{-\frac{d}{2}} |\frac{\sum_k}{\lambda}|^{\frac{1}{2}} \exp\left\{-\frac{1}{2} D^T \frac{\sum_k}{\lambda} D\right\} \\ G\left(\lambda | \frac{v_k}{2}, \frac{v_k}{2}\right) = \frac{(\frac{v_k}{2})^{\frac{v_k}{2}}}{\Gamma(\alpha)} \lambda^{\frac{v_k}{2}-1} \exp(-\frac{v_k}{2}\lambda) \end{array} \right. \quad (6)$$

$$\left\{ \begin{array}{l} p_{pri}(\mu_k, \mathbf{Q}_k^\pm) = \prod_{t=2}^T N\left((\mu_{kt} - \mu_{c(t-1)}) | u, (r\mathbf{Q}_k^\pm)^{-1}\right) W(\mathbf{Q}_k^\pm | \mathbf{S}^\mp, v) \\ W(\mathbf{Q}_k^\pm | \mathbf{S}^\mp, v) = \frac{1}{G} |\mathbf{Q}_k^\pm|^{\frac{v-d-1}{2}} \exp(-\frac{1}{2} \text{tr}(\mathbf{S}^\pm \mathbf{Q}_k^\pm)) \\ \mu_{kt} \sim \mu_{k(t-1)} + \mathcal{N}(u, Q) \end{array} \right. \quad (7)$$

### B. EM for the Learning of Model Parameters

Parameter  $\lambda$  of HSMM can be estimated via the following likelihood maximum of the observation sequence  $O_{1:T}$ .

$$L(\lambda) = \sum_{S_{1:T}} \sum_d (S_{[t+1:t+d]}, O_{t+1:t+d}; \lambda) \quad (20)$$

where  $\Sigma_{S_{1:T}}$  is the sum of all possible state sequence and  $\Sigma_d$  is the sum of all duration from time  $t$ .

Let  $\lambda^{(k)}$  denote the current value of  $\lambda$  at iteration  $k$ . The conditional expectation of the complete-data log-likelihood is given by

$$\begin{aligned} Q(\lambda | \lambda^{(k)}) \\ = E \left\{ P \left( S_{[t+1:t+d]}, O_{t+1:t+d}; \lambda | O_{t+1:t+d}; \lambda^{(k)} \right) \right\} \end{aligned} \quad (21)$$

The EM algorithm [26] maximizes  $L(\lambda)$  by iteratively maximizing iteratively maximizing over  $\lambda, \lambda^{(k+1)}$  is updated as

$$\lambda^{(k+1)} = \arg \max_{\lambda} \{Q(\lambda | \lambda^{(k)})\} \quad (22)$$

When the iteration of the EM algorithm increases, the sequence of re-estimated parameters  $\lambda^{(k)}$  converges to a local maximum of  $L(\lambda)$ . The conditional expectation  $Q(\lambda | \lambda^{(k)})$  can be rewritten as a sum of some terms, as shown in (23). Its each term depends on a given subset of parameters.

$$\begin{aligned} Q(\lambda | \lambda^{(k)}) \\ = Q_{\pi} (\{\pi_{j,d}\} | \lambda^{(k)}) + \sum_{i=1}^M Q_a (\{\alpha_{(i,h)(j,d)}\} | \lambda^{(k)}) \\ + \sum_{i=1}^M Q_b (\{b_{j,d}(v_{k_1:k_d})\} | \lambda^{(k)}) + \sum_{i=1}^M Q_d (\{p_i(d)\} | \lambda^{(k)}) \end{aligned} \quad (23)$$

with

$$\begin{aligned} Q_b (\{b_{j,d}(v_{k_1:k_d})\} | \lambda^{(k)}) \\ = \sum_{j=1}^M \sum_{t=0}^D P \left[ v_{k_1:k_d}, O_{t+1:t+d} \mid S_{[t+1:t+d]} = j, \lambda^{(k)} \right] \\ \times \log (b_{j,d}(v_{k_1:k_d})) \end{aligned} \quad (24)$$

Next, this work proposes a variable  $Z_{jt}$  indicating the mixture component and a variable  $\Lambda_{jkt}$  to (24). Its  $Q$  function is updated as

$$\begin{aligned} Q(e | \theta^{(i)}) \\ = \sum_{j=0}^{J-1} \sum_{t=0}^{T-1} \sum_{k=0}^{K-1} p(s_t = j, Z_{jt} = k, \Lambda_{jkt} = \lambda_{jkt} | x_0^{t-1}, \theta^{(i)}) \\ \dots \log \left( c_{jk} S \left( x_t \mid Z_{jt} = k, \mu_{jk}, \sum_{jk}, v_{jk} \right) \right) \end{aligned} \quad (25)$$

$$\begin{aligned} p(s_t = j, Z_{jt} = k, \Lambda_{jkt} = \lambda_{jkt} | x_0^{t-1}, \theta^{(i)}) \\ = p(s_t = j | x_0^{t-1}, \theta^{(i)}) p(Z_{jt} = k | s_t = j, x_0^{t-1}, \theta^{(i)}) \\ p(\Lambda_{jkt} = \lambda_{jkt} | s_t = j, Z_{jt} = k, x_0^{t-1}, \theta^{(i)}) \end{aligned} \quad (26)$$

Via the Gaussian scale mixture model,  $S(x_t | \mu_{jk}, \sum_{jk}, v_{jk})$  can be presented as follow:

$$\begin{aligned} \log S \left( x_t \mid Z_{jt} = k, \mu_{jk}, \sum_{jk}, v_{jk} \right) \\ = \log N \left( x_t \mid \mu_{jk}, \frac{\sum_{jk}}{\lambda_{jkt}} \right) + \log \left( \lambda_{jkt} \mid \frac{v_{jk}}{2}, \frac{v_{jk}}{2} \right) \end{aligned} \quad (27)$$

with

$$\begin{aligned} \log N \left( x_t \mid \mu_{jk}, \frac{\sum_{jk}}{\lambda_{jkt}} \right) &= \frac{D}{2} \log (2\pi) - \frac{1}{2} \log |\sum_{jk}| + \\ &\quad \frac{D}{2} \log (\lambda_{jkt}) - \frac{1}{2} \lambda_{jkt} \Delta_{x_t}^2 (\mu_{jk}, \sum_{jk}) \end{aligned} \quad (28)$$

$$\begin{aligned} \log \left( \lambda_{jkt} \mid \frac{v_{jk}}{2}, \frac{v_{jk}}{2} \right) &= -\log \Gamma \left( \frac{v_{jk}}{2} \right) + \frac{v_{jk}}{2} \log \left( \frac{v_{jk}}{2} \right) - \\ &\quad \frac{v_{jk}}{2} \lambda_{jkt} + \left( \frac{v_{jk}}{2} - 1 \right) \log (\lambda_{jkt}) \end{aligned}$$

After plugging these terms into the expectation of the complete log likelihood function (27), its posterior can be computed as

$$\begin{aligned} p(s_t = j | x_0^{t-1}, \theta^{(i)}) \\ = L_j(t) \end{aligned} \quad (29)$$

$$\begin{aligned} p(Z_{jt} = k | s_t = j, x_0^{t-1}, \theta^{(i)}) \\ = \frac{p(x_0^{t-1} | s_t = j, Z_{jt} = k, \theta^{(i)}) p(Z_{jt} = k | s_t = j, \theta^{(i)})}{\sum_k p(x_0^{t-1} | s_t = j, Z_{jt} = k, \theta^{(i)}) p(Z_{jt} = k | s_t = j, \theta^{(i)})} \\ = \frac{c_{jk}^{(i)} T \left( x_t \mid \mu_{jk}^{(i)}, \sum_{jk}^{(i)}, v_{jk}^{(i)} \right)}{\sum_{k=1}^{K-1} c_{jk}^{(i)} T \left( x_t \mid \mu_{jk}^{(i)}, \sum_{jk}^{(i)}, v_{jk}^{(i)} \right)} \end{aligned} \quad (30)$$

$$\begin{aligned} p(\Lambda_{jkt} = \lambda_{jkt} | s_t = j, Z_{jt} = k, x_0^{t-1}, \theta^{(i)}) \\ = g \left( \frac{v_{jk}^{(i)} + D}{2}, \frac{v_{jk}^{(i)} + \Delta_{x_t}^2 (\mu_{jk}^{(i)}, \sum_{jk}^{(i)})}{2} \right) \end{aligned} \quad (31)$$

From (26), it is necessary to compute the expectations of the hidden variable  $\Lambda_{jkt}$  and  $\log(\Lambda_{jkt})$ . Because the posterior of  $\Lambda_{jkt}$  is a Gamma distribution, its expectation is

$$\begin{aligned} \lambda_{jkt} &= E \left( \Lambda_{jkt} = \lambda_{jkt} \mid s_t = j, Z_{jt} = k, x_0^{T-1}, \theta^{(i)} \right) \\ &= \frac{v_{jk}^{(i)} + D}{v_{jk}^{(i)} + \Delta_{x_t}^2 (\mu_{jk}^{(i)}, \sum_{jk}^{(i)})} \end{aligned} \quad (32)$$

The E-step is to compute Eqs. (29)–(32). The partial derivatives of (25) can be set to zero, and some equations can be obtained. Their solutions are the following M-step:

$$\begin{cases} c_{jk}^{(i)} = \frac{\sum_{t=0}^{T-1} \tilde{L}_{jk}(t)}{\sum_{t=0}^{T-1} \tilde{L}_j(t)} \\ \mu_{jk}^{(i)} = \frac{\sum_{t=0}^{T-1} \tilde{L}_{jk}(t) \tilde{\lambda}_{jkt} x_t}{\sum_{t=0}^{T-1} \tilde{L}_{jk}(t) \tilde{\lambda}_{jkt}} \\ \sum_{jk}^{(i)} = \frac{\sum_{t=0}^{T-1} \tilde{L}_{jk}(t) \tilde{\lambda}_{jkt} (x_t - \mu_{jk}^{(i+1)}) (x_t - \mu_{jk}^{(i+1)})}{\sum_{t=0}^{T-1} \tilde{L}_{jk}(t)} \end{cases} \quad (33)$$



Fig. 5. The flight simulator and monitoring system.

TABLE I  
PERFORMANCE COMPARISON AMONG DIFFERENT MODELS

Benchmark	index	GMM	SMM	DSMM
thyroid	MSE	4.0578	3.7525	<b>3.4506</b>
	WSS/BSS	4.0767	3.0863	<b>2.8057</b>
seed	MSE	5.5924	4.4052	<b>3.5002</b>
	WSS/BSS	1.4538	0.9143	<b>0.7061</b>
iris	MSE	3.4512	4.8157	<b>1.6730</b>
	WSS/BSS	1.3682	3.7963	<b>0.9530</b>
gauss	MSE	2.0995	3.1042	<b>0.5714</b>
	WSS/BSS	1.2756	4.0023	<b>0.4302</b>
bezdekiris	MSE	3.6458	2.7734	<b>1.2735</b>
	WSS/BSS	1.4322	1.0250	<b>0.5663</b>

with

$$\tilde{L}_{jk}(t) = \frac{L_j(t) c_{jk}^{(i)} T(x_t | \mu_{jk}^{(i)}, \sum_{jk}^{(i)}, v_{jk}^{(i)})}{\sum_{k=1}^{K-1} c_{jk}^{(i)} T(x_t | \mu_{jk}^{(i)}, \sum_{jk}^{(i)}, v_{jk}^{(i)})} \quad (34)$$

#### IV. EXPERIMENT

In the experimental section, this work uses a public data set and a real-world example to illustrate the advantages of our model. **Example 1. Performance testing of DSMM**

Student distribution is an infinite approximation of Gaussian distribution, which is very suitable for the identification of outlier data. This work has intentionally added some noise and outlier points in some benchmark data to test DSMM as shown in Table I.

The clustering quality of mixture models is evaluated through WSS and BSS, where WSS stands for Within Sum of Squares and BSS stands for Between Sum of Squares. For the results from clustering, smaller WSS and larger BSS imply better classification performance. Thus, this work proposes an index WSS/BSS. The smaller the WSS/BSS, the better the clustering ability of the model.

As shown in Table I, five benchmark data sets from UCI machine learning repository [50] are used to test our model, WSS/BSS of our model is the best among three mixture models. MSE of our model in these benchmark data sets is the best. Specially, in dataset iris and gauss, the performance provided by SMM is weaker compared with that of GMM, while that of DSMM is the best. Maybe the components of these two datasets contain more Gaussian distribution. Note that GMM is more suitable for the detection and identification of components with Gaussian distribution. The DSMM provides a dynamic overlap recognition mechanism for the continuous cluster. It is able detect the similarity between cluster space and present more classification accuracy. From these tested results, the proposed DSMM model has the ability to deal with the identification of latent variables (or stages).

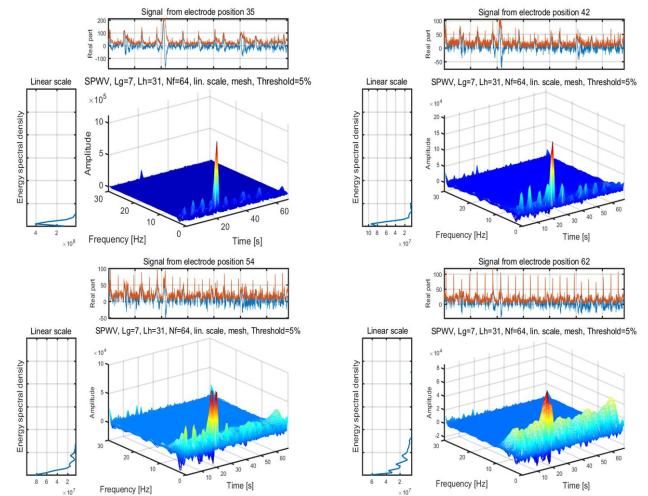


Fig. 6. Time-frequency distribution from SPWVD.

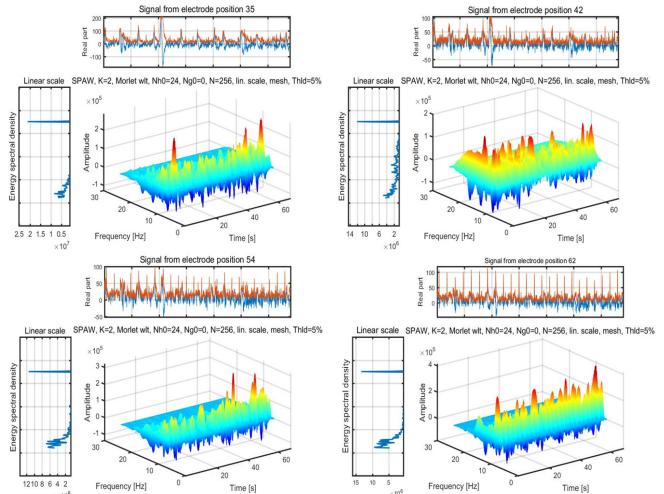


Fig. 7. Time-frequency distribution from SPAWVD.

#### Example 2. Pilot fatigue detection

##### A. Experiment Setup

The used flight dynamics model was the C919 commercial aircraft, as shown in Fig. 5. The pilots' EEG signals were recorded by the 64-channel EEG system during the flight until they were tired. The signal is sampled at 160 Hz. HDP-t-HSMM is implemented in the MATLAB 2016a environment. To prove its validity, a 5-fold cross-validation was used to optimize the parameters of the test model. More experimental details can refer to [53].

##### B. Cognitive Status Inference

The EEG signal usually contains four main rhythms. Here, we use a smoothed pseudo-affine Wigner-Ville distribution (SPAWVD) [46] and smoothed pseudo Wigner-Ville distribution (SPWVD) to process rhythm signals and present their time-frequency distributions, as shown in Figs. 6 and 7. It is clear that the three-dimensional feature distribution of Fig. 6 is not very prominent, while the three-dimensional feature distribution of Fig. 7 is more prominent, which shows local multi-peak phenomenon. Obviously, more local fine features are

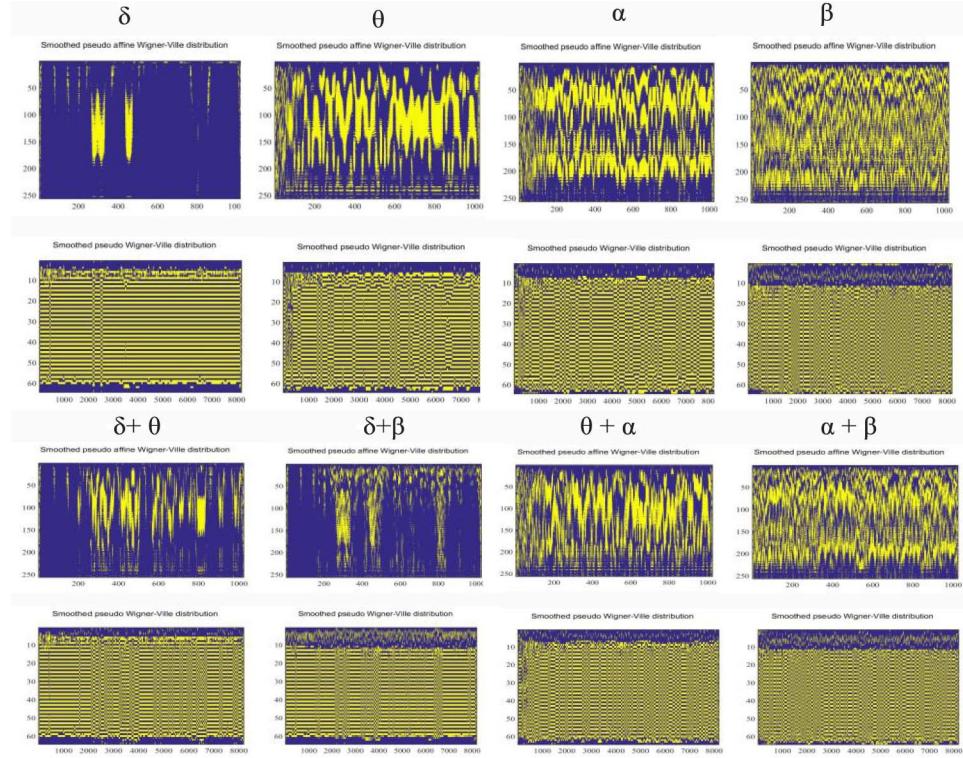


Fig. 8. Phase angle spectrum of four rhythms and their combination decomposed by SPAWVD and SPWVD, up: SPAWVD, bottom: SPWVD.

captured by SPAWVD. This is also the advantage of using affine technology.

Previous studies have found that changes of four rhythms can express the activity of human brain [1], [3], [5]. For example,  $\delta$  and  $\theta$  activity increase, while  $\beta$  decreases during mental fatigue [3]. Their power spectrum is also developed into four indicators, namely,  $(\alpha + \theta) / \beta$ ,  $(\theta + \delta) / (\alpha + \beta)$ ,  $\alpha / \beta$  and  $\theta / \beta$ . The occurrence and growth of fatigue can be detected by them. There are some conclusions are listed as follows: 1) the decrease of  $\alpha$  during fatigue is less than that of  $\beta$  [5]; 2) when  $\alpha$  and  $\beta$  activities decrease,  $(\alpha + \theta) / \beta$  and  $\alpha / \beta$  increase [6]; 3) the four indicators increase after long drive [7]. The indicators above are all based on the traditional FFT. FFT can only provide the average spectrum for a given period of time, and cannot provide instantaneous spectrum information. SPAWVD can provide the time-frequency information of EEG signal. Therefore, the SPAWVD is used as input to the brain activity indicators. This work also expanded the original four indicators to eight indicators as follows:

$$\left\{ \begin{array}{l} (p_\alpha(\omega) + p_\theta(\omega)) / p_\beta(\omega) \\ p_\alpha(\omega) / p_\beta(\omega) \\ (p_\alpha(\omega) + p_\theta(\omega)) / (p_\alpha(\omega) + p_\beta(\omega)) \\ p_\theta(\omega) / p_\beta(\omega) \\ (p_\delta(\omega) + p_\theta(\omega)) / (p_\alpha(\omega) + p_\beta(\omega)) \\ (p_\alpha(\omega) + p_\theta(\omega)) / (p_\delta(\omega) + p_\beta(\omega)) \\ (p_\delta(\omega) + p_\theta(\omega)) / (p_\delta(\omega) + p_\beta(\omega)) \\ (p_\delta(\omega) + p_\alpha(\omega)) / (p_\delta(\omega) + p_\beta(\omega)) \end{array} \right. \quad (35)$$

where  $P_s(\omega)$  represents the magnitude of the power spectral density of the SPAWVD from the waves.

SPAWVD and SPWVD are used to extract the instantaneous amplitude and frequency of the rhythm signal. Their phase angle spectrum is shown in Fig. 8. Below each rhythm symbol, two phase angle spectrums are derived from SPAWVD and SPWVD. Clearly, SPAWVD has more sparse capability than SPWVD to present the phase characteristics of the time-frequency distribution. Both  $(\alpha + \theta)$  and  $(\alpha + \beta)$  have clearer phase characteristics than the original  $\alpha$  and  $\beta$ . Other composite indicators also show this phenomenon. Therefore, through the combination of a single rhythm and a compound rhythm, the defined eight indicators contain more features than the original four indicators. All extracted spectral information contains some basic psychological stages such as waking (non-fatigue), mental disorder (micro fatigue) and drowsiness (extreme fatigue). These indicators may be highly correlated, and make it difficult to use them to detect brain cognitive behavior. It is necessary to remove some irrelevant features through a Treelet transform [44]. It is ideal as a dimension reduction and feature selection tool. It not only constructs clusters or groups of variables, but also constructs the functions of the data. At the top of Fig. 9, the classification basis vectors are shown via its energy matrix. Smaller sparse base units are used to represent its spectrum information. In the middle of Fig. 9, the components of (35) are extracted by Treelet transform. It can be seen that the linear independence of the two maximum energy trees on the left is better than that on the right. The corresponding feature distribution is more prominent, as shown in the bottom of Fig. 9.

The EEG signals from each channel in the 10-20 EEG signal acquisition system are different. Therefore, when researchers study human psychological behavior, the hierarchical model can provide a parameter sharing mechanism of behavioral

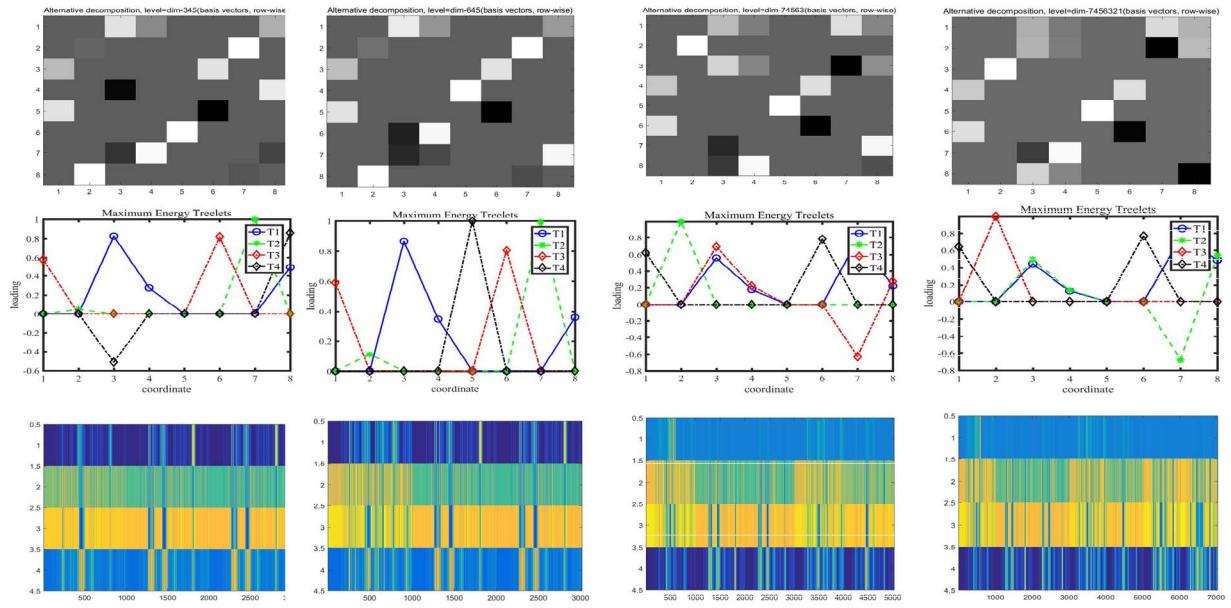
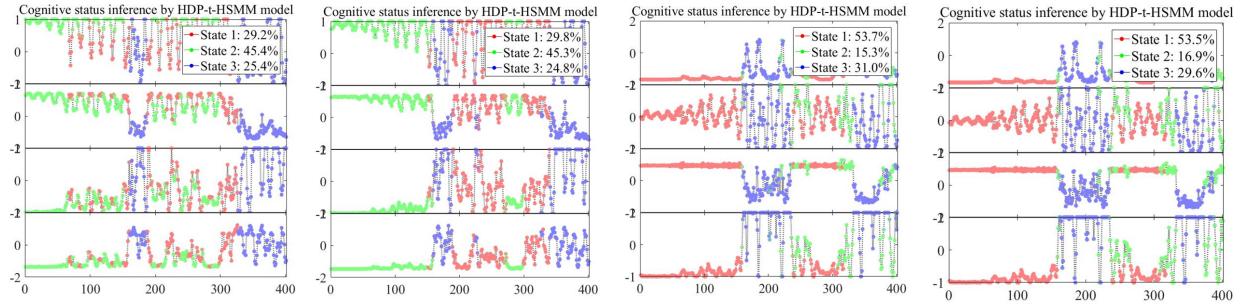


Fig. 9. Features represent and reduction, Top: energy matrix of eight features from Eq. (12); Middle: maximum energy treelet; Bottom: four prominent features provided by treelet.



modeling between the whole brain and each EEG signal. Therefore, this work proposes an HDP-t-HSMM to detect the brain cognitive activity. The cognitive inference results from 4 pilots of them are shown in Fig. 10. The four curves of each subgraph in Fig. 10 respectively represent the hidden state distribution of the four indicators. In the two pictures on the left, the fatigue cognitive characteristics of the two pilots are relatively similar. State 2 is the latent state at the start time, which represents the normal state, the latent part is state 1, which represents the micro fatigue state, and more outliers are represented as state 3, which represents the fatigue state. In the two figures on the right, state 1 represents the state of slight fatigue, state 2 represents the normal state, and state 3 represents the fatigue state.

Due to adequate rest for each pilot, the cognitive degree is distinct with three states. State transition and emission probability of the corresponding four models are shown in Fig. 11. It is clear that pilots in state 1 are the most reliable. Pilots with state 2 maybe face mental confusion (micro-fatigue) state. Pilots with state 3 feel drowsiness (extreme-fatigue). Their brain activity is sudden and drastic. One possible reason is

that the pilot is facing complicated flight operation, which is the phenomenon of over-consumption of brain energy. Flight management system need monitor this state to prevent the improper handling of pilot. This need impose additional stimuli means such as vibration from seat to help pilots get more excitements. It is very useful to air force pilots.

Brain cognitive inference of one pilot through HDP-t-HSMM with different outlier distribution is shown in Table II, and their average detection precision from all pilots is shown in Table III. Table II presents the detection results of cognitive behavior under different numbers of freedom degrees. Numbers of freedom degrees is an important parameter of the student mixture model, which affects the generalization ability of the model. When the number of freedom degree is equal to 8, the maximum workload state is detected. When it is equal to 6, micro-fatigue state is recognized. When it is equal to 4, extreme-fatigue state is detected. These numbers of freedom degree are of practical value for pilots' monitoring. Table III shows the detection results from different HMM models for the detection of all pilots' cognitive level. State 1 has an average detection

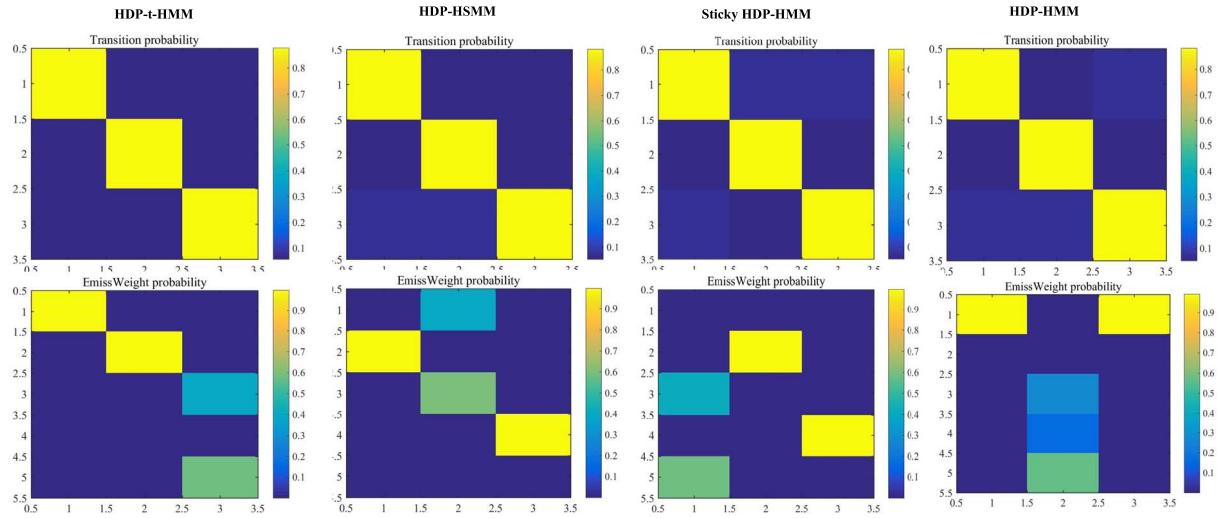


Fig. 11. State transition and emission probability of four models.

TABLE II  
BRAIN COGNITIVE INFERENCE OF ONE PILOT UNDER DIFFERENT OUTLIER DISTRIBUTION

Mental status	Number of degrees of freedom of the outlier distribution								
	1	2	3	4	5	6	7	8	9
State 1	46.5%	46.8%	47.2%	46.0%	46.0%	45.7%	45.4%	<b>51.8%</b>	45.0%
State 2	23.0%	23.1%	24.4%	23.1%	24.6%	<b>24.7%</b>	24.4%	22.3%	25.8%
State 3	30.5%	30.1%	28.4%	<b>30.9%</b>	29.4%	29.6%	29.2%	25.9%	29.2%

TABLE III  
THE AVERAGE DETECTION PRECISION OF ALL PILOTS

State	HDP-HMM	Sticky HDP-HMM	HDP-HSMM	HDP-t-HSMM
State 1	42.7%	46.8%	47.3%	<b>53.2%</b>
State 2	26.4%	24.6%	24.2%	<b>21.4%</b>
State 3	30.9%	28.6%	28.5%	<b>25.4%</b>

TABLE IV  
DETECTION ACCURACY OF DIFFERENT MODELS

Model	Data 1	Data 2	Data 3	Data 4	Data 5	Data 6	Data 7
k-means	0.681	0.724	0.785	0.736	0.650	0.692	0.706
HMM	0.815	0.825	0.862	0.814	0.815	0.825	0.805
HSMM	0.832	0.873	0.905	0.865	0.852	0.865	0.854
HDP-HMM	0.825	0.881	0.892	0.874	0.846	0.853	0.847
HDP-HSMM	0.840	0.914	0.930	0.895	<b>0.902</b>	0.883	0.871
HDP-t-HSMM	<b>0.872</b>	<b>0.935</b>	<b>0.952</b>	<b>0.926</b>	0.859	<b>0.904</b>	<b>0.916</b>

precision of 53.2% in our model. It indicates the proposed model has more capability to detect work state of pilots. For state 3, the proposed model presents an average detection precision of 25.4%. This indicates our model has strongest ability to the latent state detection of pilots, compared with other HMM models. The results indicate our model is excellent for the latent variables learning task like brain cognitive detection.

Above, this work provides an analysis of the latent state detection ability of each model for fatigue cognitive indicators. Next, this work divides the data collected in this mission

into 7 data sets to test the cognitive state recognition capabilities of our model and its peers, as shown in Table IV, as a simple and easy-to-use clustering algorithm, K-Means has a significant decrease in recognition accuracy compared with other algorithms in this experiment. Because HMM does not consider the residence time of the state, and result in rapid switching between states, its recognition accuracy is significantly lower than HSMM. On data sets 1, 3, 5, 6, and 7, the recognition accuracy of HSMM is higher than that of HDP-HMM. The recognition accuracy on all data sets of HDP-HSMM is higher than that of HDP-HMM. Except

for Data Set 5, the recognition accuracy of HDP-HSMM is higher than that of HDP-HMM. On all data sets, the recognition ability of HDP-HSMM is higher than HSMM, which verifies the advantage of HDP-HSMM to automatically infer the number of hidden states.

### C. Discussion

The current hidden Markov model faces two main problems: 1) the number of hidden states needs to be set, and 2) the state residence time is not considered. In order to solve them, this work defines a residual life HSMM, which models the residence time distribution and automatically infers the number of hidden states through hierarchical Dirichlet process. This work extracts some rhythm signals from EEG signals and construct fatigue features as input to the proposed model. Finally, through the hidden state detection experiment of brain cognitive indicators, the detection effects of different models are compared, and it is verified that the HDP-HSMM model performs well in the hidden state reasoning of brain cognitive indicators.

## V. CONCLUSION

This work addresses the problem of brain cognitive behavior inference and proposes a solution within the framework of HDP-*t*-HSMM. The advantages of this work can be summarized as follows:

The first advantage is that SPAWVD with Kaiser window function is used to extract the instantaneous spectrum features of major rhythms and their combinations.

The second advantage is that a HSMM with remaining service life is proposed to model the dynamic behavior of brain cognitive, and provide the latent states detection of rhythms or their combinations.

The third advantage is that it builds a Hierarchical multi-layer learning network with Dirichlet Process prior to detect brain cognitive states. HDP-*t*-HSMM is able to capture the dynamic relationship of brain cognitive behavior. SMM also can detect emission probability with more outliers.

A possible extension of this work is to study the use of other transient feature learning algorithms and dynamic HSMM modeling techniques such as the HDP-AR-HSMM model.

## REFERENCES

- [1] Y.-T. Liu, Y.-Y. Lin, S.-L. Wu, C.-H. Chuang, and C.-T. Lin, "Brain dynamics in predicting driving fatigue using a recurrent self-evolving fuzzy neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 2, pp. 347–360, Feb. 2016.
- [2] R. Bose, H. Wang, A. Dragomir, N. V. Thakor, A. Bezerianos, and J. Li, "Regression-based continuous driving fatigue estimation: Toward practical implementation," *IEEE Trans. Cognit. Develop. Syst.*, vol. 12, no. 2, pp. 323–331, Jun. 2020.
- [3] A. Subasi, "Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients," *Expert Syst. Appl.*, vol. 28, no. 4, pp. 701–711, May 2005.
- [4] R. Foong *et al.*, "Assessment of the efficacy of EEG-based MI-BCI with visual feedback and EEG correlates of mental fatigue for upper-limb stroke rehabilitation," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 3, pp. 786–795, Mar. 2020.
- [5] B. T. Jap, S. Lal, P. Fischer, and E. Bekiaris, "Using EEG spectral components to assess algorithms for detecting fatigue," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2352–2359, Mar. 2009.
- [6] H. J. Eoh, M. K. Chung, and S.-H. Kim, "Electroencephalographic study of drowsiness in simulated driving with sleep deprivation," *Int. J. Ind. Ergonom.*, vol. 35, no. 4, pp. 307–320, Apr. 2005.
- [7] W. Li, Q.-C. He, X.-M. Fan, and Z.-M. Fei, "Evaluation of driver fatigue on two channels of EEG data," *Neurosci. Lett.*, vol. 506, no. 2, pp. 235–239, Jan. 2012.
- [8] S. Ren, J. Li, F. Taya, J. de Souza, N. V. Thakor, and A. Bezerianos, "Dynamic functional segregation and integration in human brain network during complex tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 547–556, Jun. 2017.
- [9] Z. Gao *et al.*, "EEG-based spatio-temporal convolutional neural network for driver fatigue evaluation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2755–2763, Sep. 2019.
- [10] W. Sun, X. Zhang, S. Peeta, X. He, and Y. Li, "A real-time fatigue driving recognition method incorporating contextual features and two fusion levels," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 12, pp. 3408–3420, Dec. 2017.
- [11] A. Chaudhuri and A. Routray, "Driver fatigue detection through chaotic entropy analysis of cortical sources obtained from scalp EEG signals," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 185–198, Jan. 2020.
- [12] P. Zarjam, J. Epps, and N. H. Lovell, "Beyond subjective self-rating: EEG signal classification of cognitive workload," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 4, pp. 301–310, Dec. 2015.
- [13] R. Huang, Y. Wang, Z. Li, Z. Lei, and Y. Xu, "RF-DCM: Multi-granularity deep convolutional model based on feature recalibration and fusion for driver fatigue detection," *IEEE Trans. Intell. Transp. Syst.*, early access, Sep. 2, 2020, doi: [10.1109/TITS.2020.3017513](https://doi.org/10.1109/TITS.2020.3017513).
- [14] J. A. Blanco *et al.*, "Quantifying cognitive workload in simulated flight using passive, dry EEG measurements," *IEEE Trans. Cognit. Develop. Syst.*, vol. 10, no. 2, pp. 373–383, Jun. 2018.
- [15] B. Mandal, L. Li, G. S. Wang, and J. Lin, "Towards detection of bus driver fatigue based on robust visual analysis of eye state," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 545–557, Mar. 2017.
- [16] A.-M. Olteteau, Z. Falomir, and C. Freksa, "Artificial cognitive systems that can answer human creativity tests: An approach and two case studies," *IEEE Trans. Cognit. Develop. Syst.*, vol. 10, no. 2, pp. 469–475, Jun. 2018.
- [17] J. Li, Y. Wang, L. Zhang, A. Cichocki, and T.-P. Jung, "Decoding EEG in cognitive tasks with time-frequency and connectivity masks," *IEEE Trans. Cognit. Develop. Syst.*, vol. 8, no. 4, pp. 298–308, Dec. 2016.
- [18] C. Zhang, L. Sun, F. Cong, and T. Ristaniemi, "Spatio-temporal dynamical analysis of brain activity during mental fatigue process," *IEEE Trans. Cognit. Develop. Syst.*, early access, Feb. 27, 2020, doi: [10.1109/TCDS.2020.2976610](https://doi.org/10.1109/TCDS.2020.2976610).
- [19] P. Qi *et al.*, "EEG functional connectivity predicts individual behavioural impairment during mental fatigue," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 9, pp. 2080–2089, Sep. 2020.
- [20] C.-T. Lin, C.-H. Chuang, Y.-C. Hung, C.-N. Fang, D. Wu, and Y.-K. Wang, "A driving performance forecasting system based on brain dynamic state analysis using 4-D convolutional neural networks," *IEEE Trans. Cybern.*, early access, Aug. 20, 2020, doi: [10.1109/TCYB.2020.3010805](https://doi.org/10.1109/TCYB.2020.3010805).
- [21] S. Wang, J. Gwizdka, and W. A. Chaovallitwongse, "Using wireless EEG signals to assess memory workload in the *n*-back task," *IEEE Trans. Human-Machine Syst.*, vol. 46, no. 3, pp. 424–435, Jun. 2016.
- [22] D. Dvorak, A. Shang, S. Abdel-Baki, W. Suzuki, and A. A. Fenton, "Cognitive behavior classification from scalp EEG signals," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 729–739, Apr. 2018.
- [23] F.-Y. Wang, "Parallel control and management for intelligent transportation systems: Concepts, architectures, and applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 3, pp. 630–638, Sep. 2010.
- [24] T. Shen, J. Wang, C. Gou, and F.-Y. Wang, "Hierarchical fused model with deep learning and type-2 fuzzy learning for breast cancer diagnosis," *IEEE Trans. Fuzzy Syst.*, vol. 28, no. 12, pp. 3204–3218, Dec. 2020.
- [25] C.-T. Lin, S.-F. Tsai, and L.-W. Ko, "EEG-based learning system for online motion sickness level estimation in a dynamic vehicle environment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 10, pp. 1689–1700, Oct. 2013.
- [26] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc., B, Methodol.*, vol. 39, no. 1, pp. 1–38, 1977.
- [27] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical Dirichlet processes," *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1566–1581, Dec. 2006.

- [28] M. J. Beal, Z. Ghahramani, and C. E. Rasmussen, "The infinite hidden Markov model," in *Machine Learning*. Cambridge, MA, USA: MIT Press, 2002, pp. 29–245.
- [29] E. B. Fox, E. B. Sudderth, M. I. Jordan, and A. S. Willsky, "An HDP-HMM for systems with state persistence," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 312–319.
- [30] S. Yu, *General Hidden Semi-Markov Model*. Amsterdam, The Netherlands: Elsevier, 2016.
- [31] A. Bargi, R. Y. D. Xu, and M. Piccardi, "AdOn HDP-HMM: An adaptive online model for segmentation and classification of sequential data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 3953–3968, Sep. 2018.
- [32] A. H. H. N. Torbati and J. Picone, "A doubly hierarchical Dirichlet process hidden Markov model with a non-ergodic structure," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 1, pp. 174–184, Jan. 2016.
- [33] T. Fuse and K. Kamiya, "Statistical anomaly detection in human dynamics monitoring using a hierarchical Dirichlet process hidden Markov model," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 3083–3092, Nov. 2017.
- [34] T. Taniguchi, S. Nagasaka, K. Hitomi, K. Takenaka, and T. Bando, "Unsupervised hierarchical modeling of driving behavior and prediction of contextual changing points," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 4, pp. 1746–1760, Aug. 2015.
- [35] L. Du, M. Chen, J. Lucas, and L. Carin, "Sticky hidden Markov modeling of comparative genomic hybridization," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5353–5368, Oct. 2010.
- [36] E. Fox, E. Sudderth, M. Jordan, and A. Willsky, "Bayesian nonparametric methods for learning Markov switching processes," *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 43–54, Nov. 2010.
- [37] A. H. H. N. Torbati, J. Picone, and M. Sobel, "A left-to-right HDP-HMM with HDPM emissions," in *Proc. 48th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2014, pp. 1–6.
- [38] A. Bargi, R. Y. Da Xu, and M. Piccardi, "An online HDP-HMM for joint action segmentation and classification in motion capture data," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 1–7.
- [39] W. Hu, G. Tian, Y. Kang, C. Yuan, and S. Maybank, "Dual sticky hierarchical Dirichlet process hidden Markov model and its application to natural language description of motions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2355–2373, Oct. 2018.
- [40] D. B. Springer, L. Tarassenko, and G. D. Clifford, "Logistic regression-HSMM-based heart sound segmentation," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 822–832, Apr. 2016.
- [41] T. Liu, K. Zhu, and L. Zeng, "Diagnosis and prognosis of degradation process via hidden semi-Markov model," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 3, pp. 1456–1466, Jun. 2018.
- [42] H. Zhang, W. Ni, X. Li, and Y. Yang, "Modeling the heterogeneous duration of user interest in time-dependent recommendation: A hidden semi-Markov approach," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 48, no. 2, pp. 177–194, Feb. 2018.
- [43] S.-Z. Yu and H. Kobayashi, "Practical implementation of an efficient forward-backward algorithm for an explicit-duration hidden Markov model," *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1947–1951, May 2006.
- [44] A. B. Lee, B. Nadler, and L. Wasserman, "Treelets—An adaptive multi-scale basis for sparse unordered data," *Ann. Appl. Statist.*, vol. 2, no. 2, pp. 435–471, Jun. 2008.
- [45] G. Sikander and S. Anwar, "Driver fatigue detection systems: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2339–2352, Jun. 2019.
- [46] E. Q. Wu *et al.*, "Nonparametric Bayesian prior inducing deep network for automatic detection of cognitive status," *IEEE Trans. Cybern.*, early access, Mar. 20, 2020, doi: [10.1109/TCYB.2020.2977267](https://doi.org/10.1109/TCYB.2020.2977267).
- [47] L. He, D. Hu, M. Wan, Y. Wen, K. M. Deneen, and M. Zhou, "Common Bayesian network for classification of EEG-based multi-class motor imagery BCI," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 46, no. 6, pp. 843–854, Jun. 2016.
- [48] C. Liu, Y. Fu, J. Yang, X. Xiong, H. Sun, and Z. Yu, "Discrimination of motor imagery patterns by electroencephalogram phase synchronization combined with frequency band energy," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 3, pp. 551–557, 2017.
- [49] A. Rakshit, A. Konar, and A. K. Nagar, "A hybrid brain-computer interface for closed-loop position control of a robot arm," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 5, pp. 1344–1360, Sep. 2020.
- [50] E. Q. Wu *et al.*, "Self-paced dynamic infinite mixture model for fatigue evaluation of pilots' brains," *IEEE Trans. Cybern.*, early access, Dec. 7, 2020, doi: [10.1109/TCYB.2020.3033005](https://doi.org/10.1109/TCYB.2020.3033005).
- [51] X. Li and Z. Wang, "A HMM-based mandarin Chinese singing voice synthesis system," *IEEE/CAA J. Autom. Sinica*, vol. 3, no. 2, pp. 192–202, Apr. 2016.
- [52] W. Dong and M. Zhou, "Gaussian classifier-based evolutionary strategy for multimodal optimization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 6, pp. 1200–1216, Jun. 2014.
- [53] E. Q. Wu *et al.*, "Novel nonlinear approach for real-time fatigue EEG data: An infinitely warped model of weighted permutation entropy," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 6, pp. 2437–2448, Jun. 2020.

**Edmond Q. Wu** (Member, IEEE) received the Ph.D. degree in control theory and application from Southeast University, Nanjing, China, in 2009. He is currently an Associate Professor with the Key Laboratory of System Control and Information Processing, Ministry of Education, Shanghai Jiao Tong University, Shanghai, China. His research interests include deep learning, cognitive modeling, and industrial information processing.

**Li-Min Zhu** (Member, IEEE) received the B.E. (Hons.) and Ph.D. degrees from Southeast University, Nanjing, China, in 1994 and 1999, respectively, all in mechanical engineering. He is currently the "Cheung Kong" Chair Professor, the Head of the Department of Mechanical Engineering, and the Vice Director of the State Key Laboratory of Mechanical System and Vibration. His research interests include mechanical signature analysis for machine and manufacturing process monitoring, CNC machining and 3D measurement of complex shaped parts, and control, sensing, and instrumentation for micro/nano manufacturing. He has been an Associate Editor of the *IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING*. He is also a Technical Editor of the *IEEE/ASME TRANSACTIONS ON MECHATRONICS*, and the Editorial Board Member of the *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* and the *International Journal of Intelligent Robotics and Applications*.

**Gui-Jiang Li** is currently a Professor with The First Aircraft Institute, Xi'an, China. He is involved in equipment reliability, and health management of military aircraft. He is also a member of the Airborne Autotest Expert Group.

**Hong-Jun Li** is currently with First Army Representative Office, Xi'an. His research interests include signal and information processing, and avionics systems.

**Zhiri Tang** (Student Member, IEEE) received the B.Sc. and M.Eng. degrees from Wuhan University, Wuhan, China, in 2017 and 2019, respectively. He is currently pursuing the Ph.D. degree with the Department of Computer Science, City University of Hong Kong. His research interests include neural computing and computer vision.

**Ruihan Hu** received the Ph.D. degree from the School of Physics and Technology, Wuhan University. He is currently with the Guangdong Key Laboratory of Modern Control Technology, Guangdong Institute of Intelligent Manufacturing, Guangzhou, China. His research interests include spiking neural networks and multimodal learning.

**Gui-Rong Zhou** graduated from Northwest Polytechnic University in 1981. He was a Deputy Designer of the Institute of Helicopter Design of Aviation Industry. He is currently a Professor with the COMAC Shanghai Aircraft Design and Research Institute as a Deputy Chief Engineer on the C-919. He participated in the research and development of several models, such as Zhi-9 and Zhi-10, and enjoyed the special allowance of the Government of the State Council. In 2008, he was a Deputy Designer of C919 large passenger aircraft model. His research interests include the design of avionics systems, and prognosis and health management.