INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. 2000; **14**:829–848

Pattern classification of time-series EMG signals using neural networks

Toshio Tsuji[1,*,†] Osamu Fukuda[1], Makoto Kaneko[1] and Koji Ito[2]

[1]*Department of Industrial and Systems Engineering, Hiroshima University, 1-4-1 Kagamiyana, Higashi Hiroshima shi 739 8527, Japan*
[2]*Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Japan*

SUMMARY

This paper proposes a pattern classification method of time-series EMG signals for prosthetic control. To achieve successful classification for non-stationary EMG signals, a new neural network structure that combines a common back-propagation neural network with recurrent neural filters is used. A convergence time of the network learning can be regulated by a new learning method based on dynamics of a terminal attractor. The experiments of pattern classification and prosthetic control are carried out for several subjects including an amputee. It is shown from the results that the proposed method improves learning/classification ability for stationary and non-stationary EMG signals during a series of continuous motions. Copyright © 2000 John Wiley & Sons, Ltd.

1. INTRODUCTION

Since every motion of a human operator is realized by muscular contraction controlled by the central nervous system (CNS), EMG signals accompanied by muscular contraction involve information on muscles contributing to the motion. If the information on human operator's intended motion can be extracted from the EMG signals, it may be used as a new interface tool for virtual reality and teleoperation devices, or as a communication tool for a handicapped person. For example, in the case of a physically handicapped person who has his or her upper limb amputated by an accident, if the CNS and a part of the muscles which have actuated the original limb still remained after amputation, it can be expected that the natural feeling of prosthetic control similar to that of the original limb is realized using the EMG signals.

Until now, many investigations of EMG pattern classification were carried out for prosthetic control. In the early days, a linear prediction model for EMG signals such as an autoregressive (AR) model [1–4] was frequently used. However, it is very difficult to achieve high classification performance stably, especially for rapid movements, because of non-linear characteristics and

* Correspondence to: Toshio Tsuji, Department of Industrial and Systems Engineering, Hiroshima University, 1-4-1 Kagamiyana, Higashi Hiroshima shi 739 8527, Japan
† E-mail: tsuji@huis.hiroshima-u.ac.jp

Received 8 February 2000
Accepted 5 April 2000

large variability of the EMG signals. The non-linear characteristics are caused by interference between different muscles, and changes of signal sources and paths to recording electrodes. Also, the EMG signals change largely depending on muscle fatigue, sweating and changes of electrode location.

On the other hand, several EMG pattern classification methods using neural networks have been proposed [5–9]. The neural networks can acquire any non-linear mapping of training data through learning. For example, Kelly *et al.* [5] proposed a pattern classification method which combines a back-propagation neural network [10] and a Hopfield neural network. This method can acquire a mapping from the EMG patterns measured from one pair of electrodes to four motions of elbow and wrist joints. Also, Hiraiwa *et al.* [6] used a back-propagation neural network for estimation of five finger motions. They reported that the five finger motions, joint torque and angles were successfully estimated. Koike *et al.* [8] constructed a forward dynamics model of the human arm between EMG signals and arm trajectories in their form of neural network. In their experiments, four joint angles, one at the elbow and three at the shoulder, were estimated from surface EMG signals of 12 flexor and extensor muscles during posture control in 3D space. However, in the case of EMG pattern classification using common back-propagation neural networks, the networks need a large number of training data and learning iterations as well as a large scale of the structure [5–8].

Then, Tsuji *et al.* [9] proposed a probabilistic neural network for the EMG pattern classification problem in order to improve the back-propagation learning. This network can construct the statistical model of the EMG signals through learning, and improve classification ability. However, this network does not take into account time-varying characteristics of the time-series EMG signals, so that the classification performance may decrease depending on the non-stationarity of the EMG signals. For time-series EEG signals, we have proposed a pattern classification method utilizing a new network structure which combines a probabilistic neural network and a recurrent neural filter [11–14]. The time-varying characteristics can be taken into account by the neural filter, and this refinement improves classification ability. This study, however, dealt with a few of very simple classification problems which include only two states such as eye opening/closing and turning on/off of the artificial flash light in front of the eyes. Also it was difficult for the network to converge the learning process because of the complex network to converge the learning process because of the complex network structure.

The purpose of this paper is classification of non-stationary EMG signals during the continuous motions in a short-time period. It is too difficult for only one network to learn non-stationary EMG signals because of the local minima problem, accuracy of the classification and so on. Therefore, signal processing is divided into two parts: (1) pattern processing in each time interval, and (2) filtering based on the time history of the pattern series. Then, corresponding to two parts, a new network structure which combines the BPN and the NF is proposed. As a result, the whole process can be executed efficiently although the structure may become rather complicated. For the learning of the proposed network structure, this paper invents a new learning method which is called *terminal learning* in order to make the learning process smooth. The dynamics of the energy function always converges stably to the equilibrium point in finite time using this method. Also, it is possible to synchronize the learning process of multiple networks. The convergence time is always less than the prespecified upper limit, so that the mental stress of the operator may be reduced while waiting for the convergence. Moreover, the entropy of the network output is defined and used in order to suspend the classification for the network output with high entropy, which results in reduction of misclassification.

This paper is organized as follows: The structure of the proposed network and the terminal learning method are explained in Section 2. The experiments of pattern classification and prosthetic control are conducted in Section 3 using two kinds of time-series EMG signals: one is stationary EMG signals which are measured while keeping one motion, and the other is non-stationary EMG signals measured while changing motions continuously. Finally Section 4 concludes the paper.

2. PATTERN CLASSIFICATION METHOD

2.1. A network structure

Figure 1 shows the structure of the proposed network, which consists of a pre-processing part, a back-propagation neural network (BPN), a neural filter (NF) and a discrimination part. First, the EMG signal $\mathrm{EMG}_i(t)$ $(i = 1, \ldots, L)$ is pre-processed and converted into the input feature vector $\mathbf{X}(n) \in \Re^L$ $(n = 1, \ldots, N)$. Next, the BPN receives it and outputs ${}^{(D)}\mathbf{O}(n) \in \Re^K$ $(n = 1, \ldots, N)$. Then, the NF modifies this output taking into account the time-varying characteristics of the EMG signals. The outputs of the NF are normalized to make the sum of K outputs equal 1. Thus, they are considered as the posteriori probability. Finally, the Bayes decision rule is used to determine a specific class that corresponds to the most-likely motion intended by a human operator out of K candidates of the motions.

(1) *Pre-processing part:* The time-series EMG signals $\mathrm{EMG}_i(t)$ $(i = 1, \ldots, L)$ measured from the L electrodes are digitized by an A/D converter (sampling frequency, 1 kHz; and quantization, 12 bits) after they are amplified, rectified and filtered out through the Butterworth filter (cut-off frequency: f_{cut} Hz). Then these digitized signals are sampled every $t_{\mathrm{d}} = 10$ ms and normalized to make the sum of L channels equal 1. The sampling data are used as the input feature vector $\mathbf{X}(n) = [X_1(n), X_2(n), \ldots, X_L(n)]^{\mathrm{T}}$ for the BPN.

It should be noted that the characteristics of the input feature vector change depending on the cut-off frequency of the Butterworth filter in pre-processing part, which influences the network learning and classification. Therefore, the cut-off frequency must be settled appropriately in order to extract the time-varying characteristics of the time-series EMG signals.

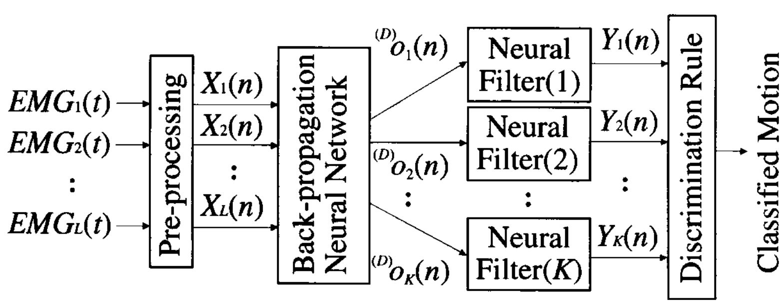


Figure 1. Structure of the proposed network.

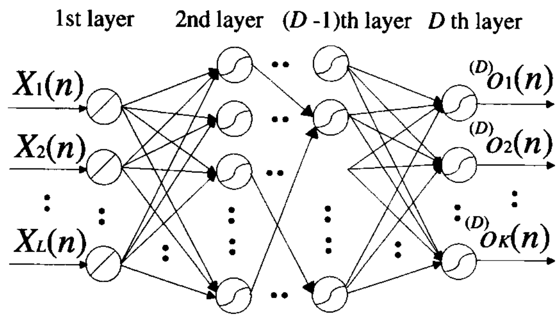


Figure 2. The back-propagation neural network.

(2) *Back-propagation neural network:* Figure 2 shows the structure of the BPN, which is of a feedforward type with D layers. The input feature vectors at each time are classified one by one using this network. The ith layer consists of M_i units for $i = 1, 2, \ldots, D$. The number of units in the first layer and the Dth layer are $M_1 = L$, $M_D = K$. In the first layer, the identity function is used for the activation function of each unit.

The units in the first layer receive the nth input feature vector $\mathbf{X}(n)$ as ${}^{(1)}\mathbf{S}(n)$ and outputs the same value as ${}^{(1)}\mathbf{o}(n)$. The input/output of the ith unit in the dth layer are defined as

$$ {}^{(d)}s_i(n) = \sum_{j=1}^{M_{d-1}} {}^{(d-1,d)}W_{ij}^{(d-1)}o_j(n) \tag{1} $$

$$ {}^{(d)}o_i(n) = \frac{1}{1 + \exp(-\,{}^{(d)}s_i(n))} \tag{2} $$

where ${}^{(d)}s_i(n)$ and ${}^{(d)}o_i(n)$ $(d = 2, 3, \ldots, D)$ denotes the input/output of the ith unit in the dth layer, and ${}^{(d-1,d)}W_{ij}$ denotes the weight coefficient between the jth unit in the $(d-1)$th layer and the ith unit in the dth layer.

(3) *Neural filter:* The outputs of the BPN may vary largely because of fluctuation of the EMG. Therefore, the NF is connected to the Dth layer of the BPN in order to take into account the time history of the EMG signals. The NFs receive the output of the BPN and makes it smoother. The characteristics of the NF can be changed flexibly through learning. This is a different feature from the conventional digital filters.

In order to deal with the time-series signals, several kinds of network structure may be suitable for the NF. In our case, the NF deals with a simple signal processing with only one input/output data, so the simple and compact structure is desirable. Lo [14] used a recurrent multilayer perceptron (RMLP) with one hidden layer of fully interconnected neurons for neural filtering. He called such a RMLP the *neural filter*. Simulation results showed that the NF with only a few hidden neurons consistently outperforms the extended Kalman filter. This NF is incorporated into the proposed network.

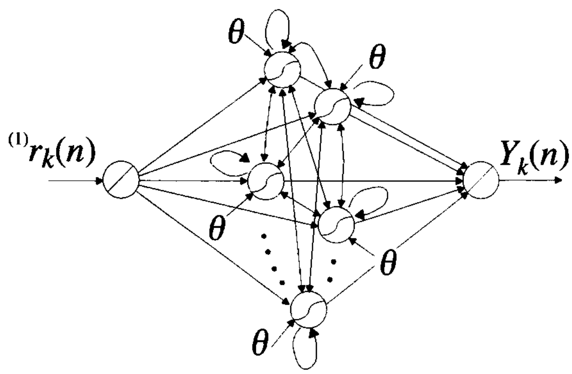


Figure 3. The neural filter.

The structure of the neural filter is shown in Figure 3 [14]. The unit in the first layer receives the nth output $^{(D)}o_k(n)$ of the BPN, and send $^{(1)}v_k(n)$ to the second layer. The identity function is used for the activation function in the first layer.

The second layer consists B fully interconnected units. Each unit in the second layer receives the nth output of the first layer and the $(n-1)$th output of the second layer. Also, each unit in this layer has the bias input ($\theta = 1$). This layer keeps the internal representation so that the time history of the input data can be taken into consideration. The input to the unit b, $^{(2)}r_k^b(n)$, and the output, $^{(2)}v_k^b(n)$, are defined as

$$^{(2)}r_k^b(n) = \sum_{a=1}^{B} {}^{(2,2)}u_k^{ab(2)}v_k^a(n-1) + {}^{(1,2)}u_k^{b(1)}v_k(n) + {}^{(\theta)}u_k^b \tag{3}$$

$$^{(2)}v_k^b(n) = \frac{1}{1+\exp(-{}^{(2)}r_k^b(n))} \tag{4}$$

where $^{(2,2)}u_k^{ab}$, $^{(1,2)}u_k^b$ and $^{(\theta)}u_k^b$ denote the weight coefficients between the ath and the bth unit in the second layer, between the unit in the first layer and the bth unit in the second layer, and between the bias input and the bth unit in the second layer, respectively. The unit in the third layer receives the output of all the units in the second layer. The input $^{(3)}r_k(n)$ and the output $Y_k(n)$ of the unit is defined as

$$^{(3)}r_k(n) = \sum_{b=1}^{B} {}^{(2,3)}u_k^{b(2)}v_k^b(n) \tag{5}$$

$$Y_k(n) = \frac{1}{1+\exp(-{}^{(3)}r_k(n))} \tag{6}$$

where $^{(3)}r_k(n)$ and $Y_k(n)$ denote the input and the output in the third layer, and $^{(2,3)}u_k^b$ denotes the weight coefficient between the bth unit in the second layer and the unit in the third layer.

(4) *Discrimination part:* First, the K outputs of the NF are normalized to make the sum of the outputs equal to 1 and considered as the *a posteriori* probability. Then the entropy is calculated from this posteriori probability, and the classification is performed using this entropy [7]. When the BPN receives an input feature vector $\mathbf{X}(n)$ and each NF outputs $Y_k(n)$ $(k = 1, 2, \ldots, K)$, the entropy $H_{\mathrm{nf}}(n)$ is defined as

$$H_{\mathrm{nf}}(n) = -\sum_{k=1}^{K} p(k \mid \mathbf{X}(n)) \log p(k \mid \mathbf{X}(n)) \tag{7}$$

$$p(k \mid \mathbf{X}(n)) = \frac{Y_k(n)}{\sum_{k'=1}^{K} Y_{k'}(n)} \tag{8}$$

The entropy indicates, or may be interpreted as, a risk of misclassification. For example, if the entropy is above the determination threshold θ_{d}, the determination should be suspended since large entropy means that the network output is ambiguous. On the other hand, if the entropy is less than θ_{d}, the Bayes decision rule is used to determine the specific class. Thus, possible misclassification is expected to be reduced.

2.2. Learning rule

In this section, the learning algorithm of the proposed network is explained. If the teacher signal is given only to the output units in the NF, the error may *back-propagate* from the NF to the BPN, so that the learning is performed for both the networks at the same time. However, the appropriate error back-propagation between the NF and the BPN cannot be guaranteed because of a serial structure of two networks. Therefore, we introduce the following two-step learning schedule. The learning is efficiently divided into the BPN and the NF so that the function of each network becomes clear.

In the case of the back-propagation learning based on the steepest descent, the operator cannot predict the convergence time of learning in advance, so that the mental stress of the operator while waiting for the convergence may be increased. Therefore, this paper newly develops the terminal learning method in order to make the learning process smooth. In the terminal learning, the dynamics of the energy function always converges stably to the equilibrium point in finite time. The equilibrium point is a kind of terminal attractor which was discovered by Zak [15]. In the terminal learning, the upper limit of the convergence time can be prespecified. This reduces the mental stress of the operator while he/she is waiting for the learning convergence. Also, the learning of multiple networks can be synchronized roughly using this method.

(1) *Learning schedule:* First, the BPN is trained using the training data which are sampled from the stationary EMG signals. The teacher signal $T_{nn}^{k}(n)$ $(k = 1, \ldots, K)$ is given for each output unit corresponding to the operator's motion.

After the learning of the BPN, another set of the input feature vectors $\mathbf{X}(n)$ is given and the BPN outputs ${}^{(D)}o_k(n)$ $(k = 1, \ldots, K)$, where the weight coefficients in the BPN are not modified. Then, each NF is trained using this output data and the teacher signal $T_{\mathrm{nf}}^{k}(n)$ $(k = 1, \ldots, K)$ given for each output unit in order to construct a kind of the digital filter.

(2) *Learning rule of the BPN:* The teacher vector $\mathbf{T}_{nn}(n) = [T^1_{nn}(n), \ldots, T^k_{nn}(n), \ldots, T^k_{nn}(n)]^{\mathrm{T}}$ is prepared for the nth feature vector $\mathbf{X}(n)$. In order to speed up the learning, the teacher signal is given: $T^k_{nn}(n) = 1.1$ for the particular unit and $T^k_{nn}(n) = -0.1$ for all the other units. As an energy function of the network learning, we use

$$E_{nn} = \frac{1}{2} \sum_{n=1}^{N} \sum_{k=1}^{K} ({}^{(D)}o_k(n) - T^k_{nn}(n))^2 \tag{9}$$

and the learning is performed to minimize it. Each weight ${}^{(d-1,d)}W_{ij}$ is considered as a time dependent continuous variable and its time derivative is defined as

$${}^{(d-1,d)}\dot{W}_{ij} = -\eta_{nn}\gamma_{nn} \frac{\partial E_{nn}}{\partial^{(d-1,d)}W_{ij}} \tag{10}$$

$$\gamma_{nn} = \frac{E^{\alpha}_{nn}}{\sum_{d=2}^{D} \sum_{i=1}^{M_d} \sum_{j=1}^{M_{d-1}} \left(\dfrac{\partial E_{nn}}{\partial^{(d-1,d)}W_{ij}} \right)^2} \tag{11}$$

where $\eta_{nn} > 0$ is a positive learning rate and $\alpha (0 < \alpha < 1)$ is a constant. The time derivative of the energy function E_{nn} can be calculated as

$$\begin{aligned} \dot{E}_{nn} &= \sum_{d=2}^{D} \sum_{i=1}^{M_d} \sum_{j=1}^{M_{d-1}} \left(\frac{\partial E_{nn}}{\partial^{(d-1,d)}W_{ij}} {}^{(d-1,d)}\dot{W}_{ij} \right) \\ &= -\eta_{nn} E^{\alpha}_{nn} \leqslant 0. \end{aligned} \tag{12}$$

It can be seen from (12) that E_{nn} is a monotonically non-increasing function, and it always converges stably to the equilibrium point (the global minimum or one of local minima). In this case, the convergence time $t_{\mathrm{f}nn}$ can be calculated as

$$\begin{aligned} t_{\mathrm{f}nn} &= \int_0^{t_{\mathrm{f}nn}} \mathrm{d}t = \int_{E_{nn}(0)}^{E_{nn}(t_{\mathrm{f}nn})} \frac{\mathrm{d}E}{\dot{E}} \\ &= \frac{E_{nn}(0)^{1-\alpha} - E_{nn}(t_{\mathrm{f}nn})^{1-\alpha}}{\eta_{nn}(1-\alpha)} \\ &\leqslant \frac{E_{nn}(0)^{1-\alpha}}{\eta_{nn}(1-\alpha)} \end{aligned} \tag{13}$$

where $E_{nn}(0)$ is the initial value of the energy function E_{nn} calculated using initial weights, and $E_{nn}(t_{\mathrm{f}nn})$ is the final value of E_{nn} at the equilibrium point.

In this paper, learning is performed by a discrete form derived from (10):

$$ {}^{(d-1,d)}W_{ij}(t+\Delta t_{nn}) = {}^{(d-1,d)}W_{ij}(t) + \frac{\Delta t_{nn}}{2}\left({}^{(d-1,d)}\dot{W}_{ij}(t) + {}^{(d-1,d)}\dot{W}_{ij}(t+\Delta t_{nn})\right) \tag{14} $$

Weight modification is carried out at every sampling time Δt_{nn}.

Also the entropy calculated from the output of the BPN is used during the learning in order to speed it up. The entropy $H_{nn}(n)$ can be derived in the same manner as (7), (8):

$$ H_{nn}(n) = -\sum_{k=1}^{K} p(k \mid \mathbf{X}(n)) \log p(k \mid \mathbf{X}(n)) \tag{15} $$

$$ p(k \mid \mathbf{X}(n)) = \frac{{}^{(D)}O_k(n)}{\sum_{k'=1}^{K} {}^{(D)}O_{k'}(n)} \tag{16} $$

The network learning is continued until all of the entropy is less than the convergence threshold θ_ε and there is no misclassification for all training data.

(3) *Learning rule of the NF:* Next, the terminal learning of the NF is explained. The energy function for the kth NF is defined as

$$ {}^{(k)}E_{\mathrm{nf}} = \frac{1}{2}\sum_{n=1}^{N} (Y_k(n) - T_{\mathrm{nf}}^{k}(n))^2 \tag{17} $$

where $T_{\mathrm{nf}}^{k}(n)$ is the teacher signal for the output of the kth NF. The teacher signal is given: $T_{\mathrm{nf}}^{k}(n) = 1.0$ for the particular unit and $T_{\mathrm{nf}}^{k}(n) = 0.0$ for all the other units. The learning is performed to minimize this sum of the squared errors. The weight coefficients between the second and third layers are modified using the error back-propagation learning. The time derivative of the weight coefficient ${}^{(2,3)}\dot{u}_k^b$ can be derived as

$$ {}^{(2,3)}\dot{u}_k^b = -\eta_{\mathrm{nf}}\gamma_{\mathrm{nf}} \frac{\partial\,{}^{(k)}E_{\mathrm{nf}}}{\partial\,{}^{(2,3)}u_k^b} \tag{18} $$

$$ \gamma_{\mathrm{nf}} = \frac{{}^{(k)}E_{\mathrm{nf}}^{\beta}}{\psi} \tag{19} $$

$$ \psi = \sum_{b=1}^{B}\left(\frac{\partial\,{}^{(k)}E_{\mathrm{nf}}}{\partial\,{}^{(2,3)}u_k^b}\right)^2 + \sum_{a=1}^{B}\sum_{b=1}^{B}\left(\frac{\partial\,{}^{(k)}E_{\mathrm{nf}}}{\partial\,{}^{(2,2)}u_k^{ab}}\right)^2 + \sum_{b=1}^{B}\left(\frac{\partial\,{}^{(k)}E_{\mathrm{nf}}}{\partial\,{}^{(1,2)}u_k^b}\right)^2 + \sum_{b=1}^{B}\left(\frac{\partial\,{}^{(k)}E_{\mathrm{nf}}}{\partial\,{}^{(\theta)}u_k^b}\right)^2 \tag{20} $$

where $\eta_{\mathrm{nf}} > 0$ denotes the learning rate and $\beta (0 < \beta < 1)$ is a constant.

For the weight coefficients in the first and second layers, the error back-propagation through time [10] is used because of the interconnection in the second layer. In this case, the time derivative of the weight coefficient ${}^{(2,2)}\dot{u}_k^{ab}$ in the second layer can be calculated as follows:

$$
{}^{(2,2)}\dot{u}_k^{ab} = -\eta_{\mathrm{nf}}\gamma_{\mathrm{nf}} \frac{\partial^{(k)}E_{\mathrm{nf}}}{\partial^{(2,2)}u_k^{ab}} \tag{21}
$$

Note that the time history of the input data is considered back to C time steps for computation of $\partial^{(k)}E_{\mathrm{nf}}/\partial^{(2,2)}u_k^{ab}$. The time derivative of the weight coefficient ${}^{(1,2)}\dot{u}_k^{b}$, ${}^{(\theta)}\dot{u}_k^{b}$ can be derived in the same manner as (21). The time derivative of the energy function ${}^{(k)}E_{\mathrm{nf}}$ can be calculated as

$$
\begin{aligned}
{}^{(k)}\dot{E}_{\mathrm{nf}} &= \sum_{b=1}^{B}\left(\frac{\partial^{(k)}E_{\mathrm{nf}}}{\partial^{(2,3)}u_k^{b}}{}^{(2,3)}\dot{u}_k^{b}\right) + \sum_{a=1}^{B}\sum_{b=1}^{B}\left(\frac{\partial^{(k)}E_{\mathrm{nf}}}{\partial^{(2,2)}u_k^{ab}}{}^{(2,2)}\dot{u}_k^{ab}\right) \\
&\quad + \sum_{b=1}^{B}\left(\frac{\partial^{(k)}E_{\mathrm{nf}}}{\partial^{(1,2)}u_k^{b}}{}^{(1,2)}\dot{u}_k^{b}\right) + \sum_{b=1}^{B}\left(\frac{\partial^{(k)}E_{\mathrm{nf}}}{\partial^{(\theta)}u_k^{b}}{}^{(\theta)}\dot{u}_k^{b}\right) \\
&= -\eta_{\mathrm{nf}}E_{\mathrm{nf}}^{\beta} \leqslant 0
\end{aligned} \tag{22}
$$

The convergence time $t_{f\mathrm{nf}}$ can be easily calculated (see (13)), and the weight modification is carried out by a discrete form derived from (18), (21) at every sampling time Δt_{nf}.

3. EXPERIMENTS

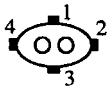
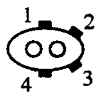
3.1. Experimental conditions

The experiments of pattern classification and prosthetic control are performed for four subjects: one is an amputee, and the others are normal. The amputee was amputated at the forearm, 6 cm from the left wrist joint, about six months ago. The BPN has $D = 3$ layers, and the second layer of the NF consists of $B = 8$ units. The parameters of the terminal learning are set as $t_{fnn} = t_{f\mathrm{nf}} = 1$, $\Delta t_{nn} = \Delta t_{\mathrm{nf}} = 0.001$, $\alpha = \beta = 0.7$, respectively.

The time-series EMG signals are measured from $L = 4$ dry-type electrodes (Imasen lab.). The electrode positions are shown in Table I, and the EMG signals were measured by the bipolar derivation method. When the proposed network is used in a practical situation, it is difficult for the operator to attach the electrodes on the selected muscles. Therefore, in this paper, the locations where the electrodes are attached are not specified. For variable locations of the electrodes, the proposed system can adapt itself using the neural network. Possible electrode locations for EMG measurement may be limited for the case of amputees because of the damage level of the muscles. Considering this point, for the normal subjects, that is the pseudo-amputees, some experiments were carried out under the electrode locations shifted at the unilateral (see Table I) in order to simulate the situation in which the cross-type location is not available. On the other hand, for the amputee used in the experiment, the cross-type electrode locations were available, so there was no need for us to try any other electrode locations.

The experiments are conducted using two kinds of time-series EMG signals: one is stationary EMG signals and the other non-stationary ones. In the former case, the subject was asked to keep one motion, and the stationary EMG signals were measured. On the other hand, in the latter case,

Table I. Electrode locations and the measured muscles.

Electrode locations	No. 1	No. 2	No. 3	No. 4
	Extensor Digitorum Communis	Extensor Carpi Ulnaris	Flexor Carpi Radialis	Palmaris Longus
	Extensor Digitorum Communis	Extensor Digitorum Profundus	Flexor Carpi Ulnaris	Flexor Carpi Radialis

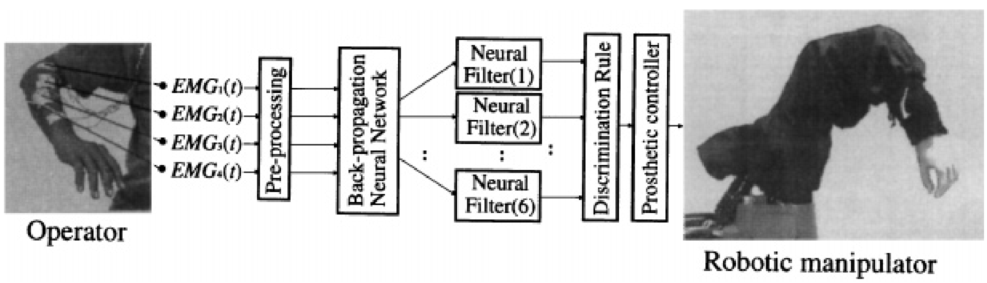


Figure 4. Control system for prosthetic hand.

the subject was asked to change the motions continuously in approximately four second time periods, while non-stationary EMG signals were measured. The purpose of the experiments was explained to all the subjects, and the experiments were carried out with the content of the subjects. This paper deals with the classification of non-stationary EMG signals during the six specified motions. First, in Section 3.2, the experiments for stationary EMG signals are carried out in order to regulate the network and learning parameters. Then, in Section 3.3, the time history of the classification results is taken into consideration using the NF in order to improve the classification performance. Also, comparison experiments are carried out using the conventional BPN and the proposed method in order to make the difference clear between two methods.

Figure 4 shows the prosthetic control system used in the experiments, which consists of the pre-processing part, the BPN, the NF, the discrimination part, the robotic manipulator [16] and its controller. The robotic manipulator has eight degrees of freedom and three of them are controlled according to the output of the discrimination part, which results in six different motions, that is, wrist flexion, wrist extension, wrist pronation, wrist supination, hand grasping and hand opening. The end-effector of the robotic manipulator is a prosthetic hand (Imasen lab.), which is almost of the same size as the human hand. It is driven by the ultra sonic motors (SHINSEI Corp.) so that the motor noise is significantly reduced.

The mean values and the standard deviations of the classification results for 10 kinds of initial weights are shown in the following figures and tables except for Figures 5 and 10.

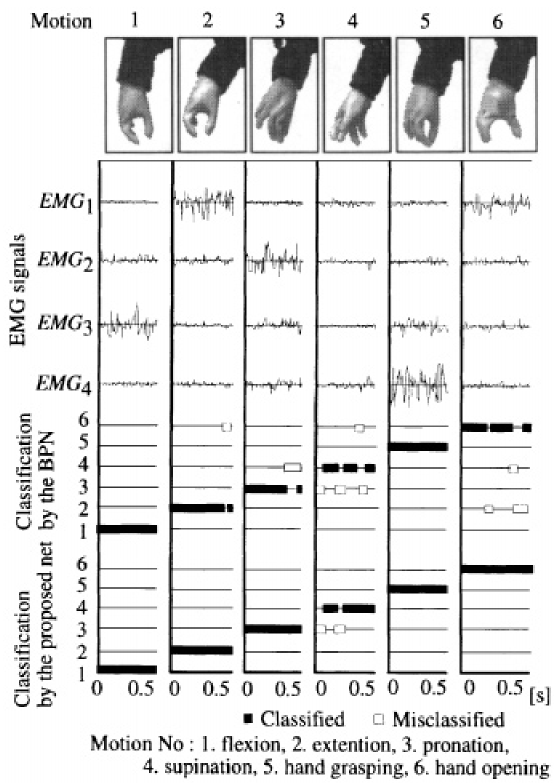


Figure 5. An example of prosthetic control using stationary EMG signals.

3.2. Experimental results for stationary EMG signals

Figure 5 shows an example of prosthetic control using stationary EMG signals (0.5 for each motion). In this figure, the motion pictures, the time-series EMG signals ($EMG_1, \ldots, EMG_4$), the classification results of the BPN, the classification results of the proposed network are shown. The stationary EMG signals used in the figure are measured from the amputee.

The convergence and determination thresholds are settled as $\theta_\varepsilon = 0.2$, $\theta_d = 0.35$, respectively, and the cut-off frequency $f_{cut} = 9$ Hz. The second layer of the BPN consists of $M_2 = 10$ units. The time history of the input data is considered back to $C = 3$ time steps in the error back-propagation learning through time.

The prosthetic hand can be controlled according to the classification results of the proposed network. A few misclassified data are observed at the beginning of the wrist supination. In this experiment, however, sufficient classification is achieved using the BPN. Next, we examine the classification ability of the BPN and determine the appropriate parameters of the BPN.

(1) *Effect of the network parameters:* First, we examine the effect of the convergence threshold θ_ε on learning/classification ability. Experiments are carried out using the BPN, and the time-series EMG signals measured from the amputee are used.

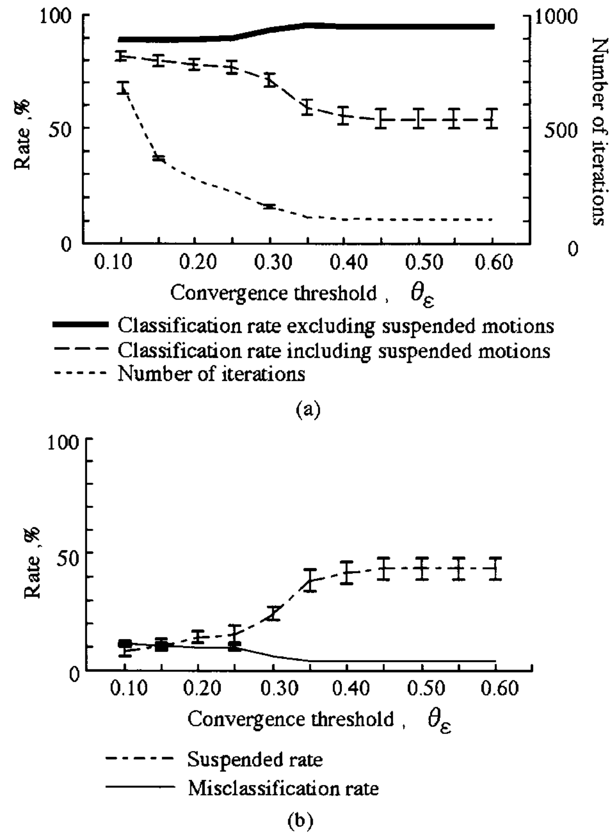


Figure 6. Effect of the convergence threshold θ_ε on the learning/classification ability of the BPN.

In the experiments, the convergence threshold θ_ε is changed from 0.1 to 0.6. The determination threshold is $\theta_d = 0.25$, and the second layer of the BPN consists of $M_2 = 10$ units. The cut-off frequency is settled as $f_{cut} = 1$ Hz.

Figure 6(a) shows the classification rate including/excluding suspended motion and the number of iterations, and Figure 6(b) shows the suspended and the misclassification rates. It can be seen that the number of learning iteration decreases as the convergence threshold θ_ε increases, and the suspended rate considerably increases near $\theta_\varepsilon = 0.3$. Under the determination threshold $\theta_d = 0.25$, the network keeps the classification rate excluding suspended motions too high for the convergence threshold θ_ε to be less than 0.3.

Next, we examine the effect of the determination threshold θ_d on classification ability. In the experiments, the determination threshold θ_d is changed from 0.05 to 0.6. The convergence threshold is set as $\theta_\varepsilon = 0.20$, and the second layer of the BPN consists of $M_2 = 10$ units. The cut-off frequency $f_{cut} = 1$ Hz is used.

Figure 7 shows the classification results. It can be seen that the suspended rate decreases as the determination threshold θ_d increases, and the classification rates including and excluding the suspended motions are almost the same for $\theta_d \geqslant 0.4$. On the other hand, the misclassification

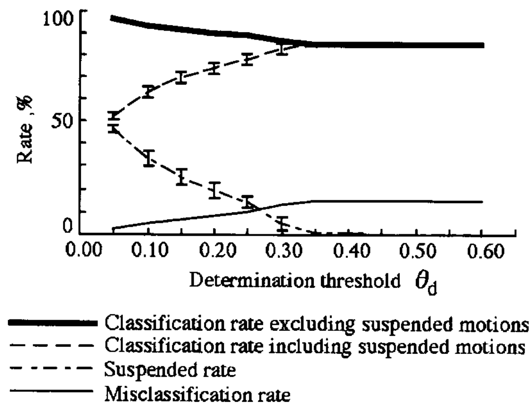


Figure 7. Effect of the determination threshold θ_d on the classification ability of the BPN.

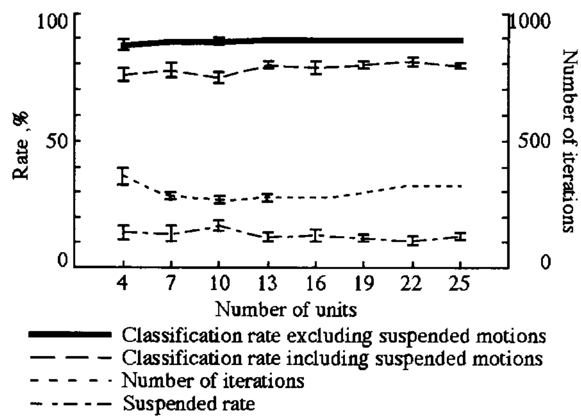


Figure 8. Effect of the number of units in the second layer M_2 on the learning/classification ability of the BPN.

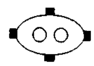
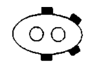
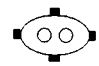
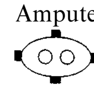
decreases by suspending classification for $\theta_d \leqslant 0.4$, and the classification rate excluding suspended motion increases. Therefore, if the operator sets the determination threshold appropriately, it is expected that the misclassification can be removed.

Also, Figure 8 shows the classification results of the BPN, which indicates the changes of the learning/classification ability with the number of the units in the second layer M_2 from 4 to 25. The convergence and determination thresholds are set as $\theta_\varepsilon = 0.2$, $\theta_d = 0.25$, respectively, and the cut-off frequency is $f_{cut} = 1$ Hz.

In the experiments, only small improvement of the learning/classification ability depending on the number of the units in the second layer is observed. Note that for a small number of units in the second layer, the learning does not converge frequently.

Based on the above classification results, the network parameters are determined as follows: the convergence and determination thresholds are $\theta_\varepsilon = 0.2$, $\theta_d = 0.25$, respectively, and the number of

Table II. Classification results of stationary EMG signals for three subjects.

Experiments	No. 1	No. 2	No. 3	No. 4
Subjects	Normal A	Normal A	Normal B	Amputee
Electrode locations				
Classification rate excluding suspended motions (%)	99.86 ± 0.34	99.98 ± 0.05	98.65 ± 0.53	88.89 ± 1.26
Classification rate including suspended motions (%)	93.27 ± 1.01	94.23 ± 1.26	69.17 ± 2.73	77.62 ± 2.70
Suspended rate (%)	6.60 ± 1.12	5.75 ± 1.28	29.88 ± 2.90	12.67 ± 3.24

the second layer of the BPN is $M_2 = 10$ units. Table II shows the experimental results for three subjects, where the location of the electrodes is shown in the table. The classification results of the BPN are shown, and the cut-off frequency in the pre-processing part is set as $f_{\mathrm{cut}} = 1$ Hz.

The network can classify six motions with an accuracy of about 90 per cent. In the experiment nos. 3 and 4, however, the suspended rate increases high, so that the classification rate including suspended motions decreases remarkably.

(2) *Effect of the terminal learning:* If a human subject performs continuous motions in a few seconds, the sampled EMG patterns must be changed considerably depending on time. It is necessary to use a sampled pattern as an input to the NN which contains frequency components of the continuous motions. Therefore, we examine the changes of the classification rate and learning iterations depending on the cut-off frequency f_{cut} of the Butterworth filter in the pre-processing part.

The experiments are carried out using the common back-propagation learning rule shown in Figure 9(a) and the terminal learning rule shown in Figure 9(b). Both experiments are performed using the BPN and the time-series EMG signals measured from the amputee.

In Figure 9(a), the network keeps classification rate high for any cut-off frequency f_{cut}. The number of learning iterations, however, increases remarkably as f_{cut} increases. In Figure 9(b), as the terminal learning is used, the convergence time is always less than the upper limit of a prespecified time $t_{\mathrm{f}nn}/\Delta t_{nn} = 1000$, while the classification rate decreases slightly.

Table III shows the experimental results for three subjects using the common back-propagation learning rule and the proposed learning rule. The cut-off frequency is $f_{\mathrm{cut}} = 9$ Hz.

Both the networks can perform high classification rates. However, in the case of the common back-propagation learning rule, the learning iterations for the amputee's data increase remarkably as shown in Table III(a). On the other hand, in the case of the terminal learning, the convergence time is always less than the upper limit of the prespecified number of iterations 1000.

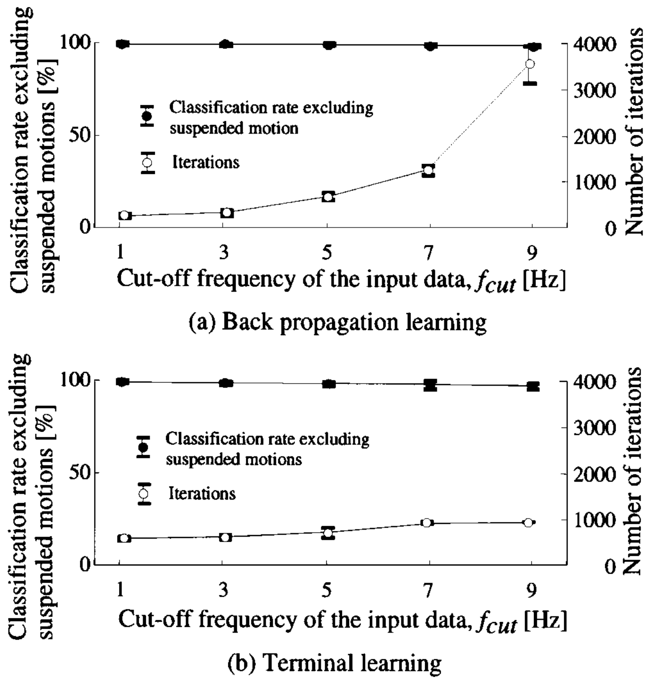


Figure 9. Effect of the terminal learning on the learning/classification ability of the BPN.

Table III. Effect of the terminal learning on the learning/classification ability of the BPN.

Experiments	No. 1	No. 2	No. 3	No. 4
Subjects	Normal A	Normal A	Normal B	Amputee
Electrode locations				
(a) *Back propagation learning*				
Classification rate including suspended motions (%)	95.1 ± 1.3	96.3 ± 0.8	89.7 ± 0.8	93.4 ± 0.5
Misclassification rate (%)	0.0 ± 0.0	0.1 ± 0.1	0.8 ± 0.4	1.3 ± 0.4
Suspended rate (%)	4.9 ± 1.3	3.6 ± 0.9	9.5 ± 0.9	5.2 ± 0.7
Number of iterations	769.2 ± 56.8	760.0 ± 45.7	621.5 ± 81.0	3488.4 ± 2201.1
(b) *Back propagation learning using* TA				
Classification rate including suspended motions (%)	92.7 ± 1.0	94.3 ± 1.3	90.7 ± 0.8	86.7 ± 1.4
Misclassification rate (%)	0.0 ± 0.0	0.0 ± 0.1	1.2 ± 0.4	0.4 ± 0.2
Suspended rate (%)	7.3 ± 0.9	5.6 ± 1.3	9.2 ± 0.7	13.3 ± 1.3
Number of iterations	982.4 ± 26.4	990.8 ± 19.6	710.7 ± 48.6	1000.0 ± 0.0

3.3. Experimental results for non-stationary EMG signals

Figure 10 shows an example of prosthetic control using non-stationary EMG signals. The non-stationary EMG signals used in the figure are measured from the amputee. The suspended data are also plotted as one of the classification results. If the sum of the squared EMG_i $(i = 1, \ldots, 4)$ is less than the threshold, it is considered that no motion appears. The determination threshold is $\theta_d = 0.25$, and the time history of the input data is considered back to $C = 3$ time steps during the learning of the NF.

In the results with the BPN, many misclassified data are observed, especially in the latter half and immediately after switching the motions. On the other hand, the proposed method considerably improves the classification performance using the NF. This is the effect of the NF that can be adapted for the time history of the classification results.

Next, we examine the changes of the classification rates with the cut-off frequency of the Butterworth filter in the pre-processing part as shown in Figure 11. It can be seen from the figure that the classification rate becomes worse with low cut-off frequency $f_{cut} = 1$ Hz, since the time-varying characteristics of the EMG signals caused by the motions are filtered out in this

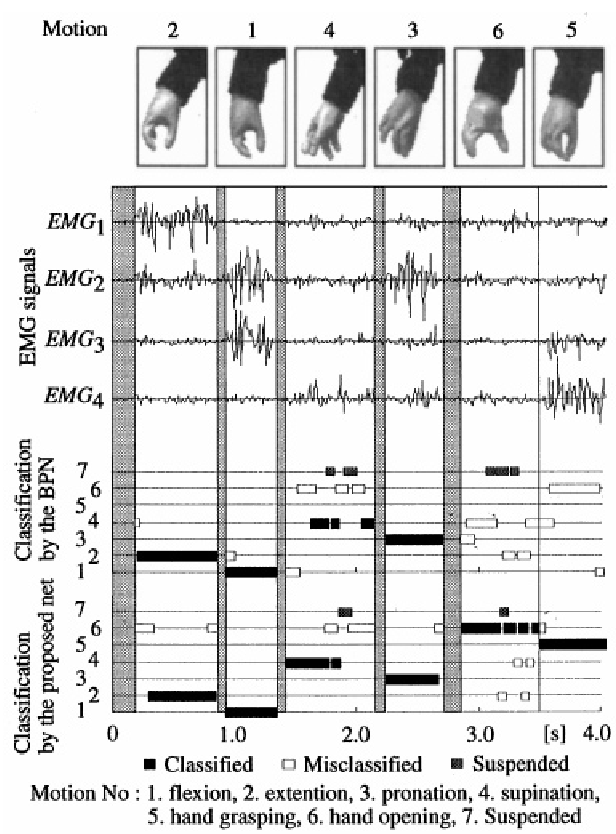


Figure 10. An example of prosthetic control using non-stationary EMG signals.

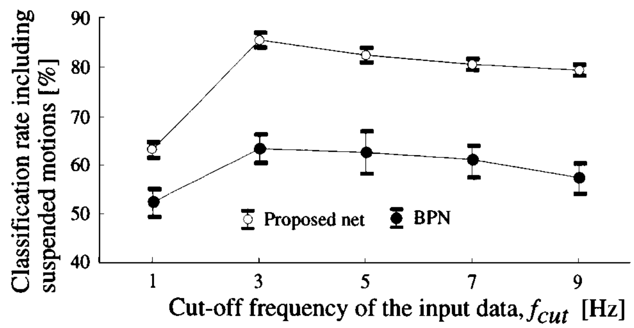


Figure 11. Effect of the cut-off frequency f_{cut} on the classification ability of the proposed network.

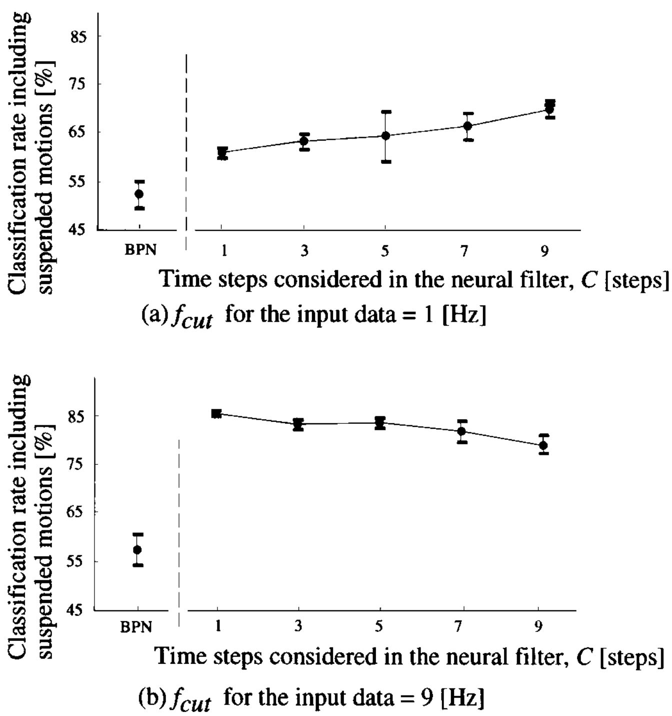


Figure 12. Effect of the time steps C of the BPTT on the classification ability of the proposed network.

case. The NF that can take into account the time history of the output of the BPN improves the classification accuracy. In Figure 11, the classification rates increased by 10–20 per cent using the NF.

Figure 12 indicates changes of the classification rate with the number of the time steps C considered in the NF. The classification rate of the BPN is also shown in the figure, and the

Table IV. Classification results of non-stationary EMG signals for two subjects.

Experiments	No. 1	No. 2	No. 3	No. 4
Subjects	Normal C	Normal C	Amputee	Amputee
Electrode locations				
(a) *BPN*				
Classification rate including suspended motions (%)	52.3 ± 1.4	49.6 ± 1.7	58.4 ± 2.0	50.2 ± 1.7
Misclassification rate (%)	2.2 ± 1.6	1.9 ± 1.6	1.1 ± 1.7	2.1 ± 1.5
Suspended rate (%)	45.5 ± 2.0	48.5 ± 1.7	40.5 ± 1.4	47.7 ± 1.3
Number of iterations (NNP)	884.5 ± 19.1	984.6 ± 32.1	1000.0 ± 0.0	792.3 ± 42.1
(b) *Proposed network*				
Classification rate including suspended motions (%)	77.8 ± 1.4	75.2 ± 1.7	81.6 ± 0.9	85.2 ± 2.0
Misclassification rate (%)	1.1 ± 1.2	2.1 ± 1.5	2.0 ± 0.8	2.0 ± 1.7
Suspended rate (%)	21.1 ± 0.9	22.7 ± 1.3	17.4 ± 0.7	12.8 ± 1.3
Number of iterations (NNP)	884.5 ± 19.1	984.6 ± 32.1	1000.0 ± 0.0	792.3 ± 42.1
Number of iterations (NFP)	1000.0 ± 0.0	1000.0 ± 0.0	1000.0 ± 0.0	1000.0 ± 0.0

cut-off frequency f_{cut} used is 1 and 9 Hz. In the case of $f_{\text{cut}} = 1$ Hz, the classification rates tend to improve with increase in the time steps. On the other hand, in the case of $f_{\text{cut}} = 9$ Hz, no improvement of the classification rates depending on the time steps is observed. This may be due to the fact that the learning became difficult for $f_{\text{cut}} = 9$ Hz, in which the filtered EMG signals varied largely.

To examine the effect of the NF on the classification result, experiments are carried out for two subjects. The electrodes for subject C are located on two different positions shown in Table IV. In the experiments nos. 3 and 4, the EMG signals are measured from the amputee on different days. Table IV shows the experimental results, where the cut-off frequency and the time steps of the BPTT are $f_{\text{cut}} = 9$ Hz and $C = 3$ time steps, respectively. The proposed network can achieve high classification performance for all experiments and the misclassification rate is quite small under the determination threshold $\theta_d = 0.25$. On the other hand, the classification rates by the BPN decreased by 20–30 per cent for all the subjects because the learning and classification become considerably difficult for non-stationary EMG signals. Note that the convergence time is always less than the upper limit of a prespecified time.

4. CONCLUSIONS

The present paper proposed a new neural network for pattern classification problems of the time-series EMG signals. To classify the non-stationary EMG signals accurately, the proposed network is combined to form two different neural networks: one is a common back-propagation neural network, and the other a recurrent neural filter. Also, the terminal learning method is newly developed in order to regulate the convergence time. To examine the classification ability of the proposed network, experiments were performed for several subjects using the stationary/non-stationary EMG signals. The results obtained here are summarized as follows:

- The network structure and its learning method have been newly developed in order to classify non-stationary EMG signals accurately while changing motions continuously.
- The operator can prespecify the learning convergence time using the terminal learning method.
- The continuous motions of the operator can be estimated from the time-series EMG signals with sufficient accuracy.
- The prosthetic hand can be controlled using the operator's EMG signals.
- The neural filter can take into account the time history of the input signal, and improve the classification accuracy. Especially, for non-stationary EMG signals, the classification rates increased by 20–30 per cent using this NF.
- Ambiguous classifications can be suspended based on the entropy of the network output. If the operator sets the determination threshold appropriately, misclassification can be eliminated.

In this paper, the NFs are connected to output units of the BPN individually in order to clarify the role of each network and to simplify the network learning, so that mutual relations among the BPN's outputs need not be considered. If we use the NF with multiple input/output units which is connected to the output units of the BPN, this relation can be considered and the classification ability may be improved. Also, the determination suspension method based on the entropy may suspend correct classification if the operator sets the small value as the determination threshold. Therefore the appropriate threshold must be determined depending on the measured EMG signals. If this threshold can be set automatically, it should be effective for some practical applications.

REFERENCES

1. Graupe D, Magnussen J, Beex AAM. A microprocessor system for multifunctional control of upper limb prostheses via myoelectric signal identification. *IEEE Transactions on Automatic Control* 1978; **23**(4):538–544.
2. Lee S, Saridis GN. The control of a prosthetic arm by EMG pattern recognition. *IEEE Transactions on Automatic Control* 1984; **29**(4):290–302.
3. Doershuk PC, Gustafson DE, Willsky AS. Upper extremity limb function discrimination using EMG signal analysis. *IEEE Transactions on Biomedical Engineering* 1983; **30**(1):18–29.
4. Triolo RJ, Moskowitz GD. The theoretical development of a multichannel time-series myoprocessor for simultaneous limb function detection and muscle force estimation. *IEEE Transactions on Biomedical Engineering* 1989; **36**(10): 1004–1017.
5. Kelly MF, Paker PA, Scott RN. The application of neural networks to myoelectric signal analysis: a preliminary study. *IEEE Transactions on Biomedical Engineering* 1990; **37**(3):211–230.
6. Hiraiwa A, Shimohara K, Tokunaga Y. EMG pattern analysis and classification by neural network. *Proceedings of IEEE International Conference on System, Man and Cybernetics*, 1989;1113.
7. Tsuji T, Ichinobe H, Ito K, Nagamachi M. Discrimination of forearm motions from EMG signals by error back propagation typed neural network using entropy. *Transactions of the Society of Instrument and Control Engineers* 1993; **29**(10):1213–1220 (in Japanese).

8. Koike Y, Kawato M. Estimation of arm posture in 3D-space from surface EMG signals using a neural network model. *IEICE Transactions on Information and Systems* 1994; **E77-D**(4):368–375.
9. Tsuji T, Mori D, Ito K. Motion discrimination method from EMG signals using statistically structured neural networks. *Transactions on IEE Japan* 1992; **112-C**(8):465–473 (in Japanese).
10. Rumelhart DE, McClelland JL, Williams RJ. Learning internal representations by error propagation. In *Parallel Distributed Processing*, *Vol. I*, Rumelhart DE, McClelland JL, the PDP Research Group (eds). MIT Press: Cambridge, 1986; 318–362.
11. Tsuji T, Ichinobe H, Fukuda O, Kaneko M. A maximum likelihood neural network based on a log-linearized gaussian mixture model. *Proceedings of IEEE International Conference on Neural Networks*, 1995; 2479–2484.
12. Fukuda O, Tsuji T, Kaneko M. Pattern classification of EEG signals using a log-linearized gaussian mixture neural networks. *IEEE International Conference on Neural Networks*, 1995; 2479–2484.
13. Tsuji T, Fukuda O, Ichinobe H, Kaneko M. A log-linearized gaussian mixture network and its application to EEG pattern classification. *IEEE Transactions on System*, *Man and Cybernetics-Part C*: *Applications and Reviews* 1999; **29**(1):1–13.
14. Lo JTH. Synthetic approach to optimal filtering. *IEEE Transactions on Neural Networks*. 1994; **5**(5):803–811.
15. Zak M. Terminal attractors for addressable memory in neural networks. *Physics Letters A* 1988; **133**:18–22.
16. Fukuda O, Tsuji T, Otsuka A, Kaneko M. EMG-based human-robot interface for rehabilitation aid. *Proceedings of IEEE International Conference on Robotics and Automation*, 1998; 3492–3497.

