Chapter 4

Bayesian Sequential Learning for EEG-Based BCI Classification Problems

S. Shigezumi, H. Hara, H. Namba, C. Serizawa,
Y. Dobashi, A. Takemoto, K. Nakamura and
T. Matsumoto

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/56146

1. Introduction

Non-invasive Brain-Computer Interfaces (BCIs) have been an active research area where several different methods have been developed. They include Electroencephalography (EEG), Near-infrared Spectroscopy (NIRS), functional-MRI (fMRI) among others [1]. Of those BCIs, EEG is one of the most studied methods. This is mainly due to its fine temporal resolution, ease of use, and relatively low set-up cost. Each BCI method naturally has its own advantages and disadvantages. EEG is no exception.

One of the main disadvantages of an EEG-based BCI is its susceptibility to noise, which motivates development of a variety of machine learning algorithms for decoding EEG signals, and there have been significant advancements in the area [2].

One way of categorizing machine-learning algorithms for BCI is **batch** mode and **sequential (online)** mode. In the batch mode learning, the collectively acquired EEG data from a subject is divided into two subsets: training data and test data. The former is used for training the machine-learning algorithm, whereas the latter is used to evaluate the algorithm's capability to predict the subject's intention [2]-[4]. There are several facets in batch mode learning which call for improvements:

1. First, it is non-trivial to decide how much data should be used for training and how much data should be left for testing. It should also be noted that the number of necessary training data may depend on each subject.

2. Second, with the batch mode learning, by definition, one cannot perform sequential evaluations of predictive performance as time evolves.

3. Third, the batch mode learning presumes that the data is stationary, i.e., the subject's physical condition and/or the environment around the subject does not change over the period of the experiment.

In contrast, the sequential learning algorithm considered in this study starts learning with the very first single trial datum and proceeds with the learning each time a single trial datum arrives within a Bayesian framework.

This paper proposes a Bayesian sequential learning algorithm for steady-state visual evoked potential (SSVEP) classification problems in a principled manner. In particular, the paper performs the following:

a. Evaluation of the *sequential posterior distribution* of unknown parameters each time a trial is performed.

b. Computation of the *sequential predictive distribution* of the class label at each trial based on the posterior distribution obtained above.

c. *Automatic hyperparameter learning,* where hyperparameter in this study corresponds to the search region volume in the unknown parameter space.

d. Sequential evaluation of the error between the true label and the predicted label.

e. Sequential evaluation of *marginal likelihood* which quantifies the reliability of the prediction at each trial.

f. Experiments are performed on a four class problem in addition to two class problem, where the extension from the latter to the former is nontrivial.

g. Formulate the problem using nonlinear model to capture potential nonlinearities which can be easily extended to more difficult problems.

2. Related work

There are three ingredients in this study: (i) SSVEP, (ii) Sequential (Online) learning, and (iii) Sequential Monte Carlo implementations. The descriptions that follow will be given in terms of these keywords. For the batch mode learning, we cite the survey paper reported in [2] instead of citing individual papers.

Allison et al. [5] performed a demographic study of several different BCI methods and showed that an SSVEP-based BCI spelling system was competitive for different age groups, as well as different gender groups, with little experience, under noisy environments. It is also reported that most subjects stated that they did not consider the flickering stimuli annoying and would use or recommend such a BCI system. In [6], and also in [7], an SSVEP-based orthosis control system with an LED light source is proposed. The flickering frequencies were 6 Hz, 7 Hz, 8

http://dx.doi.org/10.5772/56146

Hz, and 13 Hz in the former, and 8 Hz and 13 Hz in the latter. Classification was performed using the second- and third-order higher harmonics in addition to the fundamental frequency component. In [8], an SSVEP-based speller is proposed. After Principal Component Analysis (PCA), the probability of each frequency component is estimated using a particular information matrix. The speller introduces a selection based on a decision tree and an undo command for correcting eventual errors. In [9], EEG signals are represented in the spatial-spectral-temporal domain by a wavelet transform. It also uses a multi-linear discriminative subspace by employing general tensor discriminant analysis (GTDA). The classification is conducted by support vector machine (SVM). Reference [10] proposes a biphasic stimulation technique to solve the issue of phase drifts in SSVEP in phase-tagged systems. The Kalman filter is used in [11] to decode neural activity and estimate the desired movement kinematics, where the filter gain is approximated by its steady-state form, which is computed offline before real-time decoding commences. Canonical correlation analysis (CCA) is used in [12] to analyze the frequency components of SSVEP, where the correlations between the target oscillation waveforms, as well as their higher order harmonics, and those of the acquired SSVEP waveforms are calculated. It is demonstrated that the scheme performed better than a fast Fourier transform-based spectrum estimation method. CCA is used in an online manner by updating the parameters each time data arrives. An online learning scheme called Stochastic Meta Descent (SMD), which is a generalization of the gradient descent algorithm, is proposed in [13]. The paper also discusses various aspects of errors incurred in online learning algorithms.

The Subject Specific Classification Model is discussed in [14], where model Gaussian parameters are updated online after an initial learning of the Subject Independent Classification Model from a pool of subjects. The data was taken from P300, which is one component of EEG.

Martinez et al. [4] propose an SSVEP-based online BCI system with visual neurofeedback. The algorithm is different from the one proposed in this paper. There is a report on Bayesian sequential learning for EEG signals [6], where the Sequential Monte Carlo is used for implementation. In addition to the task differences between [6] and this study, there are several algorithmic distinctions between the two. First, the basis function used in the former is linear with respect to the associated parameters, whereas the latter uses a basis function in which parameters appear in a nonlinear manner. Second, the parameter that controls the size of the parameter search region is fixed in the former, whereas it is learned automatically in the latter. These two differences, at least within our experience, are important for achieving better performance.

In addition, no extension for multi-class classification problems was performed in the former, whereas both algorithmic extension and a learning experiment by using multi-class real EEG data are performed in the latter. Our proposed algorithm is based on part of earlier work [16] of the authors' research group for a different application. Preliminary results on a two-class classification problem were reported by the present group in [17]. The current paper gives a full account of the results by expanding several parts of that conference paper. First, a four-class classification problem was formulated and tested experimentally. Second, the algorithm was further improved by incorporating an Effective Sample Size and Rao-Blackwellisation. Third, more detailed discussions are added.

3. Subjects and data acquisition

Of Event Related Potentials used in BCI, the target quantities considered in this study are SSVEPs, which are natural responses to visual stimulation at specific frequencies. These frequency components and their higher-order harmonics can be observed in the occipital region [4]. It is known that SSVEPs are often useful in research because of the reasonable signal-to-noise ratio and relative immunity to artifacts [5].

In an attempt to perform two-class and four-class classification problems, we gathered two sets of SSVEP data. The settings that we describe below for the two experiments are the same except for the number of stimuli and their frequencies. EEG data were recorded by a Polymate (Nihon Koden, Tokyo) with six active electrodes (O1, OZ, O2, O9, IZ, O10) according to the international 10-10 system and referenced to the left earlobe with a digitization rate of 500 Hz. Even though the highest flickering frequency was 10 Hz, we considered second and third order harmonics in one of the experiments reported below. Our original intention was to examine the harmonics higher than three even though they were not reported in this paper. In order to have a wide margin for the Nyquist frequency we chose 500 Hz. Five volunteers (aged 21-23 years) participated in the present study. All subjects were healthy, with no past history of psychiatric or neurological disorders.

Written informed consent was obtained from each subject on forms approved by the ethical committee of Waseda University.

Each subject was seated in a comfortable chair 60 cm in front of the monitor in an electrically shielded and dimmed room. The flow of task events is shown in Figure 1. The stimulus for the two-class problem is illustrated in Figure 2, whereas that of the four-class problem is illustrated in Figure 3.

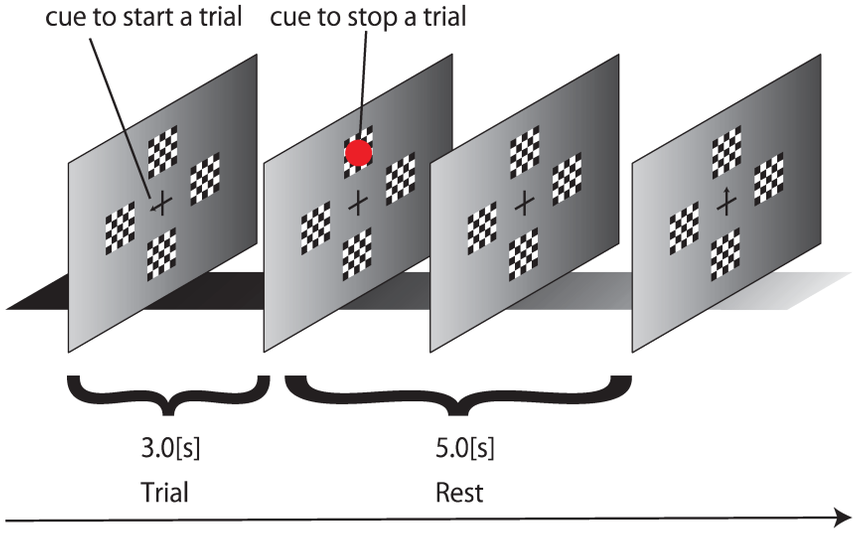

Figure 1. Task flow for the four-class classification problem

http://dx.doi.org/10.5772/56146

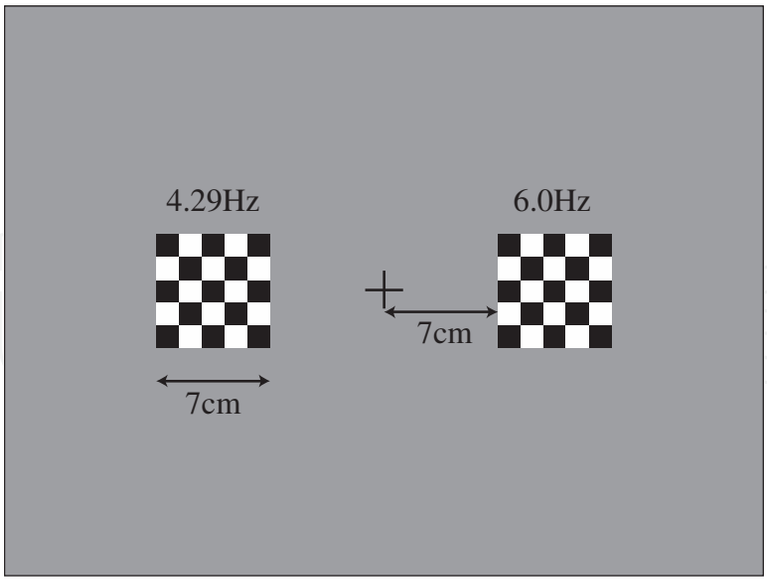

Figure 2. Monitor display for the two class classification problem

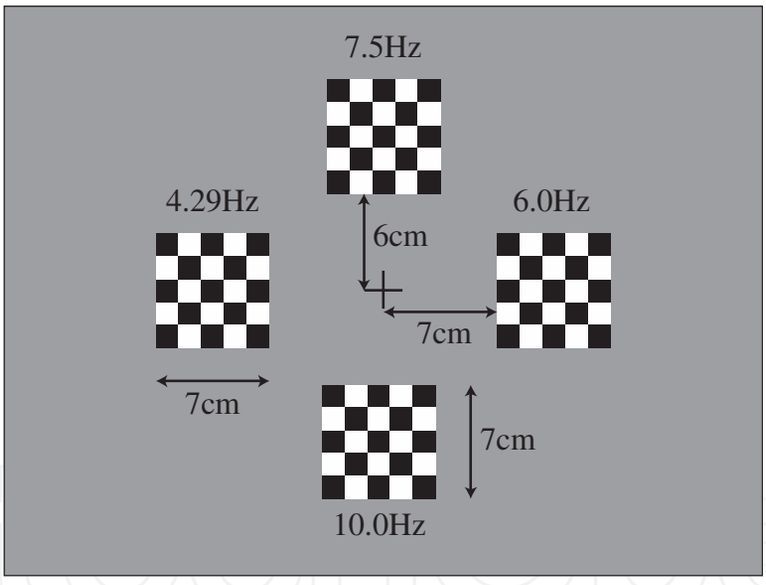

Figure 3. Monitor display of the four-class classification problem

In the two-class problem, two flickering checkerboard stimuli (left and right) were presented on the monitor, whereas in the four-class problem, four flickering checkerboard stimuli (left, right, top, and bottom) were presented. In addition, a fixation cross was placed at the center, which the subject was usually asked to fixate at.

In the two-class problem, the left stimulus was a checkerboard flickering with frequency of 4.29 Hz, whereas the right stimulus flickered at frequency of 6.00 Hz. In the four-class problem, there were additional stimuli, one at the top with frequency of 7.50 Hz, and one at the bottom

with frequency of 10.0 Hz. There were three reasons for selecting these particular frequencies. First, it is known that SSVEPs are discernible in approximately the 4.0 Hz - 50 Hz band [18][19]. Second, higher harmonics of a particular frequency component should not overlap with the fundamental frequency component. Such overlap could give rise to a problem when one considers multi-class classification problem where multiple frequencies are involved as is described in section **6**. Third, since the monitor refresh rate is 60 Hz, choice of the flickering frequencies were restricted by 60/positive integer. We chose 60/14=4.2857…. which we approximated by 4. 29.

The subject usually fixated at a central fixation cross. When an arrow replaced the cross, the subject should move his or her eyes to the checkerboard indicated by the arrow for 3.0 s, after which a red circle is shown so that the subject would know when to rest for 5.0 s. This sequence was one trial, and trials were repeated twenty times, constituting one session. The direction of the arrow was selected at random. Each subject completed 600 trials, or 30 sessions. The measurements were performed with a Polymate AP1124, a multi-purpose portable bio-amplifier recording device, manufactured by TEAC Corporation, Tokyo, Japan. The device is equipped with 24 channels with a maximum sampling frequency of 1 kHz. In addition to electroencephalograms (EEGs), eyeball movement and other external signals can be measured. The dimensions are W90 mm x H 44 mm x D 158 mm, the weight is 300 g, and the device is powered by battery.

4. Algorithm

This section gives a description of the proposed sequential learning algorithm. It consists of several aspects: (i) the basis function to fit the data, (ii) the likelihood function, (iii) sequential parameter learning, and (iv) sequential hyperparameter learning. The actual predictive values are given by the predictive distribution of the target class labels, which will be described in **4.3**. In order to improve the learning capabilities, Rao-Blackwellisation will also be described. We begin with a two-class classification problem followed by a multi-class classification problem. A schematic diagram of the proposed algorithm is given in Figure 4.

4.1. Two-class classification problem

Let $x_k \in R^d$ be the feature vector at the k-th trial, where d represents the dimension of x_k which, in our paper, is the DFT spectrum of a single trial EEG. Let $y_k \in \{0,1\}$ be the binary class label of each trial, where 0 corresponds to the right flickering image and 1 corresponds to the left flickering image. Our purpose here is to learn parameters associated with the basis function, to be defined shortly, and predict the subject's intention given SSVEP data, each time datum arrives.

4.1.1. Basis function and classifier

Consider the parameterized family of nonlinear basis functions $f(\bullet)$ defined by:

http://dx.doi.org/10.5772/56146

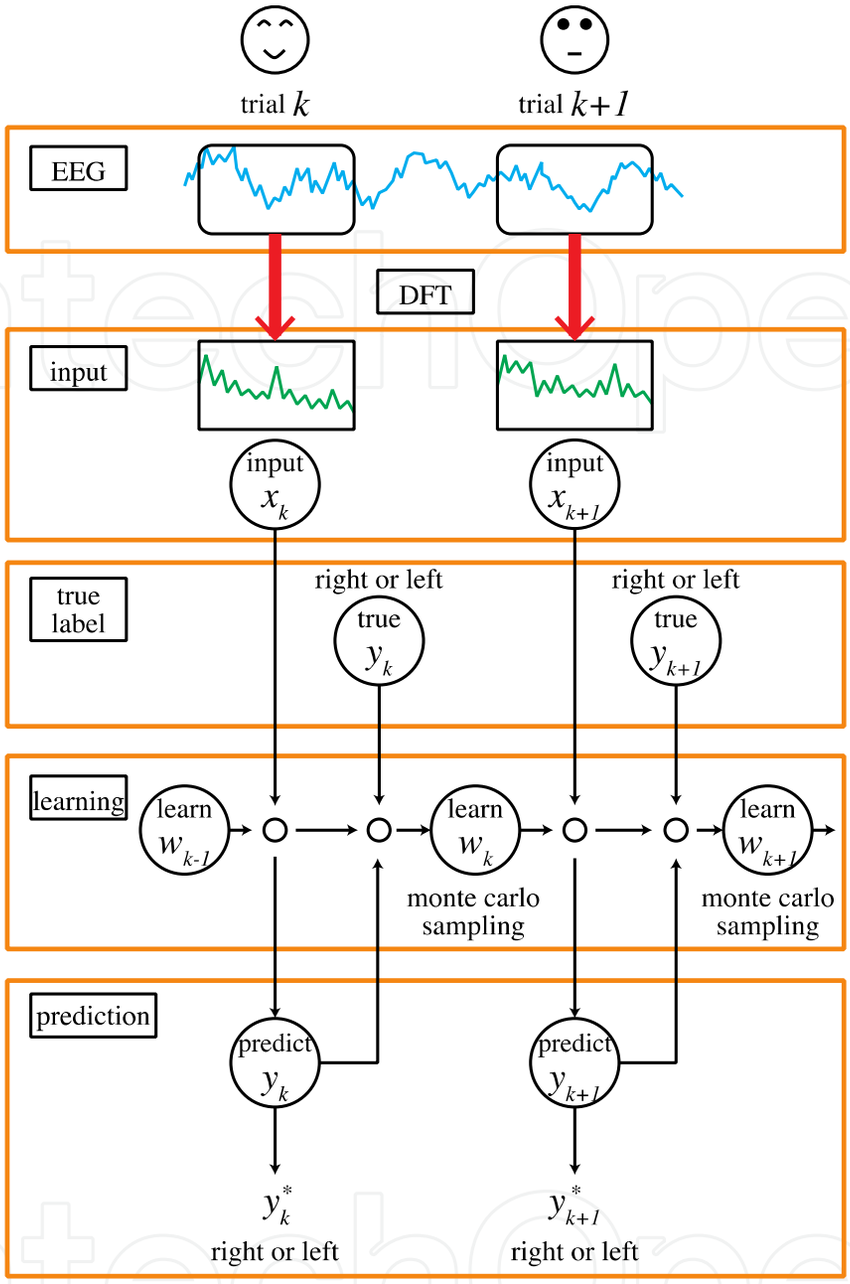

Figure 4. Schematic diagram of the proposed algorithm

$$f(x_k;\omega_k)=\sum_{j=1}^{h} v_{k,j}\sigma\left(\sum_{i=1}^{d} u_{k,ij}x_{k,i}+u_{k,0j}\right)+v_{k,0}, \tag{1}$$

where $u_k:=(u_{k,0},\ \cdots,\ u_{k,d})^T\in R^{h(d+1)}$, $u_{k,i}:=(u_{k,i1},\ \cdots,\ u_{k,ih})^T\in R^h$, $v_k:=(v_{k,0},\ \cdots,\ v_{k,h})^T\in R^{h+1}$, $\omega_k=(u_k,\ v_k)$.

The function $\sigma(\bullet)$ is a sigmoidal function defined by $\sigma(a)=\frac{1}{1+\exp(-a)}$, where h represents the number of hidden units.

Other popular basis functions often work as well. It should be noted that this basis function is nonlinear with respect to u_k as well as x_k, which enables capturing of potential nonlinearities involved.

In order to associate the quantity defined by (1) with the class label, consider:

$$P(y_k \mid x_k, \omega_k) := Be(y_k; \Phi(f(x_k; \omega_k))), \tag{2}$$

whereΦ is a function which monotonically maps the real numbers onto [0,1]. Several choices of Φ are possible. One is:

$$\Phi(u) := \frac{1}{1 + \exp(-u)}, \tag{3}$$

while another is:

$$\Phi(u) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} \exp(-a^2/2) da. \tag{4}$$

We tested both functions and found them to work equally well for our SSVEP learning. In what follows, we will report our results with (4) by introducing a latent variable z_k and considering

$$P(y_k \mid x_k, \omega_k) = \int P(y_k, z_k \mid x_k, \omega_k) dz_k, \tag{5}$$

$$P(y_k, z_k \mid x_k, \omega_k) = P(y_k \mid z_k) P(z_k \mid x_k, \omega_k), \tag{6}$$

$$P(y_k \mid z_k) := Be(I(z_k \geq 0)), \tag{7}$$

$$P(z_k \mid x_k, \omega_k) := N(z_k; f(x_k; \omega_k), 1.0), \tag{8}$$

where $I(A)$ represents an indicator function defined as 1 when A is *true* and 0 when A is *false*.

4.1.2. Parameter search stochastic dynamics

In order to perform sequential learning, we perform a sequential stochastic search of the parameter ω_k each time trial data is acquired:

$$P(\omega_k \mid \omega_{k-1}, \gamma_k) := \frac{1}{Z_\omega(\gamma_k)} \exp\left(-\frac{\gamma_k \left\| \omega_k - \omega_{k-1} \right\|^2}{2}\right) \tag{9}$$

where Z_ω represents the normalization constant. This amounts to searching for a new value ω_k based on the previous value ω_{k-1}, but in a random walk manner. This is a first-order Markov process, so that the parameters of the distant past are naturally forgotten because of the noise,

http://dx.doi.org/10.5772/56146

whereas the parameters of the immediate past tend to be taken into account with higher weights. This stochastic parameter search is reflected in the posterior distributions (20) given sequential data. Since this transition probability is Gaussian, it involves γ_k, which is the reciprocal of the variance parameter. More specifically, if γ_k is small, the parameter search region for ω_k will be large, whereas if γ_k is large, the search region will be small.

4.1.3. Automatic hyperparameter search stochastic dynamics

Our experiences tell us that automatic adjustment of γ_k is often important in order to achieve better performance. γ_k is often called a hyperparameter since it controls the behavior of the target parameter ω_k. We perform the following automatic stochastic search of γ_k :

$$P(\gamma_k \mid \gamma_{k-1}) := \frac{1}{\sqrt{2\pi}\gamma_k \sigma_h} \exp\left(-\frac{(\log \gamma_k - \log \gamma_{k-1})^2}{2\sigma_h{}^2}\right), \tag{10}$$

There are at least two reasons for this transition probability to be log-normal. One is that γ_k needs to be positive, and another is to cover a large range of values in the hyperparameter space. It should be noted that (10) is also a first-order Markov process, so that the hyperparameters of the immediate past are taken into account, whereas the hyperparameters of the distant past tend to be forgotten. In order to explain the importance of such hyperparameter learning, consider Figure 5. Letting $\theta_k := (\omega_k, \gamma_k)$ the blue region represents the likelihood function landscape in the θ-space, where the darker the blue color, the higher the likelihood function value. The white diamonds represent samples $\theta_{t-1}^{(i)} \sim P(\theta_{t-1} \mid x_{1:t-1}, y_{1:t-1})$, the yellow diamonds $\theta_t^{*(i)} \sim P(\theta_t^{(i)} \mid \theta_{t-1}, \gamma_t^{(i)})$, $\gamma_t^{(i)}$:*large*, and the light-green diamonds $\theta_t^{*(i)} \sim P(\theta_t^{(i)} \mid \theta_{t-1}, \gamma_t^{(i)})$, $\gamma_t^{(i)}$:*small*. Now suppose that the likelihood function landscape in the θ -space changed by a relatively large amount, as shown by the pink region, where the darker the color, the higher the likelihood. The yellow diamonds are scarce in the pink region, so that it is difficult to find θ samples that give rise to meaningful likelihood function values. This is due to the fact that $\gamma_t^{(i)}$ is large, so that the search region is restricted. If $\gamma_t^{(i)}$ is relatively small, on the other hand, then the green θ samples might capture at least a part of the pink region where the likelihood function values are meaningful. The proposed hyperparameter learning scheme automatically learns appropriate $\gamma_t^{(i)}$ values from the sequential data and lets the algorithm find reasonable θ samples.

4.1.4. Rao-blackwellised SMC

In this paper, we implemented not only the standard SMC but also Rao-Blackwellised SMC (RBSMC) for the purpose of performance improvement. Rao-Blackwellisation is a statistical variance reduction strategy for the Monte Carlo method [25], [26]. It is a combination of analytical integration (marginalization) and the Monte Carlo method. In order to explain this,

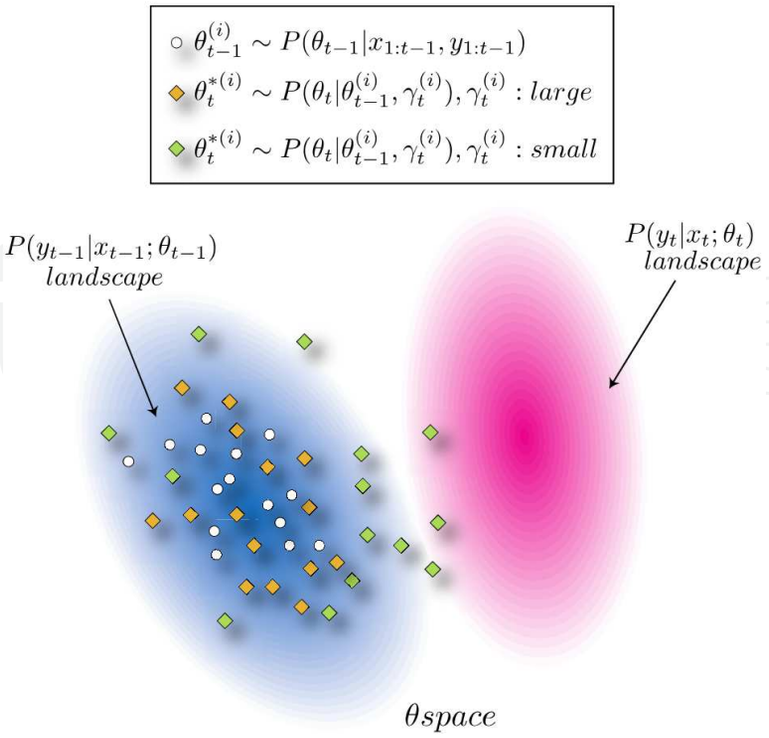

Figure 5. The proposed hyperparameter learning scheme automatically finds the appropriate region in the θ-space. The blue region indicates the likelihood function landscape at time t - 1, whereas the pink region indicates the likelihood function landscape at time t. The darker the color, the higher the likelihood function value. The white diamonds represent samples $\theta_{t\text{-}1}^{(j)} \sim P\left(\theta_{t\text{-}1} \middle| x_{1:t\text{-}1}, y_{1:t\text{-}1}\right)$, the yellow diamonds $\theta_t^{*(j)} \sim P\left(\theta_t^{(j)} \mid \theta_{t\text{-}1}, \gamma_t^{(j)}\right)$, $\gamma_t^{(j)}$:*large*, and the light-green diamonds $\theta_t^{*(j)} \sim P\left(\theta_t^{(j)} \mid \theta_{t\text{-}1}, \gamma_t^{(j)}\right)$, $\gamma_t^{(j)}$:*small*. The proposed scheme automatically learns appropriate γ values so that it can capture appropriate θ samples in relatively high-likelihood regions in the θ-space.

recall the parameters associated with the basis function (1), write $\omega_k := \left(u_k,\ v_k\right)$, and decompose the stochastic search dynamics (9) into two parts:

$$P\left(u_k \mid u_{k\text{-}1},\ \gamma_k\right) := \frac{1}{Z_{u_k}(\gamma_k)} \exp\left(-\frac{\gamma_k \left\| u_k - u_{k\text{-}1} \right\|^2}{2}\right), \tag{11}$$

$$P\left(v_k \mid v_{k\text{-}1},\ \delta_k\right) := \frac{1}{Z_{v_k}(\delta_k)} \exp\left(-\frac{\delta_k \left\| v_k - v_{k\text{-}1} \right\|^2}{2}\right), \tag{12}$$

where there are two hyperparameters γ_k and δ_k. The corresponding hyperparameter stochastic search dynamics will be given by:

$$P\left(\gamma_k \mid \gamma_{k\text{-}1}\right) := \frac{1}{\sqrt{2\pi}\gamma_k \sigma_h} \exp\left(-\frac{(\log \gamma_k - \log \gamma_{k\text{-}1})^2}{2{\sigma_h}^2}\right), \tag{13}$$

$$P\left(\delta_k \mid \delta_{k\text{-}1}\right) := \frac{1}{\sqrt{2\pi}\delta_k \sigma_h} \exp\left(-\frac{(\log \delta_k - \log \delta_{k\text{-}1})^2}{2{\sigma_h}^2}\right). \tag{14}$$

http://dx.doi.org/10.5772/56146

Since the basis function is linear with respect to v_k, the Rao-Blackwellisation can be conducted with the data augmentation of Z_k [26], where the likelihood function $P(y_k \mid x_k, \omega_k)$ is to be integrated out with respect to v_k, which, in turn, gives rise to smaller variances. A specific implementation of this particular Rao-Blackwellisation will be given in subsection **5.2**.

4.2. Multi-class classification problem

This section attempts to generalize the results of the previous section to multi-class problems. Although we will restrict ourselves to a four-class problem, the method can, in principle, be applied to cases with more than four classes.

Let $x_k \in R^d$ be the feature vector at the k-th trial, which is the power spectrum obtained through DFT over the trial period, where d stands for the dimension of x_k. Let $y_k \in \{1,2, 3,4\}$ denote the class labels at each trial, where left corresponds to label 1, right to label 2, top to label 3, and bottom to label 4. Our goal is to learn the parameters associated with the basis function described below in an attempt to predict the subject's intention.

4.2.1. Basis function

Consider the basis function $f_q(\bullet)$ defined by (15), which is nonlinear with respect to not only x_k but also the parameter vector ω_k. There are Q outputs associated with the basis function, where Q is the number of class labels, which is four in this paper. We have:

$$f_q(x_k;\omega_k)=\sum_{j=1}^{h} v_{k,jq}\sigma\left(\sum_{i=1}^{d} u_{k,ij}x_{k,i}+u_{k,0j}\right)+v_{k,0q}, \tag{15}$$

where $u_k := (u_{k,0}, \cdots, u_{k,d})^T \in R^{h(d+1)}$, $u_{k,i} := (u_{k,i1}, \cdots, u_{k,ih})^T \in R^h$, $v_k := (v_{k,0}, \cdots, v_{k,h})^T \in R^{Q(h+1)}$, $\omega_k = (u_k, v_k)$.

4.2.2. Multinomial logistic model

This paper assumes the Multinomial Logistic Model for the target problems, where it is assumed that the error $_{k,q}$ in each term follows an independently identically distributed logistic distribution. By introducing a latent variable $z_{k,q}$, we write:

$$z_{k,q} := f_q(x_k;\omega_k) + C_{k,q} + {}_{k,q}, \tag{16}$$

$$C_{k,q} := -\log \sum_{i\neq q}^{Q} \exp(f_i(x_k;\omega_k)), \tag{17}$$

where $C_{k,q}$ represents the score of the term controlled by the outputs of the other class labels. It follows from (15) that the probability of y_k belonging to class q is described by:

$$P(y_k = q \mid z_{k,q}) = \frac{\exp(f_q(x_k; \omega_k))}{\sum_{i=1}^{Q} \exp(f_i(x_k; \omega_k))} = \sigma(f_q(x_k; \omega_k) + C_{k,q}), \tag{18}$$

The predicted label y_{pred} is the label q_{max} that has the maximum value of $P(y_k = q \mid z_{k,q})$. Using (18), the likelihood function is described by:

$$P(y_k \mid x_k, \omega_k) := \prod_{q=1}^{Q} P(y_k = q \mid z_{k,q})^{I(y_k = q)} (1 - P(y_k = q \mid z_{k,q}))^{I(y_k \neq q)}, \tag{19}$$

The function $I(\bullet)$ is again an indicator described in **4.1**. The generalization to the multi-class problem (18)-(20) appears straightforward; however, our experience tells us that the multi-class problems are much more difficult than the equations look. Experimental results are reported in **6.4**.

4.2.3. Parameter/hyperparameter search stochastic dynamics

We use the same standard Sequential Monte Carlo (SMC) used in **4.1**.

4.3. Bayesian Sequential Learning

Letting $\theta_k := (\omega_k, \gamma_k, \delta_k)$, $x_{1:k} := (x_1, \cdots, x_k)$, $y_{1:k} := (y_1, \cdots, y_k)$, one can derive its sequential posterior distribution at trial k:

$$P(\theta_k \mid x_{1:k}, y_{1:k}) = \frac{P(y_k \mid x_k, \theta_k) P(\theta_k \mid x_{1:k-1}, y_{1:k-1})}{\int P(y_k \mid x_k, \theta_k) P(\theta_k \mid x_{1:k-1}, y_{1:k-1}) d\theta_k}, \tag{20}$$

The second factor in the numerator is the predictive probability for parameter θ_k, which is given by:

$$P(\theta_k \mid x_{1:k-1}, y_{1:k-1}) = \int P(\theta_k \mid \theta_{k-1}) P(\theta_{k-1} \mid x_{1:k-1}, y_{1:k-1}) d\theta_{k-1}, \tag{21}$$

$$P(\theta_k \mid \theta_{k-1}) = P(\omega_k \mid \omega_{k-1}, \gamma_k) P(\gamma_k \mid \gamma_{k-1}). \tag{22}$$

At the $k+1$-st trial, let the EEG data x_{k+1} be given. Then the prediction at the trial amounts to computing the predictive probability for label $P(y_{k+1})$:

$$P(y_{k+1} \mid x_{1:k+1}, y_{1:k}) = \int P(y_{k+1} \mid x_{k+1}, \theta_{k+1}) P(\theta_{k+1} \mid x_{1:k}, y_{1:k}) d\theta_{k+1}. \tag{23}$$

5. Implementation

The Sequential Monte Carlo (SMC) is a powerful means of evaluating the posterior or predictive probabilities of Bayesian nonlinear or non-Gaussian models in a sequential manner. This

http://dx.doi.org/10.5772/56146

study uses the SMC to evaluate equations (20) and (23) in an attempt to evaluate the SER. The SMC first attempts to approximate the posterior distribution by an empirical distribution (delta mass) weighted by normalized importance weights (importance sampling). In order to avoid depletion of samples, caused by an increase in the variance of the weights, the SMC replaces the weighted empirical distribution by unweighted delta masses (resampling).

There are several different methods of determining when resampling should be done. We tried two of them. One method is to resample every step, and another is to perform resampling only when the Effective Sample Size (EES [22]-[24]) becomes smaller than a threshold value:

$$ESS = \frac{1}{\sum_{i=1}^{n} \left(\tilde{\Omega}_k^{(i)} \right)^2}, \tag{24}$$

where n is the number of samples, i is the index of a sample, and $\tilde{\Omega}_k$ is the normalized importance weight defined by $\tilde{\Omega}_k^{(i)}$.

The threshold value of ESS is often set at $N/2$ ([23],[24]), which we adopted.

Implementation of Standard SMC

(a) Importance Sampling step

(i) For $i = 1, ..., n$, draw samples of $\boldsymbol{\theta}_k^*$ from the proposal density Q:

$$\left\{ \boldsymbol{\theta}_k^{*(i)} \right\}_{i=1}^{n} \sim Q(\boldsymbol{\theta}_k^* | \boldsymbol{x}_{1:k}, y_{1:k}).$$

(ii) For $i = 1, ..., n$, compute the importance weight $\Omega_k^{(i)}$:

$$\begin{aligned} \Omega_k^{(i)} &\propto \tilde{\Omega}_{k-1}^{(i)} \frac{P(y_k | \boldsymbol{x}_k, \boldsymbol{w}_k^{(i)}) P(\boldsymbol{\theta}_k^{*(i)} | \boldsymbol{x}_{1:k-1}, y_{1:k-1})}{Q(\boldsymbol{\theta}_k^{*(i)} | \boldsymbol{x}_{1:k}, y_{1:k})}, \\ &= \tilde{\Omega}_{k-1}^{(i)} P(y_k | \boldsymbol{x}_k, \boldsymbol{w}_k^{(i)}). \end{aligned}$$

(iii) For $i = 1, ..., n$, compute the normalized importance weight $\tilde{\Omega}_k^{(i)}$:

$$\tilde{\Omega}_k^{(i)} = \frac{\Omega_k^{(i)}}{\sum_{j=1}^{n} \Omega_k^{(j)}},$$

where $\sum_{j=1}^{n} \tilde{\Omega}_k^{(j)} = 1$.

(iv) Calculate the ESS using (24).

$$\begin{cases} \text{if} & ESS < \left(\frac{n}{2}\right) & \text{go to (v)} \\ \text{else} & & \text{k=k+1 and go to (i)} \end{cases}$$

(b) Resampling step

(v) Resample $\left\{ \boldsymbol{\theta}_k^{*(i)} \right\}_{i=1}^{n}$ with probability $\{\tilde{\Omega}_k^{(i)}\}_{i=1}^{n}$, and then set all the normalized importance weights $\frac{1}{n}$.

Figure 6. Implementation of standard SMC.

5.1. Standard SMC

Figure 6 gives an overview of the standard SMC, where $\theta_k^* := (\omega_k, \gamma_k)$ and $Q(\bullet)$ denote the proposal distribution, which in this paper is set as $P(\theta_k^* \mid x_{1:k-1}, y_{1:k-1})$. This choice is due to its simplicity of implementation. Figure 6 demonstrates the case where resampling is done with the ESS. If resampling is performed every step, then the ESS step is simply ignored.

5.2. Rao-blackwellised SMC

In the Rao-Blackwellised SMC implementation, the marginal likelihood $P(y_k \mid z_{1:k}, u_{1:k})$ is used instead of the likelihood function $P(y_k \mid x_k, \omega_k)$, where $\Theta_k := (\omega_k, \gamma_k, \delta_k)$. This implementation is described in Figure 7.

Implementation of Rao-Blackwellised SMC

(a) Importance sampling step

(i) For $i = 1, ..., n$, draw samples of $\boldsymbol{\Theta}_k$ from the proposal density Q:

$$\left\{\boldsymbol{\Theta}_k^{(i)}\right\}_{i=1}^{n} \sim Q(\boldsymbol{\Theta}_k | \boldsymbol{x}_{1:k}, y_{1:k}).$$

(ii) For $i = 1, ..., n$, compute the importance weight $\Omega_k^{(i)}$:

$$\begin{aligned} \Omega_k^{(i)} &\propto \tilde{\Omega}_{k-1}^{(i)} \frac{P(y_k | z_{1:k}^{(i)}, \boldsymbol{u}_{1:k}^{(i)}) P(\boldsymbol{\Theta}_k^{(i)} | \boldsymbol{x}_{1:k-1}, y_{1:k-1})}{Q(\boldsymbol{\Theta}_k^{(i)} | \boldsymbol{x}_{1:k}, y_{1:k})}, \\ &= \tilde{\Omega}_{k-1}^{(i)} P(y_k | z_{1:k}^{(i)}, \boldsymbol{u}_{1:k}^{(i)}). \end{aligned}$$

(iii) For $i = 1, ..., n$, compute the normalized importance weight $\tilde{\Omega}_k^{(i)}$:

$$\tilde{\Omega}_k^{(i)} = \frac{\Omega_k^{(i)}}{\sum_{j=1}^{n} \Omega_k^{(j)}},$$

where $\sum_{j=1}^{n} \tilde{\Omega}_k^{(j)} = 1$.

(iv) For $i = 1, ..., n$, update $\boldsymbol{X}_k^{(i)}$ by using $(y_k^{(i)}, \boldsymbol{X}_{k-1}^{(i)})$. For details regarding the updating, see the Appendix, where $\boldsymbol{X}_k = (\boldsymbol{V}_{k|k}, \boldsymbol{v}_{k|k})$.

(v) Calculate the ESS using (24).

$$\begin{cases} \text{if} & ESS < \left(\frac{n}{2}\right) & \text{go to (vi)} \\ \text{else} & & \text{k=k+1 and go to (i)} \end{cases}$$

(b) Resampling step

(vi) Resample $\left\{\boldsymbol{\Theta}_k^{(i)}, \boldsymbol{X}_k^{(i)}\right\}_{i=1}^{n}$ with probability $\{\tilde{\Omega}_k^{(i)}\}_{i=1}^{n}$, and then set all the normalized importance weights $\frac{1}{n}$.

Figure 7. Implementation of Rao-Blackwellised SMC.

Here, the marginal likelihood $P(y_k \mid z_{1:k}, u_{1:k})$ can be written as:

$$P(y_k \mid z_{1:k}, u_{1:k}) = \Phi\left(\frac{z_{k|k-1}}{\sqrt{S_k}}\right)^{y_k} \left(1 - \Phi\left(\frac{z_{k|k-1}}{\sqrt{S_k}}\right)\right)^{1-y_k} \tag{25}$$

http://dx.doi.org/10.5772/56146

Details of updating $z_{k|k-1}$ and S_k are given in the Appendix.

6. Results

This section reports the results of learning experiments using the algorithms proposed in the previous sections.

6.1. Observation data

As explained in **3**, the six channels (*O1, OZ, O2, O9, IZ, O10*) located in the occipital eye field were used for our classification problems. Data were taken in 600 trials from each of five subjects (*A, B, C, D, E*). One trial lasted for 3.0 seconds. Data in the first 0.0-1.0 s was deleted in order to eliminate the effect of eyeball movements on the EEG. The raw data were filtered by 50.0 Hz notch filter. After Hanning windowing, DFT was performed by *MATLAB_R2008a* to obtain the feature vector x consisting of the power spectrum.

Figure.8 demonstrates the power spectrum of subject *D* taken from one trial when a stimulus was presented at 6.0 Hz. The particular frequency component is relatively clear. Figure.9 is from another trial of the same subject, where the target frequency component is not clearly discernible.

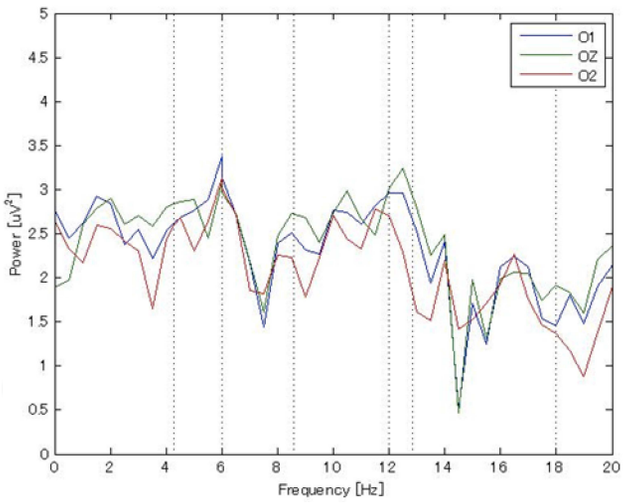

Figure 8. Frequency spectrum of Subject D. The target frequency of 6.0 Hz is reasonably discernible.

The vertical lines in the two figures indicate (from the left) 4.29Hz, 6.0Hz, 8.58(4.29 × 2) Hz, 12.0(6.0 × 2) Hz, 12.87(4.29 × 3) Hz and 18.0(6.0 × 3) Hz, respectively. It should be noted that even with SSVEP, the observed frequency components are not always identifiable by inspection. It should also be noted that SSVEP can contain higher harmonics of the target frequency [4] and that the classification accuracy may be improved by taking into account higher harmonics [4]. In order to examine the effectiveness of the higher order harmonics for our

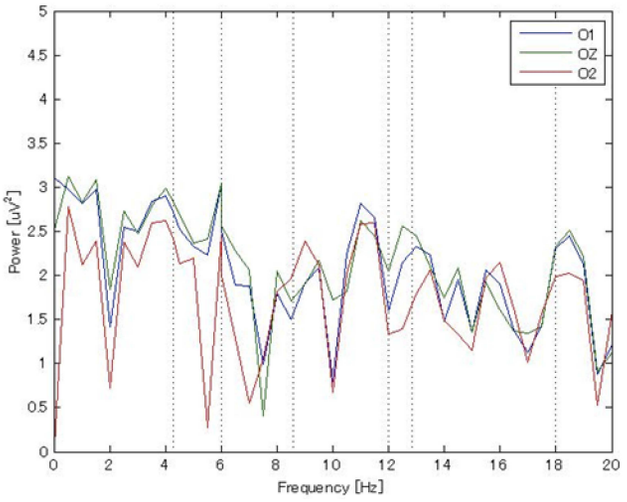

Figure 9. Frequency spectrum of the same subject at in Figure.8. The target frequency component is difficult to observe.

classification problem, this section considers the following three settings: (i) the fundamental frequency only, (ii) the second order harmonics, in addition to the fundamental frequency, and (iii) second and third order harmonics, in addition to the fundamental frequency. Since the number of channels is 6, the dimensions of our feature vectors are (i) 12, (ii) 24, and (iii) 36, respectively.

It should be noted that while more frequency components give more information, the number of parameters to be learned increases, so that learning becomes more difficult.

6.2. Experimental settings

This study examines several different versions of Sequential Monte Carlo for implementing the target sequential learning, as displayed in Table 1, where standard SMC means no hyperparameter learning and resampling is performed at every step. The abbreviated notation will be used throughout the rest of the paper. Various experimental settings are summarized in Table 2, where n denotes the number of samples; γ_0 and δ_0, the initial conditions for hyperparameters γ and δ ; σ_h, the hyper-hyperparameter; and h, the number of perceptron hidden units.

6.3. Performance evaluation criteria

We will propose three performance evaluation criteria. One is Sequential Error Rates (SER_k) defined by

$$SER_k := \frac{1}{M} \sum_{k'=k-M+1}^{k} I\left(y_{k'} \neq y_{k',pred}\right), \tag{26}$$

http://dx.doi.org/10.5772/56146

Abbreviation	Algorithm
SMC	Standard SMC
HP+SMC	SMC with hyperparameter auto-adjustment
SMCESS	SMC by calculating ESS
HP+SMCESS	SMCESS with hyperparameter auto-adjustment
RBSMC	Rao-Blackwellised SMC
HP+RBSMC	RBSMC with hyperparameter auto-adjustment
RBSMCESS	RBSMC by calculating ESS
HP+RBSMCESS	RBSMCESS with hyperparameter auto-adjustment
SMCmulti	Standard SMC for multi-class classification
HP+SMCmulti	SMCmulti with hyperparameter auto-adjustmen

Table 1. Algorithm names and their abbreviations

Algorithm	n	γ_0	δ_0	σ_h	h
SMC	1000	100.0	-	-	10
HP+SMC	1000	100.0	-	0.01	10
SMCESS	1000	100.0	-	-	10
HP+SMCESS	1000	100.0	-	0.01	10
RBSMC	1000	100.0	100.0	-	10
HP+RBSMC	1000	100.0	100.0	0.01	10
RBSMCESS	1000	100.0	100.0	-	10
HP+RBSMCESS	1000	100.0	100.0	0.01	10
SMCmulti	1000	100.0	-	-	10
HP+SMCmulti	1000	100.0	-	0.02	10

Table 2. Experimental Settings. n denotes the number of samples; γ_0 and δ_0, the initial conditions for hyperparameters γ and δ ; σ_h , the hyper-hyperparameter; and h , the number of perceptron hidden units.

where $y_k(y_{k'})$ is the true class, and $y_{k,pred}(y_{k',pred})$ is the predicted class defined by (23). Notation $I(\bullet)$ stands for an indicator described in **4**. This is the moving average of the prediction error over a window of size M. We will also compute Cumulative Error (CE)

$$CE := \sum_{k=1}^{K} I(y_k \neq y_{k,pred}), \tag{27}$$

in order to make performance comparisons with the existing methods. Another quantity we will be evaluating is the sequential marginal likelihood:

$$P(y_k \mid x_{1:k}, y_{1:k-1}) = \int P(y_k \mid x_k, \theta_k)\, P(\theta_k \mid x_{1:k-1}, y_{1:k-1}) d\theta_k \tag{28}$$

which is the marginalization of the likelihood with respect to the current predictive distribution. This quantifies the reliability of the prediction y_k with respect to $(x_{1:k}, y_{1:k-1})$. In order to explain a *rationale* behind this, recall that given data y, the likelihood $P(y \mid z)$ can be interpreted as the degree of appropriateness of z in explaining y. This, in turn, can be interpreted as the appropriateness of y in terms of z.

6.4. Experimental results

6.4.1. Two-class classification problem

a. Sequential Error Rate

Figure 10 shows the Sequential Error Rate of subject D over one session consisting of 600 trials. The algorithm was implemented by Sequential Monte Carlo together with the proposed hyperparameter learning (HP+SMC). Table 3 summarizes the Sequential Error Rates of subjects A-E, which were averages over ten learning trials.

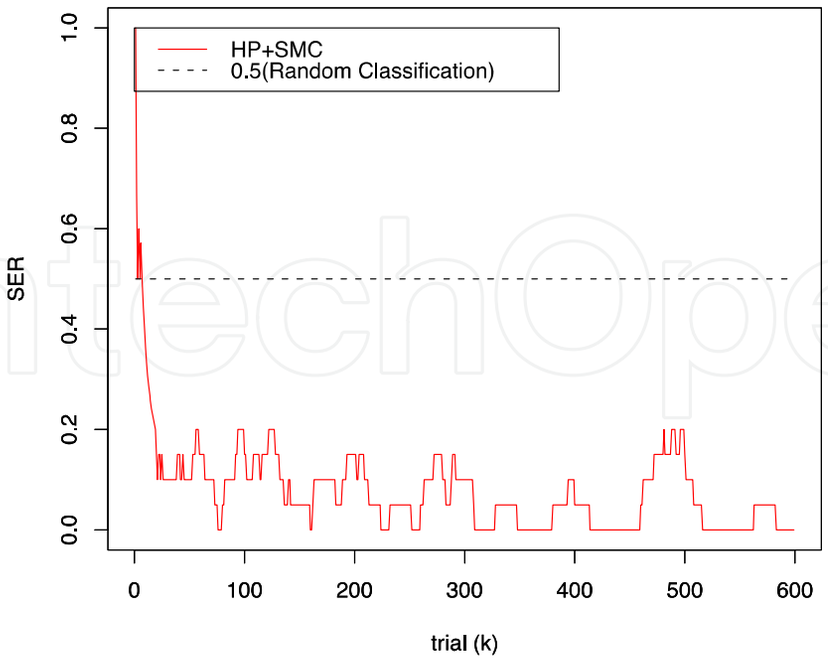

Figure 10. Sequential Error Rate of subject D with (HP+SMC), Sequential Monte Carlo together with the proposed hyperparameter learning. The dotted line at 1/2 corresponds to a random classification.

http://dx.doi.org/10.5772/56146

	A	B	C	D	E
minimum error rate	0.00	0.010	0.00	0.00	0.00
maximum error rate	0.75	0.75	0.71	0.75	0.80
average over 600 trials	0.16	0.26	0.19	0.077	0.13

Table 3. Sequential Error Rate of the subjects (M=20, HP+SMC)

b. Trajectory of hyperparameter

In Figure 11, the γ_t-trajectory (blue) is superimposed on the Sequential Error Rates (red) of Figure 10. The value of γ_t is the posterior mean. Note that there was a significant dip in γ_t around 380 and 480, due to the fact that the algorithm detected a sudden change in the data, so that it automatically widened the search region in the parameter space. Eventually, the algorithm re-started learning the parameters. This phenomenon was also discernible at around 475. The hyperparameter learning appeared functional.

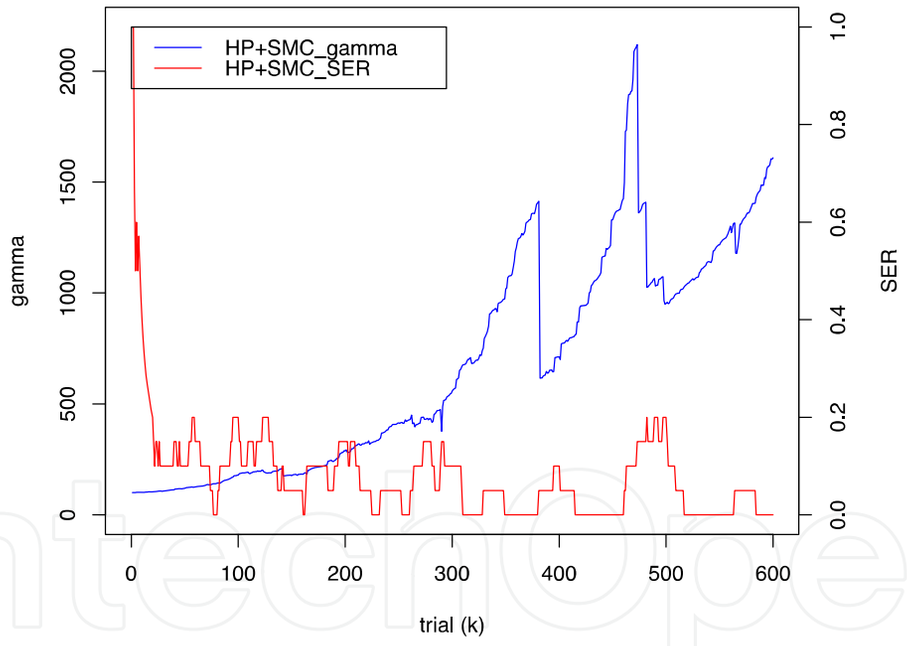

Figure 11. Trajectory of hyperparameter γ_k for Subject D with the SER in Fig.10 (HP+SMC) superimposed. The value of γ_k was its posterior mean

c. Sequential Marginal Likelihood (Reliability of the Predictions)

Figure 12 shows the negative log-Sequential Marginal Likelihood of subject D averaged over the window M=20 as was in Figure 10. Even though Figure 10 and Figure 12 are similar, the latter comes from the Bayesian concept where the latter appears slightly less abrupt. This particular quantity can be applied to the change detection problem as is done in [27].

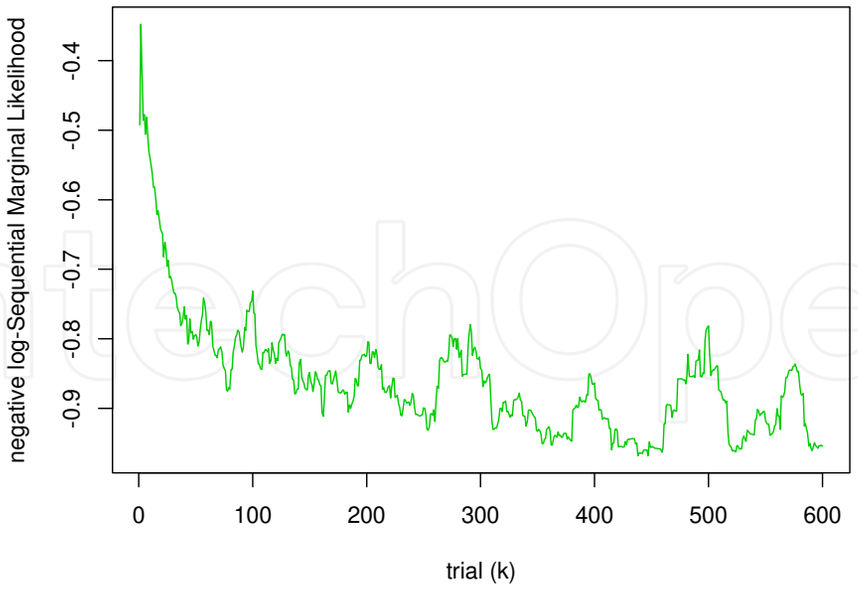

Figure 12. Negative log-Sequential Marginal Likelihood of subject D with moving average *M*=20.

d. Cumulative Error

Figure.13 shows the Cumulative Error of subject D with different algorithms, and Table.4 gives final Cumulative Errors of subjects A-E, that is, the Cumulative Errors at the last trial. These values were the averages over ten experiments.

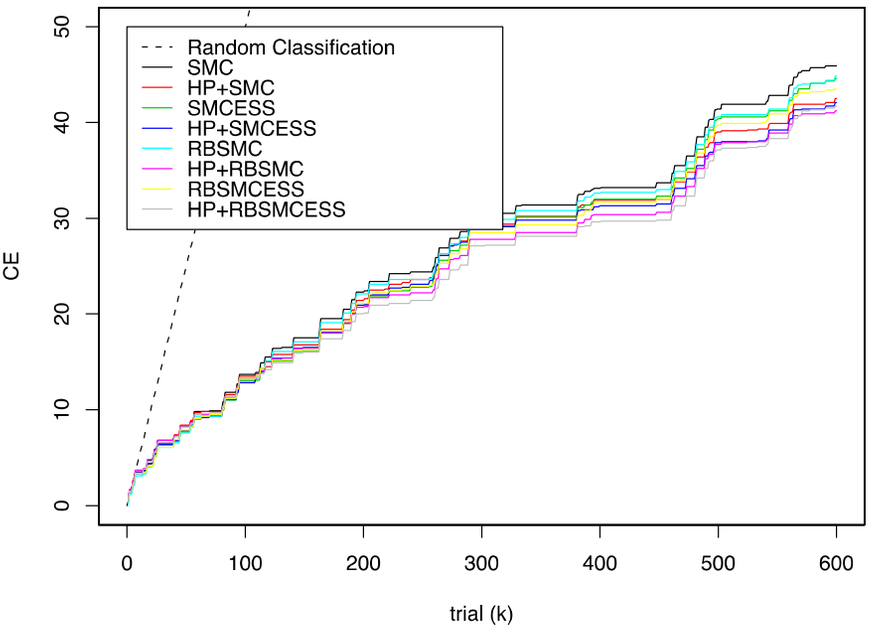

Figure 13. Cumulative Error (CE) of subject D. Different colors indicate different versions of the algorithms, as shown in Table. I.The dotted straight line at 1/2 corresponds to a random classification.

http://dx.doi.org/10.5772/56146

	A	B	C	D	E
SMC	95.50	159.2	109.3	45.90	80.10
HP+SMC	90.70	151.4	113.7	42.50	75.20
SMCESS	96.20	170.5	108.3	44.60	80.30
HP+SMCESS	92.10	155.9	110.0	42.10	71.70
RBSMC	91.90	155.8	109.1	44.80	75.20
HP+RBSMC	88.80	158.4	109.3	41.20	73.90
RBSMCESS	91.00	157.1	107.6	43.50	76.00
HP+RBSMCESS	89.40	154.6	106.3	41.70	72.70

Table 4. Final Cumulative Error

e. Effective Sample Size

Figure 14 shows the ESS trajectories (moving average over 20 trials) of subject D with several different methods.

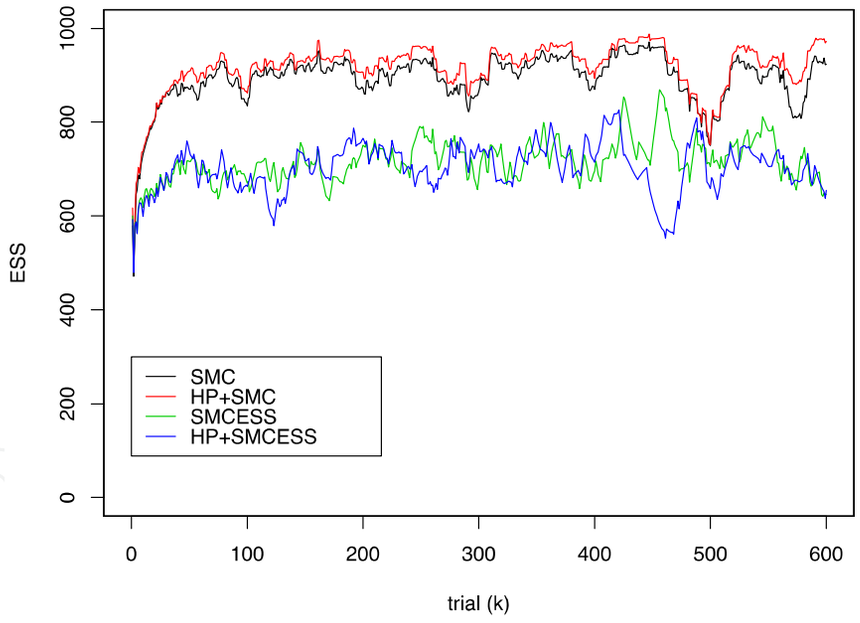

Figure 14. Trajectory of ESS of subject D.

f. Computation Time

Table 5 summarizes the computation time of the various methods averaged over ten experiments. The middle column shows the time per trial, whereas the right-most column shows the time needed for all trials. The par trial time does not contain the case where resampling is done with the ESS.

	1 step (S)	whole data (S)
SMC	0.120	72.0
HP+SMC	0.130	77.7
SMCESS	-	57.4
HP+SMCESS	-	63.9
RBSMC	0.320	192
HP+RBSMC	0.328	197
RBSMCESS	-	145
HP+RBSMCESS	-	121

Table 5. Computation time

g. Harmonic Frequency Components

Effects of the higher order harmonics were examined and are summarized in Table.6 for three cases: (i) fundamental frequency only, (ii) second-order higher harmonics in addition to the fundamental frequency, and (iii) second- and third-order higher harmonics in addition to the fundamental frequency. The numbers in the table indicate the final Cumulative Errors. The results were the averages over ten experiments.

	A	B	C	D	E
(i) fundamental	90.70	151.4	113.7	42.50	75.20
(ii) fundamental+2nd	98.20	141.1	46.50	37.60	92.80
(iii) fundamental+2nd+3rd	78.20	131.3	44.30	45.70	104.0

Table 6. Effect of Harmonics (Cumulative Error)

6.4.2. Multi-class classification problem

a. Sequential Error Rate: Figure 15 shows the Sequential Error Rates (moving average window size 20) of subject D for the the four-class problem with the HP+SMC algorithm, and Table 7 gives various values related to the Sequential Error Rates of subjects A-E averaged over 10 experiments.

	A	B	C	D	E
minimum error rate	0.23	0.39	0.25	0.15	0.58
maximum error rate	0.74	0.82	0.60	0.72	0.85
average over 600 trials	0.46	0.60	0.47	0.36	0.72

Table 7. Sequential Error Rate of the subjects (M=20, HP+SMCmulti)

http://dx.doi.org/10.5772/56146

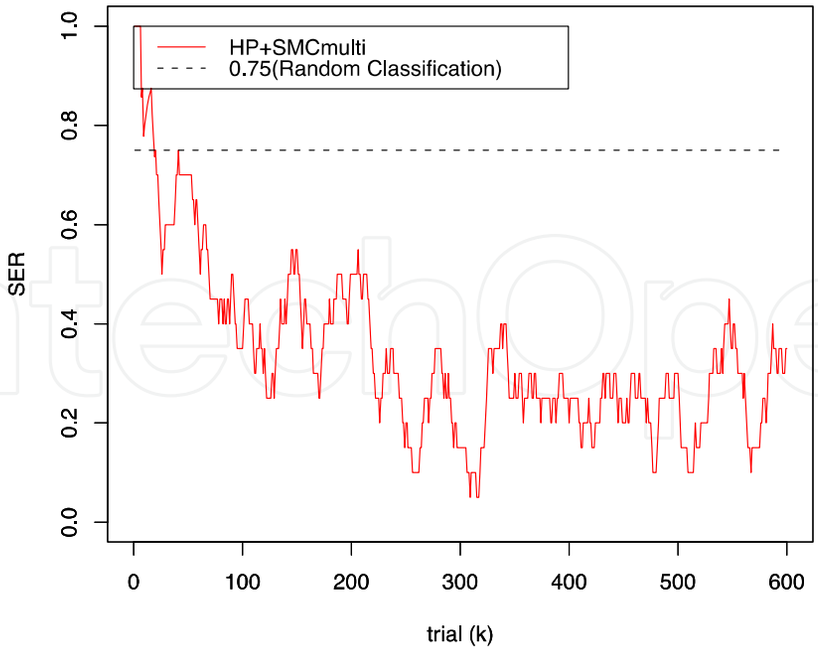

Figure 15. Sequential Error Rate of four-class classification for subject D, where the hyperparameter is learned together with SMC (HP+SMCmulti). The dotted line at 3/4 corresponds to random classification.

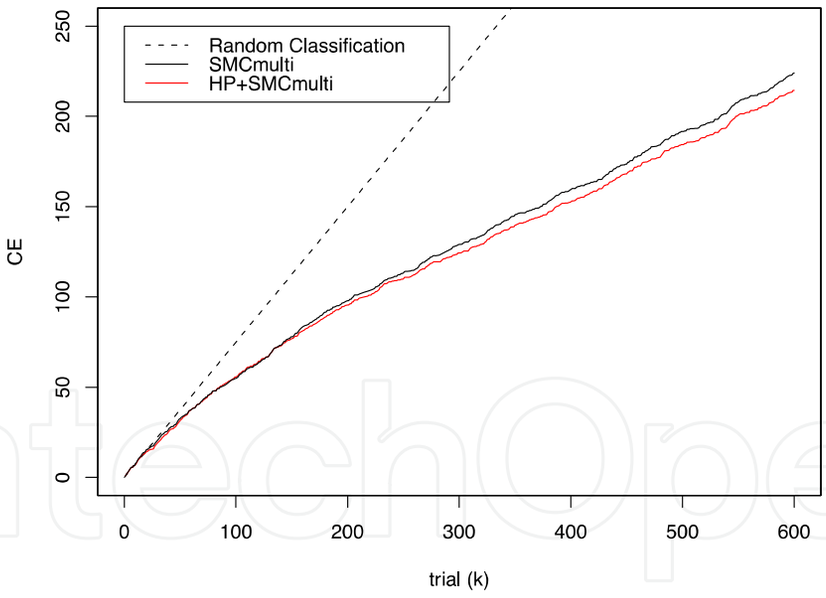

Figure 16. Cumulative Error of four-class classification for subject D with two different algorithms. One is the standard SMC without hyperparameter learning, and the other is with the proposed hyperparameter learning. The dotted straight line indicates a random classification.

b. Cumulative Error: Figure 16 shows Cumulative Errors of subject D for the four-class problem with two different algorithms. One is a standard SMC without hyperparameter learning (SMCmulti), and the other is the proposed SMC with hyperparameter learning

(HP+SMCmulti). The dotted line indicates a random classification. Table.8 summarizes the CEs for subjects A-E. These are the values averaged over ten experiments.

	A	B	C	D	E
SMC_{MULTI}	283.2	363.1	285.1	224.0	434.1
$HP + SMC_{MULTI}$	280.3	361.7	281.6	214.5	434.8

Table 8. Final Cumulative Error of four class classification

7. Discussion

7.1. Sequential error rate

We first observe that there are two errors involved in brain-computer interfaces in general, and in this study in particular. One is the error made by the brain (subject), and the other is the error made by the computer (algorithm), provided that the hardware behind the experiments is functional. Let us look at the Sequential Error Rate in Figure 10. It started decreasing immediately after the experiment began, and it had already dropped to about 0.1 at around the 20-th trial. At around the 80-th trial, the Sequential Error Rate became almost 0. One possible interpretation of this is that, if we can assume that the subject does not make an error during these 80 trials, then the SER trajectory represents the process of how the computer learns the classification problem. Recall that there are h(d+2)+1 parameters in (1), which in this case is 141. In addition, the hyperparameter γ needs to be learned. This means that the parameter/hyperparameter landscape is vague at the beginning in a high dimensional space, so that the computer searches for posterior samples in an attempt to find appropriate parameter values for better classification. It should be noted that the hyperparameter γ_t is relatively flat up to trial 80 but slightly increases. Since γ_t represents the reciprocal of the size of the parameter search region, this period can be interpreted as the computer's early effort to search for parameters by slightly narrowing down the parameter space search region.

At around trial 80, the SER dropped to almost zero, so that if the subject's EEG signals were consistent with the previous ones, the computer algorithm did not need to seek different samples in the parameter space. Therefore, the ups and downs after trial 80 could be interpreted as the fact that the subject's EEG signals became slightly inconsistent with the previous ones. During this period, the computer naturally needed to search for slightly different posterior samples, so that some errors were incurred.

This was followed by several ups and downs between 0.2 and almost 0 until approximately trial 310. The subject seemed to have obtained a reasonable amount of skill for the task, so that the subject achieved almost 0. The Sequential Error Rate in the trials between 310 and 330, as well as between 350 and 380, were almost zero. However, the subject's Sequential Error Rate again increased at around trial 380. A sharp dip was observed in the hyperparameter trajectory,

http://dx.doi.org/10.5772/56146

as demonstrated in Figure 11. One possible interpretation of this is that by trial 380, the computer had found fairly good posterior samples for predictions so that the parameter search region was narrowed down; however, a sudden change was observed and the computer needed to quickly widen its stochastic parameter search region, which was indicated by the sudden drop of hyperparameter γ_t at around k=380. With this, the algorithm tried to learn parameter values different from the previous ones and eventually found better parameter values. The Sequential Error Rate again dropped to almost 0 at around trial 420, which lasted for approximately 40 trials. A similar phenomenon was discernible after around trial 480. It is important to notice that, in addition to the learning mechanism, a "forgetting" mechanism is naturally built in. Namely, 9 and 10 are first-order Markov stochastic dynamical systems, so that memories of the distant past are forgotten, whereas the more recent data are taken into account with higher weights. Note, however, that the Sequential Monte Carlo algorithm took into account several hundred parameter values instead of a single parameter value, which endowed the predictions with robustness.

A question might arise as to why was the drop in γ_t at around k=380 much more significant than, for instance, that at around k=270, where the SER increase was more significant at the latter than the former. Our interpretation is that at around k=270, $\gamma_t$is not too large, so that the parameter search region is still reasonably large, whereas at around k=380, the parameter search region was already sharp, and a more sudden change of the search region size was needed.

7.2. Sequential marginal likelihood

Since Sequential Marginal Likelihood can be interpreted as the reliability of each prediction of the subject, this quantity can also be used to evaluate subject's performance with probabilistic justification. Another potential application of this quantity is its use in change detection problem such that a significant change of this quantity would indicate occurrence of change in the subject's signal quality or/and environmental change. It should be noted that the sequential marginal likelihood is a well defined Bayesian quantity whereas such reliability index is not available in maximum likelihood method.

7.3. Rao-blackwellisation

Note that in Figure13, the best performance was achieved by HP+RBSMCESS, where the Rao-Blackwellisation and the Effective Sample Size were taken into account, in addition to the hyperparameter learning. The proposed scheme appeared functional.

Figure 13 shows the Cumulative Error of subject D, where the black dotted line shows the Cumulative Error corresponding to random classification. Since the Cumulative Error with the proposed algorithms grows slower than the random classification, the results appeared to indicate that the algorithms were functional. Figure 13 appears to indicate that the proposed γ_t-learning, as well as the Rao-Blackwellised SMC, was functional.

7.4. Effective sample size

From the trajectory of the Effective Sample Size (ESS), we observed that the ESS was generally large if resampling was performed at each step. This could be attributable to the fact that the purpose of resampling was to avoid degeneracy of samples, i.e., to bring in more diversity in the samples. Table 5 appears to indicate that the computation time with ESS was significantly reduced since it avoided sampling when ESS did not become smaller than a threshold value.

7.5. Higher-order harmonics

Taking the higher-order harmonics into account generally improved the prediction capabilities, except for subject E, as was seen in Table 6. One future research project could be to develop an algorithm to choose appropriate frequency components automatically. The number of frequency components is also related to the overfitting problem in machine learning, where the number of parameters is large compared with the number of data available, which sometimes results in performance degradation.

7.6. Multi-class classification problem

The extension of the two-class problem to the four-class classification problem discussed in Section 4.2 was nontrivial. One of the difficulties can be seen from the term $C_{k,q}$ in (17) - (19), where the values of equation (18) must be well-separated from each other for the four classes. The experimental results reported in subsection 6.4, however, appeared reasonable. The Cumulative Errors shown in Figure.16 appeared to indicate that the learning was functional.

8. Conclusion

This paper proposed Bayesian sequential learning algorithms for SSVEP sequential classification problems in a principled manner. Two experiments were conducted: one involving a two-class problem, and the other involving a four-class problem. The stimuli consisted of a flickering checkerboard at frequencies ranging from 4.29 to 10.0 Hz. The algorithms were implemented by the Sequential Monte Carlo. One of the points of the proposed algorithms was their hyperparameter learning, enabling it to automatically adjust to environmental changes, including changes in the subjects' physical conditions as well as their environments. Computation costs were also measured, which appeared to indicate that the algorithms could be implemented in real time. The proposed algorithms appeared functional. The proposed sequential algorithms are applicable to other brain signals besides EEG.

In the experiments performed in this study, the subjects were asked to look at the stimuli. A future research project is to examine sequential classification problems with covert selective attention [20], [21] where subjects are asked to pay attention to stimuli without eyeball movements. This project is in progress and will be reported in a future paper.

http://dx.doi.org/10.5772/56146

Appendix

Update Equations of $z_{k|k-1}$ and S_k

The update equations of $z_{k|k-1}$ and S_k can be summarized as follows:

$$v_{k|k-1} = v_{k-1|k-1}$$

$$V_{k|k-1} = V_{k-1|k-1} + \delta_k^{-1}$$

$$S_k = \Psi_k^T(x_k; u_k) V_{k|k-1} \Psi_k(x_k; u_k) + 1$$

$$z_{k|k-1} = \Psi_k^T(x_k; u_k) v_{k|k-1}$$

$$K_k = V_{k|k-1} \Psi_k(x_k; u_k) S_k^{-1}$$

$$v_{k|k} = v_{k|k-1} + K_k(z_k - z_{k|k-1})$$

$$V_{k|k} = V_{k|k-1} - K_k \Psi_k^T(x_k; u_k) V_{k|k-1}$$

Where $v_{k|k-1} := E[v_k \mid \mathbf{\Theta}_{1:k-1}]$, $v_{k|k} := E[v_k \mid \mathbf{\Theta}_{1:k}]$, $V_{k|k-1} := Cov[v_k \mid \mathbf{\Theta}_{1:k-1}]$, $V_{k|k} := Cov[v_k \mid \mathbf{\Theta}_{1:k}]$, $z_{k|k-1} = E[z_k \mid \mathbf{\Theta}_{1:k-1}]$, and $S_k = Var[z_k \mid \mathbf{\Theta}_{1:k-1}]$.

Acknowledgements

The authors thank A. Doucet for valuable comments.

Author details

S. Shigezumi[1], H. Hara[1], H. Namba[1], C. Serizawa[1], Y. Dobashi[1], A. Takemoto[2], K. Nakamura[2] and T. Matsumoto[1]

1 Department of Electrical Engineering and Bioscience, Waseda University, Tokyo, Japan

2 Primate Research Institute, Kyoto University, Aichi, Japan

References

[1] Niels Birbaumer ``Breaking the silence: Brain-computer interfaces (BCI) for communication and motor control*Psychophysiology*, (2006). , 43(6), 517-532.

[2] Bashashati, A, Fatourechi, M, Ward, R. K, & Birch, G. E. A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals," *J. Neural Eng.*, (2007). , 4(2), R32-R57.

[3] G. R. Mu¨ller-Putz, R. Scherer, C. Brauneis, and G. Pfurtscheller, ``Steady-state visual evoked potential (SSVEP)-based communication: impact of harmonic frequency components," *J. Neural Eng.*, vol. 2, pp. 123-130, 2005.

[4] Martinez, P, Bakardjian, H, & Cichocki, A. Fully Online Multicommand Brain-Computer Interface with Visual Neurofeedback Using SSVEP Paradigm," *Comput. Intell. Neurosci.*, , 2007, 1-9.

[5] B. Allison, T. Lu¨th, D. Valbuena, A. Teymourian, I. Volosyak, and A. Gra¨ser, ``BCI Demographics: How Many (and What Kinds of) People Can Use an SSVEP BCI?," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 2, pp. 107-115, 2010.

[6] G. R. Mu¨ller-Putz and G. Pfurtscheller, ``Control of an Electrical Prosthesis With an SSVEP-Based BCI," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 361-364, 2008.

[7] Ortner, R, Allison, B. Z, Korisek, G, Gaggl, H, & Pfurtscheller, G. An SSVEP BCI to Control a Hand Orthosis for Persons With Tetraplegia," *IEEE Trans. Neural Syst. Rehabil. Eng.*, (2011). , 19(1), 1-5.

[8] Cecotti, H, & Self-paced, A. and Calibration-Less SSVEP-Based Brain-Computer Interface Speller," *IEEE Trans. Neural Syst. Rehabil. Eng.*, (2010). , 18(2), 127-133.

[9] Li, J, Zhang, L, Tao, D, Sun, H, & Zhao, Q. A Prior Neurophysiologic Knowledge Free Tensor-Based Scheme for Single Trial EEG Classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, (2009). , 17(2), 107-115.

[10] Wu, H. Y, Lee, P. L, Chang, H. C, & Hsieh, J. C. Accounting for Phase Drifts in SSVEP-Based BCIs by Means of Biphasic Stimulation," *IEEE Trans. Biomed. Eng.*, (2011). , 58(5), 1394-1402.

[11] Malik, W. Q, Truccolo, W, Brown, E. N, & Hochberg, L. R. Efficient Decoding With Steady-State Kalman Filter in Neural Interface Systems," *IEEE Trans. Neural Syst. Rehabil. Eng.*, (2011). , 19(1), 25-34.

[12] Lin, Z, Zhang, C, Wu, W, & Gao, X. Frequency Recognition Based on Canonical Correlation Analysis for SSVEP-Based BCIs," *IEEE Trans. Biomed. Eng.*, (2007). , 54(6), 1172-1176.

[13] Buttfield, A, Ferrez, P. W, & Millan, J. R. Towards a Robust BCI: Error Potentials and Online Learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, (2006). , 14(2), 164-168.

[14] Lu, S, Guan, C, & Zhang, H. Unsupervised Brain Computer Interface Based on Inter-subject Information and Online Adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, (2009). , 17(2), 135-145.

[15] Yoon, J. W, Roberts, S. J, Dyson, M, & Gan, J. Q. Adaptive classification for Brain Computer Interface systems using Sequential Monte Carlo sampling," *Neural Netw.*, (2009). , 22, 1286-1294.

[16] Sega, K, Nakada, Y, & Matsumoto, T. Online Bayesian Learning for Dynamical Classification Problem Using Natural Sequential Prior," in *Proc. IEEE MLSP'08*, (2008). , 392-397.

[17] Hara, H, Takemoto, A, Dobashi, Y, Nakamura, K, & Matsumoto, T. Sequential Error Rate Evaluation of SSVEP Classification Problem with Bayesian Sequential Learning," The 10th IEEE International Conference on Information Technology and Applications in Biomedicine, Nov.(2010). Corfu, Greece., 2-5.

[18] Regan, D. Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine," Elsevier, New York, (1989).

[19] Danhua ZhuJordi Bieger, Gary Garcia Molina, and Ronald M. Aarts, ``A Survey of Stimulation Methods Used in SSVEP-Based BCIs, " *Computational Intelligence and Neuroscience*, Article ID 702357, 12 , 2010, 2010.

[20] Kelly, S. P, Lalor, E. C, Finucane, C, Mcdarby, G, & Reilly, R. B. Visual spatial attention control in an independent brain-computer interface," *IEEE Trans. Biomed. Eng.*, (2005). , 52(9), 1588-1596.

[21] Allison, B, Mcfarland, D, Schalk, G, Zheng, S, Jackson, M, & Wolpaw, J. Towards an independent brain-computer interface using steady state visual evoked potentials," *Clin. Neurophysiol.*, (2008). , 119(2), 399-408.

[22] A. Doucet et. al, eds., ``Sequential Monte Carlo in Practice," Springer, 2001.

[23] Del, P, & Moral, A. Doucet, and A. Jasra, ``Sequential Monte Carlo samplers," *J. Roy. Stat. Soc. Ser. B*, (2006). , 68(3), 411.

[24] Sisson, S. A, Fan, Y, & Tanaka, M. M. Sequential Monte Carlo without likelihoods," *Proc. Natl Acad. Sci. USA* 104, 17601765, (2007).

[25] Casella, G, & Robert, C. P. Rao-Blackwellization of sampling schemes", *Biometrika*, (1996). , 83, 81-94.

[26] Andrieu, C, De Freitas, N, & Doucet, A. Rao-Blackwellised Particle Filtering via Data Augmentation", *Advances in Neural Information Processing Systems*, (2001).

[27] Matsumoto, T, & Yosui, K. Adaptation and Change Detection With a Sequential Monte Carlo Scheme ", *IEEE Trans. System, Man, and Cybernetics.*, (2007). , 37(3), 592-606.

IntechOpen