EEG Discrimination using Wavelet Packet Transform and a Reduced-dimensional Recurrent Neural Network

Nan Bu, Keisuke Shima, and Toshio Tsuji

***Abstract*— This paper proposes a novel reduced-dimensional recurrent neural network (NN) for electroencephalography (EEG) discrimination. Due to time-varying characteristics of EEG signals, recurrent NN is a useful approach for EEG pattern discrimination. However, when dealing with high-dimensional data, NNs usually have problems of heavy computation burden and difficulty in training. To overcome these problems, the proposed NN incorporates a dimension-reducing stage into the network structure of a recurrent probabilistic NN. Moreover, an EEG discrimination method is developed using wavelet packet transform (WPT) and the proposed NN. EEG discrimination experiments were conducted with EEG signals measured during finger movements. The experimental results of four subjects indicate that the proposed method can achieve relatively high discrimination performance.**

I. INTRODUCTION

Electroencephalography (EEG) signals consist of neuronal action potentials recorded from the scalp and reflect central nervous system activities. Regardless of the degree of physical disability, EEG-based brain-computer interfaces (BCIs) can be used for communication and motor restoration to help paralyzed people [1]. In a BCI system, feature patterns are extracted from EEG time series that are generated by mental activities such as motor imagery and actual movements. The features are then *translated* into control commands for external systems or prosthetic devices *via* a pattern discrimination process.

A variety of feature extraction methods has been proposed in previous researches. Since EEG signals are inherently non-stationary, many studies have employed wavelet-based feature patterns [2]-[4]. The present study utilizes wavelet packet transform (WPT) to extract EEG features, which can provide higher frequency resolution than traditional wavelet transform (WT) methods [5]. For EEG discrimination, the frequency bands containing discriminative features may be different among users. In addition, discrimination may not be achieved only based on low-frequency information, which is focused on by WT methods.

On the other hand, there are a lot of EEG discrimination methods using linear discrimination analysis, support vector machine, neural networks (NNs), and so on [6]. Discrimination methods based on NN can approximate nonlinear mapping between input and output patterns, and they can be trained to adapt for variations among individuals. In particular, recurrent NNs have been investigated in order to enhance discrimination performance by considering information of time context in EEG signals [7], [8]. Haselsteiner *et al.* developed a time-dependent NN, which achieved better performance than traditional multilayer perceptron (MLP) with 4-channel EEG signals for two motor imagery patterns [7]. Also, Tsuji *et al.* proposed a recurrent log-linearized Gaussian mixture network (R-LLGMN) based on hidden Markov model (HMM) [8]. R-LLGMN has been applied to EEG pattern discrimination. In order to improve discrimination rate, many electrodes and high-dimensional frequency information are required for feature vectors. However, when dimensionality of features increases, complexity of the NN increases rapidly. Generally, training process turns to be time-consuming and difficult to converge.

In order to deal with this problem, the authors had proposed a probabilistic neural network for high-dimensional pattern discrimination and applied it to electromyogram (EMG) pattern discrimination [9]. This NN uses orthogonal transformation to project an original feature space into a lower-dimensional one, then calculation of probabilities is conducted with the lower-dimensional features based on a Gaussian mixture model (GMM).

This paper aims to propose a novel probabilistic NN for pattern discrimination of multivariate time series such as EEG signals. A dimension-reducing process is incorporated in a recurrent probabilistic neural network, which is based on GMM and HMM. Structure and learning algorithm of the proposed NN are briefly explained in Section II. Then an EEG discrimination method using the proposed NN is developed in Section III. It is expected that this method can discriminate multivariate EEG features, which are extracted using WPT, with high accuracy. In Section IV, discrimination performance of the proposed NN is evaluated through EEG discrimination experiments. Finally, Section V concludes this paper.

II. NEURAL NETWORK FOR MULTIVARIATE TIME SERIES PATTERN DISCRIMINATION

A. Network Structure

Structure of the proposed NN is shown in Fig. 1. This network is a seven-layer recurrent NN, and consists of a dimension-reducing stage and a time series pattern discrimination stage. This NN calculates posterior probabilities for multivariate time series patterns. Details of each layer are explained as follows.

1) Dimension-reducing Stage: In this stage, the input vectors $\mathbf{x}(t) \in \Re^M$ $(t = 1, 2, \cdots, T)$ are converted into feature vectors with a reduced dimension D, ${}^{(2)}\mathbf{O}(t) \in \Re^D$,

N. Bu is with the Department of Control and Information Systems Engineering, Kumamoto National College of Technology, Koshi, 861-1102 JAPAN (e-mail: bu@ieee.org)

K. Shima and T. Tsuji are with the Graduate School of Engineering, Hiroshima University, Higashi-Hiroshima, 739-8527 JAPAN

978-1-4244-6561-3/10/$26.00 ©2010 IEEE

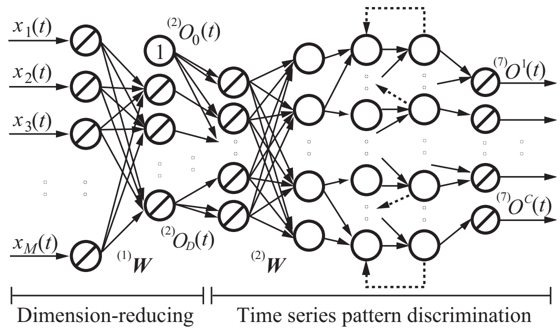

Fig. 1. Structure of the proposed probabilistic neural network.

where M is dimension of the input vectors and $D < M$. The identity function is used for units' activation in the first layer, and the relationships between the input units ${}^{(1)}I_m(t)$ and the output units ${}^{(1)}O_m(t)$ in this layer are defined as

$$ {}^{(1)}O_m(t) = {}^{(1)}I_m(t) = x_m(t). \tag{1} $$

where $x_m(t)$ $(m = 1, 2, \cdots, M)$ is an element of $\mathbf{x}(t)$.

In the second layer, each unit sums up the first layer's outputs, ${}^{(1)}O_m(t)$, weighted by the coefficients ${}^{(1)}\boldsymbol{W} = [{}^{(1)}w_{m,d}] \in \Re^{M \times D}(m = 1, \cdots, M; d = 1, \cdots, D)$. The second layer's inputs and outputs are defined as follows:

$$ {}^{(2)}I_d(t) = \sum_{m=1}^{M} {}^{(1)}O_m(t) {}^{(1)}w_{m,d} \tag{2} $$

$$ {}^{(2)}O_d(t) = {}^{(2)}I_d(t). \tag{3} $$

2) Time Series Pattern Discrimination Stage: This stage is based on the recurrent probabilistic NN, R-LLGMN [8]. It is composed of a Gaussian mixture model (GMM) and a hidden Markov model (HMM), so that it is able to cope with the time-varying characteristics of input features. Please refer to [8] for the correspondence of R-LLGMN to GMM and HMM.

The vector ${}^{(2)}\mathbf{O}(t)$ and an additional constant unit, defined as ${}^{(2)}O_0(t) = 1$, are processed through a non-linear computation and input into the third layer, which consists of $H = 1 + D(D + 3)/2$ units. The non-linear computation is conducted for all possible combinations of d and d',

$$ {}^{(3)}O_h(t) = {}^{(3)}I_h(t) = {}^{(2)}O_d(t) {}^{(2)}O_{d'}(t), \tag{4} $$

where $d' \geq d$ $(d, d' = 0, 1, \cdots, D)$ and $h = \frac{2D-3d+2d'+5}{2}$.

Unit $\{c, k, k', m\}$ $(c = 1, \cdots, C;\ k, k' = 1, \cdots, K_c;\ m = 1, \cdots, M_{c,k})$ in the fourth layer receives outputs of the third layer ${}^{(3)}O_h(t)$ weighted by the coefficient ${}^{(2)}w^c_{k',k,m,h}$. The relationship between the input and the output of the fourth layer is defined as,

$$ {}^{(4)}I^c_{k',k,m}(t) = \sum_{h=1}^{H} {}^{(3)}O_h(t) {}^{(2)}w^c_{k',k,m,h}, \tag{5} $$

$$ {}^{(4)}O^c_{k',k,m}(t) = \exp\Big({}^{(4)}I^c_{k',k,m}(t)\Big), \tag{6} $$

where C is the number of classes, K_c is the number of states in HMM, and $M_{c,k}$ is the number of components of the Gaussian mixture distribution in class c and state k.

The input of a unit $\{c, k, k'\}$ in the fifth layer integrates the output of units $\{c, k, k', m\}$ $(m = 1, 2, \cdots, M_{c,k})$ in the fourth layer, and the output of the sixth layer is also fed back to the fifth layer. These are expressed as follows,

$$ {}^{(5)}I^c_{k',k}(t) = \sum_{m=1}^{M_{c,k}} {}^{(4)}O^c_{k',k,m}(t), \tag{7} $$

$$ {}^{(5)}O^c_{k',k}(t) = {}^{(6)}O^c_{k'}(t-1) {}^{(5)}I^c_{k',k}(t), \tag{8} $$

where ${}^{(6)}O^c_{k'}(0) = 1.0$ is for the initial state.

The sixth layer receives the integrated outputs of unit $\{c, k, k'\}$ $(k' = 1, 2, \cdots, K_c)$. The relationships in this layer are defined as,

$$ {}^{(6)}I^c_k(t) = \sum_{k'=1}^{K_c} {}^{(5)}O^c_{k',k}(t), \tag{9} $$

$$ {}^{(6)}O^c_k(t) = \frac{{}^{(6)}I^c_k(t)}{\sum_{c'=1}^{C} \sum_{k'=1}^{K_{c'}} {}^{(6)}I^{c'}_{k'}(t)}. \tag{10} $$

Finally, unit c in the seventh layer integrates the outputs of K_c units $\{c, k\}$ $(k = 1, 2, \cdots, K_c)$ in the sixth layer.

$$ {}^{(7)}O^c(t) = {}^{(7)}I^c(t) = \sum_{k=1}^{K_c} {}^{(6)}O^c_k(t). \tag{11} $$

Here, ${}^{(7)}O^c(t)$ indicates the posterior probability of class c. In this NN, calculations in the fifth and sixth layers are associated with feedback connections. So that, time-varying features in time series can be used, and it is possible to discriminate multivariate time series using this NN.

B. Learning Algorithm

This study uses a two-step learning method to train the network's weights, ${}^{(1)}\boldsymbol{W}$ and ${}^{(2)}\boldsymbol{W}$, separately. N time series with length T, $\mathbf{x}(t)^{(n)} = [x_1(t)^{(n)}, \cdots, x_M(t)^{(n)}]^{\mathrm{T}}$ $(n = 1, \cdots, N; t = 1, \cdots, T)$, are prepared for each class as training data.

1) Dimension-reducing Stage: Firstly, the weight coefficients ${}^{(1)}\boldsymbol{W}$ are trained in the dimension-reducing stage. It is preferable that dimension-reducing transform in this stage can also enhance class separability to ease time series pattern discrimination in the following stage. In this paper, the weight coefficients ${}^{(1)}\boldsymbol{W}$ are estimated based on linear discriminant analysis (LDA) [10]. Given the training data, an evaluation function is defined to show the class separability,

$$ E_1 = \frac{{}^{(1)}\boldsymbol{W}^{\mathrm{T}} \boldsymbol{S}_b {}^{(1)}\boldsymbol{W}}{{}^{(1)}\boldsymbol{W}^{\mathrm{T}} \boldsymbol{S}_w {}^{(1)}\boldsymbol{W}}, \tag{12} $$

where $\boldsymbol{S}_b$ and $\boldsymbol{S}_w$ are the between-class scatter matrix and the within-class scatter matrix, respectively. These matrices are defined as follows,

$$ \boldsymbol{S}_b = \frac{1}{TC} \sum_{t=1}^{T} \sum_{c=1}^{C} (\overline{\mathbf{x}}_c(t) - \overline{\mathbf{x}}(t))(\overline{\mathbf{x}}_c(t) - \overline{\mathbf{x}}(t))^{\mathrm{T}}, \tag{13} $$

$$ \boldsymbol{S}_w = \frac{1}{TCN} \sum_{t=1}^{T} \sum_{c=1}^{C} \sum_{n=1}^{N} (\mathbf{x}_c(t)^{(n)} - \overline{\mathbf{x}}_c(t))(\mathbf{x}_c(t)^{(n)} - \overline{\mathbf{x}}_c(t))^{\mathrm{T}} \tag{14} $$

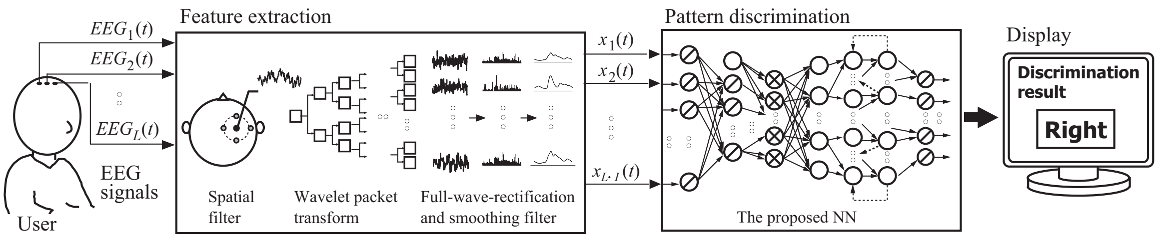

Fig. 2. A diagram of the proposed EEG discrimination method.

$$\overline{\mathbf{x}}_c(t) = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_c(t)^{(n)}, \tag{15}$$

$$\overline{\mathbf{x}}(t) = \frac{1}{CN}\sum_{c=1}^{C}\sum_{n=1}^{N}\mathbf{x}_c(t)^{(n)}, \tag{16}$$

where $\mathbf{x}_c(t)^{(n)}$ is the input vector belonging to class c. The learning process finds ${}^{(1)}\boldsymbol{W}$ that maximizes E_1.

2) Time Series Pattern Discrimination Stage: A teacher vector $\boldsymbol{T}^{(n)} = [T_1^{(n)}, \cdots, T_c^{(n)}, \cdots, T_C^{(n)}]^{\mathrm{T}}$ is defined for training data $\mathbf{x}(t)^{(n)}$. If $\mathbf{x}(t)^{(n)}$ is for class c, then $T_c^{(n)} = 1$ and $T_{\hat{c}}^{(n)} = 0$ $(\hat{c} \neq c)$ for all the other classes. The energy function for weight ${}^{(2)}\boldsymbol{W}$ is defined as

$$E_2 = -\sum_{n=1}^{N}\sum_{c=1}^{C} T_c^{(n)} \log {}^{(7)}O^c(T)^{(n)} \tag{17}$$

where ${}^{(7)}O^c(T)^{(n)}$ is the posterior probability at time T for the time series $\mathbf{x}(t)^{(n)}$. The learning process is to minimize E_2, that is, to maximize the likelihood [8].

III. EEG DISCRIMINATION METHOD USING THE PROPOSED NEURAL NETWORK

Fig. 2 depicts the proposed EEG discrimination method. After spatial filtering for noise and artifact removal, EEG signals are decomposed into frequency subbands of each channel using wavelet packet transform (WPT). These multivariate EEG feature series are then input into the proposed reduced-dimensional recurrent NN to discriminate two metal activities.

Electrodes are attached to the scalp according to the international 10-20 system. Suppose L electrodes are used for signal acquisition, a Laplacian-filtered signal $EEG'_l(t)$ $(l = 1, 2, \cdots, L)$ is defined as [11],

$$EEG'_l(t) = EEG_l(t) - \frac{1}{4}\sum_{u \in S_l} EEG_u(t), \tag{18}$$

where $EEG_l(t)$ represents the signal measured from the electrode l, while $EEG_u(t)$ $(u \in S_l)$ stands for the signal measured from one of four neighboring electrodes of electrode l (Here, S_l is the index set of neighboring electrodes). Using Eq. (18), it is possible to eliminate artifacts such as electromyogram and electro-oculogram signals, and to increase the signal-to-noise ratio [11].

Then, $EEG'_l(t)$ is decomposed using wavelet packet transform. The coefficients of WPT at jth level and vth sample can be described as

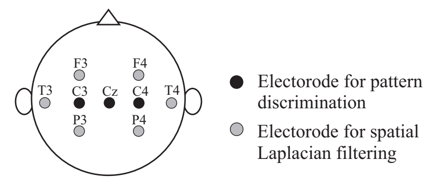

Fig. 3. The electrode placement used in this study.

$$d_l^{(j,2i-1)}(v) = \sum_s p(s-2v)\, d_l^{(j-1,i)}(s) \tag{19}$$

$$d_l^{(j,2i)}(v) = \sum_s q(s-2v)\, d_l^{(j-1,i)}(s) \tag{20}$$

where $p(v)$ and $q(v)$ represent coefficients of the wavelet function and its scaling function, respectively. The decomposition is achieved in a recursive fashion. Suffix i of $d_l^{(j,i)}(v)$ indicates the ith frequency subbands of level j. The number of frequency subbands in the jth level is 2^j. Defining the decomposition level as J, I channels of wavelet coefficients are selected from 2^J series $(I \leq 2^J)$. These data are resampled to map the wavelet coefficients on temporal axis, and are defined as $d_l^{(J,i)}(t)$ $(i = 1, \cdots, I;\ l = 1, \cdots, L)$. Full-wave rectification for each $d_l^{(J,i)}(t)$ is carried out, and the signals are then smoothed using a second-order Butterworth filter (cut-off frequency: f_c Hz). In addition, normalized signals $\tilde{d}_l^{(J,i)}(t)$, with its range as [0, 1], are converted into input vectors $\mathbf{x}(t) \in \Re^M$ $(M = L \cdot I)$.

IV. EXPERIMENTS

EEG discrimination experiments were conducted to verify the proposed method with four healthy subjects. EEG signals were measured with Ag/AgCl type electrodes, and nine electrodes were located at the motor area (see Fig. 3). EEG signals from channels C3, Cz, and C4 $(L = 3)$ were put through a spatial Laplacian filter. In the experiments, subjects were instructed to perform three tasks, (a) **Right-finger movements**: bending and stretching of the right index finger for ten sec, and resting for ten sec, (b) **Left-finger movements**: bending and stretching of the left index finger for ten sec, and resting for ten sec, and (c) **No movement**: be at rest for twenty sec. For each task 30 trials were recorded. The sampling frequency for EEG acquisition was 1 kHz.

In this study, Daubechies' wavelet was used [5]. The decomposition level in WPT J was set as 7, f_c was 0.1, and

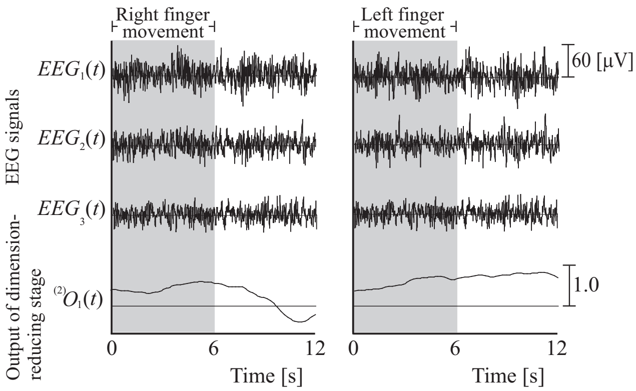

Fig. 4. Examples in the finger movement discrimination experiments (subject A; right finger movement vs. left finger movement).

TABLE I

DISCRIMINATION RESULTS OF FOUR SUBJECTS. (LEFT: LEFT FINGER MOVEMENT, RIGHT: RIGHT FINGER MOVEMENT, NO: NO MOVEMENT)

Sub.	Finger movements		
	Right - Left	Right - No	Left - No
A	73.2 ± 5.8	89.8 ± 8.0	95.2 ± 3.2
B	94.5 ± 3.3	93.9 ± 4.9	85.9 ± 5.6
C	72.1 ± 7.1	65.1 ± 10.4	74.2 ± 9.5
D	70.4 ± 7.5	71.3 ± 7.3	57.3 ± 9.2

Average ± S.D. [%]

the input vectors of the proposed NN were down-sampled to 5 [Hz]. Furthermore, among the 128 frequency subbands extracted from EEG signals, ten subbands corresponding to the frequency range of [0, 40] Hz were defined as the NN's input vector. For the NN structure, the number of reduced dimensions was set as 1 $(D = 1)$. Parameters in the time series pattern discrimination stage were set as follows: $C = 2$, $K_c = 2$ $(c = 1, 2)$, and $M_{c,K_{\acute{c}}} = 2$ $(c, \acute{c} = 1, 2)$.

Fig. 4 shows examples of right and left finger movements discrimination of subject A. The measured EEG signals and the output of dimension-reducing stage, $^{(2)}O_1(t)$, are shown in this figure. The gray areas indicate the time periods, during which the subject performed hand movements. Although, it is difficult to find significant difference of patterns directly from the EEG signals, it can be seen that outputs of the dimension-reducing stage decreased when right-finger movement was stopped. It was considered that such time-varying features in EEG signals are possible to enhance the time series discrimination.

Discrimination experiments were also conducted for right finger movements vs. no movement and left finger movements vs. no movement. Table I summarizes the discrimination results of all subjects. Mean values and standard deviations (S.D.) of the discrimination rates are calculated using 50 different sets of training data and test data, while the initial weights in the proposed NN were changed 10 times randomly in each trial. Tasks with the highest discrimination rates were marked with grey backgrounds for each subject. From the table, it can be found that the proposed method can realize relative high discrimination performance. Subjects A and B achieved discrimination rates close to or higher than 90%. On the other hand, it should be noted that subjects C and D had no experience in EEG discrimination and only had 30 min training before the experiments.

V. CONCLUSIONS

This paper aims to develop a pattern discrimination approach for BCI applications based on a novel probabilistic NN. This NN introduces the functions of dimensional reduction and time series pattern discrimination in its structure as a dimension-reducing stage and a time series pattern discrimination stage, respectively. With the proposed NN, discrimination for multivariate time series can be conducted with improved accuracy. In addition, an EEG discrimination method was developed based on the proposed NN and feature extraction using a spatial Laplacian filter and the wavelet packet transform. EEG discrimination experiments were carried out with three metal activities related to finger movements. The experimental results demonstrated the validity of the proposed method.

In the future research, we would like to improve the discrimination performance by enhancing the NN's structure and the learning algorithm. Weight coefficients in both stages of the proposed NN should be trained with a single criterion. Also, additional experiments are needed to evaluate the proposed method over more subjects. Furthermore, electrode placement and EEG signal preprocessing should be investigated.

REFERENCES

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller and T.M. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophysiol.* vol. 113, no. 6, 2002, pp. 767-791.

[2] N. Hazarika, J.Z. Chen, A.C. Tsoi and A. Sergejew, "Classification of EEG signals using the wavelet transform," *Signal Process.*, vol. 59, no. 1, 1997, pp 61-72.

[3] C.-W. Chen, C.-C.K. Lin and M.-S. Ju, "Detecting movement-related EEG change by wavelet decomposition-based neural networks trained with single thumb movement," *Clin. Neurophysiol.*, vol. 118, no. 4, 2007, pp. 802-814.

[4] D. Vautrin, X. Artusi, M.F. Lucas, D. Farina, "A novel criterion of wavelet packet best basis selection for signal classification with application to brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 11, 2009, pp. 2734-2738.

[5] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way, third edition*, Academic Press, Burlington, MA; 2008.

[6] F. Lotte, M. Congedo, A. Lècuyer, F. Lamarche and B. Arnaldi, "A review of classification algorithms for EEG-based brain-computer interfaces," *J. Neural Eng.* vol. 4, no. 2, 2007, pp. R1-R13.

[7] E. Haselsteiner and G. Pfurtscheller, "Using time-dependent neural networks for EEG classification," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 4, 2000, pp. 457-463.

[8] T. Tsuji, N. Bu, M. Kaneko and O. Fukuda, "A recurrent log-linearized Gaussian mixture network," *IEEE Trans. Neural Network.*, vol. 14, no. 2, 2003, pp. 304-316.

[9] T. Tsuji, N. Bu, J. Arita, M. Ohga, "A speech synthesizer using facial EMG signals," *Int. J. Comput. Intell. Appl.*, vol. 7, no. 1, 2008, pp. 1-15.

[10] C.M. Bishop, *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford; 1995.

[11] D.J. McFarland, L.M. McCane, S.V. David, and J.R. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroencephalogr. Clin. Neurophysiol.*, vol. 103, no. 3, 1997, pp. 386-394.