

# MYOELECTRIC PATTERN DISCRIMINATION USING AN HMM-BASED NEURAL NETWORK

Nan Bu\*, Osamu Fukuda\*\*, Satoshi Noda\*, and Toshio Tsuji\*

\*Department of the Artificial Complex Systems Engineering  
Hiroshima University, Higashi-Hiroshima, Hiroshima, 739-8527, JAPAN

\*\*Institute for Human Science and Biomedical Engineering

National Institute of Advanced Industrial Science and Technology, Namiki, Tsukuba, 305-8564, JAPAN.

e-mail: bu@bsys.hiroshima-u.ac.jp

## ABSTRACT

A pattern discrimination method of electromyogram (EMG) signal for prosthetic control is presented in this paper. This method uses a novel recurrent neural network based on the hidden Markov model. This network includes recurrent connections to enable the capability of modeling time series such as the EMG signal. Weight coefficients in the network can be learned using a well-known back-propagation through time algorithm. The pattern discrimination experiments were conducted to demonstrate feasibility and performance of the proposed method. This method successfully discriminated forearm motions using the EMG signal, and achieved considerably high discrimination performance in comparison with the other discrimination methods.

## KEY WORDS

Neural Networks, Pattern Discrimination, EMG, Recurrent Neural Network.

## 1 Introduction

A human-machine interface for prosthetic control is important to assist the disabled who have lost their manipulation capability of the upper limb. EMG signal is often used as an interface tool for prosthetic devices since the EMG signal is the manifestation of the electrical stimulations, which motor units of the arm receive from the Central Nervous System (CNS), and it indicates the activation level of motor units associated with contraction of muscle. Different motions resulting from different modes of muscle activation generate different EMG patterns. Many researchers have tried to use the EMG signal to provide control command for the prosthetic arm [1]-[11].

Various techniques have been proposed to discriminate the EMG pattern. Graupe *et al.* [1] used autoregressive (AR) models to represent the EMG signal, which was measured from a single electrode-site between Biceps and Triceps. Motion patterns were discriminated according to the parameters of the AR model. Tsuji *et al.* [2] used a multi-dimensional AR model, and discriminated the

forearm motions using the frequency and the amplitude characteristics extracted from the multi-channel EMG signals. However, these methods could not achieve high discrimination performance, because they applied a linear model to approximate the nonlinear characteristics of the EMG signal, which change largely depending on muscle fatigue, sweat, changes of electrode location, and so on.

Neural Network (NN) is suitable for modeling nonlinear data, and can cover the distinction among different conditions. Several EMG pattern discrimination methods based on NN [3]-[11] have been presented in the last decade. For example, Hiraiwa *et al.* [3], [4] used a back-propagation (BP) NN to perform pattern discrimination with frequency features. Kelly *et al.* [5] succeeded in discriminating four forearm motions (Flexion, Extension, Pronation, Supination) using the combination of BPNN and a Hopfield NN. Similar works have also been done by Koike and Kawato [6], Huang and Chen [7], etc. However, BPNN frequently used in the previous researches cannot realize high learning and discriminating performance, and a large amount of learning data as well as a great number of learning iterations is required.

On the other hand, Tsuji *et al.* put forward NN including a statistical model [8], and used the entropy which was calculated from the outputs of NN to reduce incorrect discrimination [9]. They also proposed a feedforward probabilistic NN, a Log-Linearized Gaussian Mixture Network (LLGMN) [10], which is based on the Gaussian mixture model (GMM) and the log-linear model of the probability density function. LLGMN was successfully applied to the EMG pattern classification, where eight motions of forearm were classified using EMG signals measured from several pairs of electrodes [11]. However, these methods just focused on the static features of EMG signal, so that the time-varying characteristics were not taken into account. That is why high classification performance cannot be achieved.

This paper proposes a new EMG pattern discrimination method using a recurrent NN, a Recurrent Log-Linearized Gaussian Mixture Network (R-LLGMN) [12]. R-LLGMN

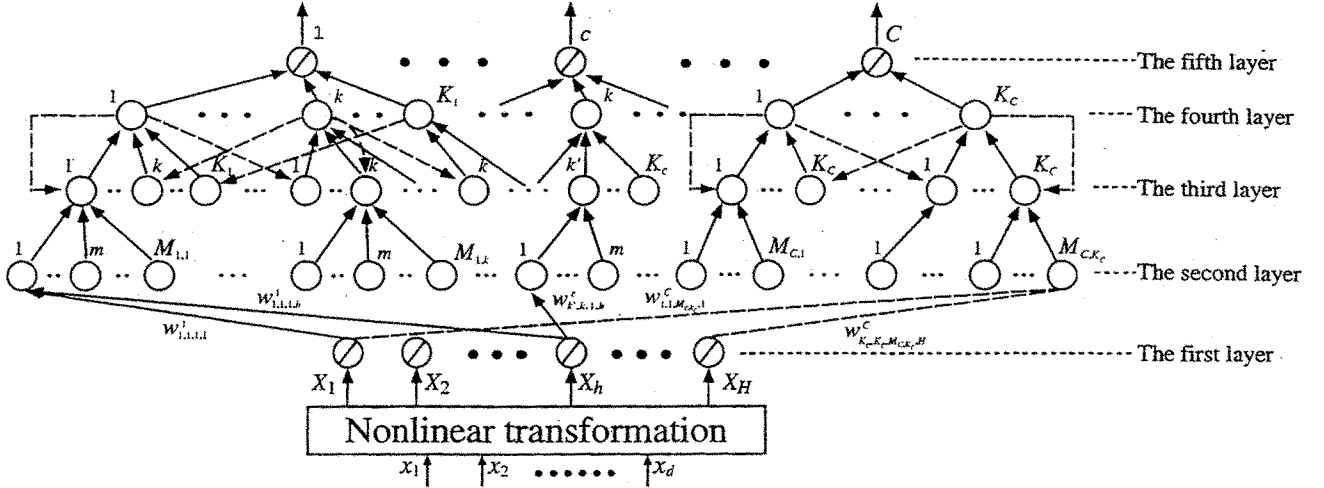


Figure 1. Structure of R-LLGMN.

is based on the algorithm of Hidden Markov Model (HMM) [13], and incorporates recurrent connections to make use of the time context in the EMG signal. With the weight coefficients well learned using a learning scheme of the back-propagation through time (BPTT) algorithm [14], R-LLGMN can calculate the *a posteriori* probability of discriminating class for each EMG pattern. Pattern discrimination experiments of the EMG signal were conducted using R-LLGMN with five subjects, and the proposed method was compared with BPNN and LLGMN in the experiments.

This paper is organized as follows: Section II explains the architecture and the learning algorithm of R-LLGMN. The EMG pattern discrimination method is described in Section III. The discrimination experiments are presented in Section IV. The final section gives the conclusion of this paper.

## 2 A Recurrent Log-Linearized Gaussian Mixture Network

A Recurrent Log-Linearized Gaussian Mixture Network (R-LLGMN) [12], which is based on GMM and the HMM algorithm, is adopted in the proposed EMG pattern discrimination method. The HMM is a well-studied technology for discrimination of time series and has been successfully applied in the field of speech recognition. The network structure and the learning algorithm of R-LLGMN are explained in the following subsections.

### 2.1 Network Structure

R-LLGMN is a five-layer recurrent NN with feedback connections between the forth and the third layer, the structure of which is shown in Fig. 1. In this structure, a continuous density HMM [13] is included. There are  $C$  classes in this model and the class  $c$  ( $c \in \{1, \dots, C\}$ ) is

composed of  $K_c$  states. The observation probability of state  $k$  ( $k \in \{1, \dots, K_c\}$ ) in class  $c$  is approximated with GMM. Suppose that, for an input vector series  $\mathbf{x}(t) \in \mathbb{R}^d$  ( $t = 1, \dots, T$ ),  $\mathbf{x}(t)$  must occur from one state  $k$  of class  $c$  in the model at any time. It is expected that the *a posteriori* probability for each class can be estimated with a well trained R-LLGMN.

First, the input vector series  $\mathbf{x}(t) \in \mathbb{R}^d$  ( $t = 1, \dots, T$ ) is pre-processed into the modified input series  $\mathbf{X}(t) \in \mathbb{R}^H$  as follows:

$$\mathbf{X}(t) = (1, \mathbf{x}(t)^T, x(t)_1^2, x(t)_1 x(t)_2, \dots, x(t)_1 x(t)_d, x(t)_2^2, x(t)_2 x(t)_3, \dots, x(t)_2 x(t)_d, \dots, x(t)_d^2)^T, \quad (1)$$

where the dimension  $H$  is determined as  $H = 1 + d(d + 3)/2$ . The vector  $\mathbf{X}(t)$  acts as the input of the first layer, and the identity function is used for activation of each unit. The output of the  $h$ th ( $h = 1, \dots, H$ ) unit in the first layer is defined as  ${}^{(1)}O_h(t)$ .

Unit  $\{c, k, k', m\}$  ( $c = 1, \dots, C; k', k = 1, \dots, K_c; m = 1, \dots, M_{c,k}$ ) in the second layer receives the output of the first layer weighted by the coefficient  $w_{k',k,m,h}^c$ . The input  ${}^{(2)}I_{k',k,m}^c(t)$  and the output  ${}^{(2)}O_{k',k,m}^c(t)$  are defined as

$${}^{(2)}I_{k',k,m}^c(t) = \sum_{h=1}^H {}^{(1)}O_h(t) w_{k',k,m,h}^c, \quad (2)$$

$${}^{(2)}O_{k',k,m}^c(t) = \exp\left({}^{(2)}I_{k',k,m}^c(t)\right), \quad (3)$$

where  $C$  is the number of discriminating classes,  $K_c$  is the number of states in class  $c$ , and  $M_{c,k}$  denotes the number of GMM components in the state  $k$  of class  $c$ .

The outputs of units  $\{c, k, k', m\}$  in the second layer are added up and input into a unit  $\{c, k, k'\}$  in the third

layer. Also, the output of the fourth layer is fed back to the third layer. These are expressed as follows:

$${}^{(3)}I_{k',k}^c(t) = \sum_{m=1}^{M_{c,k}} {}^{(2)}O_{k',k,m}^c(t), \quad (4)$$

$${}^{(3)}O_{k',k}^c(t) = {}^{(4)}O_{k'}^c(t-1) {}^{(3)}I_{k',k}^c(t), \quad (5)$$

where  ${}^{(4)}O_{k'}^c(0) = 1.0$  is for the initial phase.

The activation function in the fourth layer is described in the form as

$${}^{(4)}J_k^c(t) = \sum_{k'=1}^{K_c} {}^{(3)}O_{k',k}^c(t), \quad (6)$$

$${}^{(4)}O_k^c(t) = \frac{{}^{(4)}J_k^c(t)}{\sum_{c'=1}^C \sum_{k'=1}^{K_{c'}} {}^{(4)}I_{k',k}^{c'}(t)}. \quad (7)$$

In the fifth layer, the unit  $c$  integrates the outputs of  $K_c$  units  $\{c, k\}$  ( $k = 1, \dots, K_c$ ) in the fourth layer. The relationship in the fifth layer is defined as

$${}^{(5)}I^c(t) = \sum_{k=1}^{K_c} {}^{(4)}O_k^c(t), \quad (8)$$

$${}^{(5)}O^c(t) = {}^{(5)}I^c(t). \quad (9)$$

After only optimizing the weight coefficients  $w_{k',k,m,h}^c$  between the first layer and the second layer, this NN can estimate the *a posteriori* probability of each class. In the next subsection, the learning algorithm for this NN is described briefly.

## 2.2 Learning Algorithm

A set of input vector streams  $\tilde{\mathbf{x}}^{(n)} = (\mathbf{x}(1)^{(n)}, \mathbf{x}(2)^{(n)}, \dots, \mathbf{x}(T_n)^{(n)})$  ( $n = 1, \dots, N$ ) and the teacher vector  $\mathbf{T}^{(n)} = (T_1^{(n)}, \dots, T_C^{(n)}, \dots, T_C^{(n)})^T$  are given for the learning of R-LLGMN. It is supposed that the network acquires the characteristics of the data through learning, if for all the streams the last output of stream  $\tilde{\mathbf{x}}^{(n)}$ , namely  ${}^{(5)}O^c(T_n)$  ( $c = 1, \dots, C$ ), is close enough to the teacher signal  $\mathbf{T}^{(n)}$ . The energy function for the network is defined as

$$J = \sum_{n=1}^N J_n = - \sum_{n=1}^N \sum_{c=1}^C T_c^{(n)} \log {}^{(5)}O^c(T_n). \quad (10)$$

The learning process is to minimize  $J$ , that is, to maximize the likelihood that each teacher vector  $\mathbf{T}^{(n)}$  is obtained for the input stream  $\tilde{\mathbf{x}}^{(n)}$ . In this paper, the backpropagation-through-time (BPTT) algorithm [14] is applied in the learning rule because of the recurrent connection in R-LLGMN.

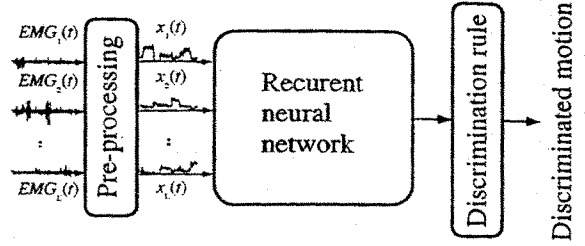


Figure 2. Structure of the proposed method.

It is supposed that the error gradient within a stream (block) is accumulated and weight modifications are only computed at the end of each block; the error is then propagated backward to the beginning of the block.

## 3 EMG Discrimination Method

A structure of the proposed discrimination method is shown in Fig. 2. This method consists of three parts in sequence: (1) EMG signal processing, (2) Recurrent neural network, and (3) Discrimination rule. First, the EMG signals are processed to extract the feature patterns. The EMG signals measured from  $L$  pairs of electrodes are rectified and filtered by a second-order Butterworth filter (cut-off frequency: 1 [Hz]). Then, they are digitized by an A/D converter with sampling frequency  $f_d$  [Hz]. Each sampled data is defined as  $EMG_l(t)$  ( $l = 1, \dots, L$ ) and is normalized to make the sum of  $L$  channels equal to 1:

$$x_l(t) = \frac{EMG_l(t) - EMG_l^{st}}{\sum_{l'=1}^L (EMG_{l'}(t) - EMG_{l'}^{st})} \quad (l = 1, \dots, L), \quad (11)$$

where  $EMG_l^{st}$  is the mean value of  $EMG_l(t)$  which is measured while relaxing the arm. The feature vector  $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_L(t)]$  is used for the input of R-LLGMN. In this paper, we assumed that the amplitude level of the EMG signal is changed in proportion to muscle force. Force information  $F_{EMG}(t)$  for input vector  $\mathbf{x}(t)$  is defined as follows:

$$F_{EMG}(t) = \frac{1}{L} \sum_{l=1}^L \frac{EMG_l(t) - EMG_l^{st}}{EMG_l^{max} - EMG_l^{st}}, \quad (12)$$

where  $EMG_l^{max}$  is the mean value of  $EMG_l(t)$  measured while keeping the maximum voluntary contraction.

For the pattern discrimination, R-LLGMN described in 2.1 is employed. Using samples labeled with the corresponding motions, R-LLGMN learns the non-linear mapping between the EMG patterns and the forearm motions. Given an EMG feature stream  $\mathbf{x}(t)$  ( $t = 1, \dots, T$ ), the output  ${}^{(5)}O^c(T)$  ( $c = 1, \dots, C$ ) presents the *a posteriori* probability of each discriminating motion.

In order to recognize whether the motion has really happened or not, the force information  $F_{EMG}(t)$  is compared with the prefixed motion appearance threshold  $M_d$ . The motion is considered to happen if  $F_{EMG}(t)$  is over than  $M_d$ . The entropy of the R-LLGMN's outputs is also calculated to present the risk of ill-discrimination. The entropy is defined as

$$H(t) = - \sum_{c=1}^C {}^5O^c(t) \log_2 {}^5O^c(t). \quad (13)$$

If the entropy  $H(t)$  is over the discrimination threshold  $H_d$ , specific motion whose probability is the largest one is determined according to the Bayes decision rule. If not, the determination is suspended.

## 4 Experiments

### 4.1 Experimental Conditions

EMG pattern discrimination experiments were conducted to examine performance of the proposed method. Experiments were held with five subjects (A, B: Amputee; C, D, E: Normal).

Subject A (male) lost his forearm about 3cm from the left wrist. EMG signals were measured from six pairs of electrodes ( $L = 6$ ) with a sampling frequency  $f_d = 60$  [Hz], and the electrodes were attached to his forearm and upper arm (Flexor Carpi Radialis (FCR), Extensor Carpi Ulnaris (ECU), Flexor Carpi Ulnaris (FCU), Biceps Brachii (BB), Triceps Brachii (TB); two pairs on FCR and one pair on the others). The subject was asked to perform six motions ( $C = 6$ ) in the order of: flexion, extension, supination, pronation, hand grasping and hand opening continuously for six seconds.

Subject B (male) lost his right hand about 15cm from the wrist. EMG signals were measured from eight pairs of electrodes attached to his forearm and upper arm ( $L = 8$ ), and  $f_d$  was set at 100 [Hz]. EMG signals during eight motions ( $C = 8$ : flexion, extension, supination, pronation, hand grasping, hand opening, co-contraction of wrist joint and co-contraction of finger part) for 20 seconds were recorded continuously.

Subject C (male), D (male) and E (male) are all normal-limbed, and six pairs of electrodes ( $L = 6$ ) were attached in the same way as subject A. EMG signals were measured (sampling frequency: 100 [Hz]) for 22 seconds, and seven motions ( $C = 7$ ) were performed with the order of hand grasping, hand opening, extension, flexion, pronation, supination and co-contraction of finger part.

In the learning process of R-LLGMN, 20 EMG patterns extracted from the EMG signals for each motion and teacher signals consisting of  $C \times 20$  patterns were

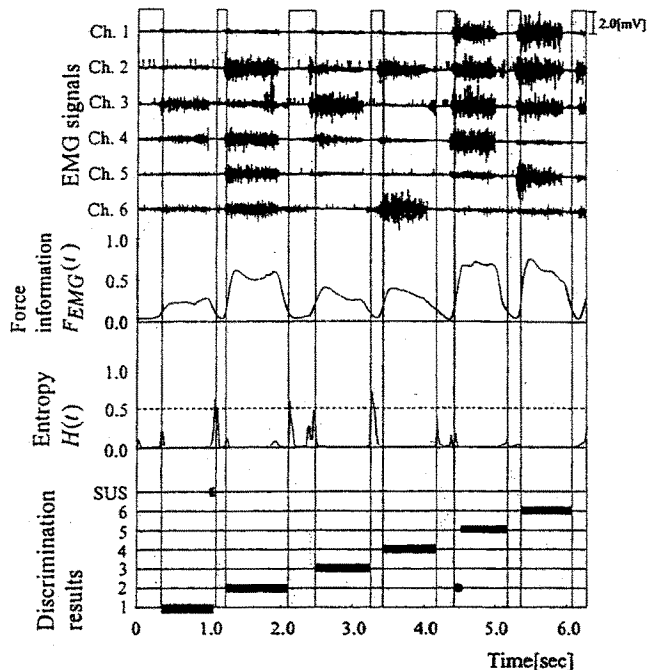


Figure 3. An example of the discrimination result.

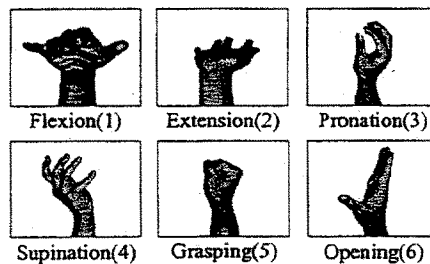


Figure 4. Motions.

used. The determination threshold  $H_d$  and the motion appearance threshold  $M_d$  were set at 0.5 and 0.2, respectively.

## 4.2 Results

### 4.2.1 Discrimination Accuracy

An example of the discrimination result of subject A is shown in Fig. 3. In this figure, six channels of the input EMG signals, the force information  $F_{EMG}(t)$ , the entropy  $H(t)$  and the discrimination results are plotted. The labels of the vertical axis in the discrimination results correspond to motions shown in Fig. 4, and SUS means that the determination was suspended. The gray areas indicate that no motion was determined because the force information is less than  $M_d$ . Incorrect determination was removed using the entropy. It can be seen from Fig. 3 that the proposed method achieves high discrimination performance even non-stationary EMG signals during the contin-

Table 1. DISCRIMINATION RESULTS OF FIVE SUBJECTS

Subject	A	B	C	D	E
R-LLGMN	99.06	89.32	93.04	93.49	92.75
Mean $\pm$ SD (%)	$\pm 0.00$	$\pm 0.37$	$\pm 0.11$	$\pm 0.00$	$\pm 0.00$
LLGMN	94.00	82.83	88.50	88.67	89.26
Mean $\pm$ SD (%)	$\pm 5.50$	$\pm 0.00$	$\pm 0.04$	$\pm 0.15$	$\pm 0.14$
BPNN	73.41	46.52	44.20	69.79	69.17
Mean $\pm$ SD (%)	$\pm 7.86$	$\pm 12.3$	$\pm 10.4$	$\pm 9.97$	$\pm 7.00$

uous motion.

Accuracy of the discrimination results for five subjects (A, B, C, D and E) was investigated as well. To verify the discrimination performance of the proposed method, comparison experiments with LLGMN and BPNN were conducted. The determination thresholds of LLGMN were set as the same values of R-LLGMN. On the other hand, BPNN had four layers (two hidden layers), and the units of each layer were set as 8, 10, 10 and 8, respectively. Each output of BPNN corresponds to each motion, so that it was normalized to make the sum of all outputs equal 1.0 and regarded as the *a posteriori* probability of each motion. The same determination thresholds were also used for the discrimination of BPNN. The learning procedure of BPNN continued until the sum of the square error became less than 0.01, where the learning rate was 0.01. However, if the sum of the square error after 50,000 iterations was still not less than 0.01, the learning procedure was stopped. In all the three methods, 10 different sets of initial weights (all randomized between [0, 1]) were used.

The mean values and the standard deviations of the discrimination rates are shown in Tables I. It can be seen that R-LLGMN achieved the best discrimination rate with a small standard deviation among all the three methods.

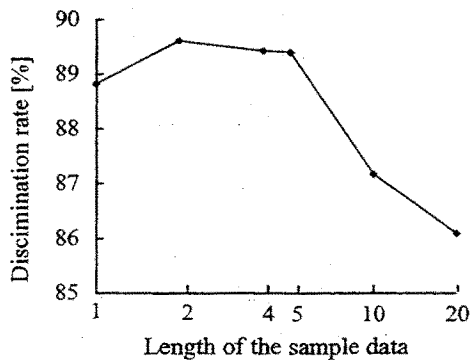


Figure 5. Discrimination rates for various data lengths (Subject B).

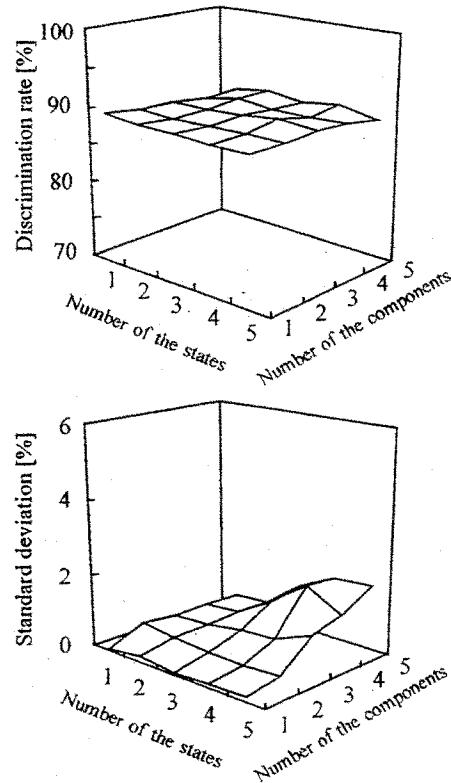


Figure 6. Discrimination rates for different model sizes (Subject B).

#### 4.2.2 Changes of Discrimination Rate with Various Conditions

The discrimination results were examined by changing the experimental conditions such as the length of sample data and parameters in the continuous density HMM. First, experiments were performed using various lengths of sample data. For each sample data, R-LLGMN was trained with 10 different sets of initial weights, which were randomly chosen in [0, 1]. The mean values of the discrimination rates for each length are shown in Fig. 5, where, the standard deviations are all very small, almost close to 0. It can be seen from Fig. 5 that the discrimination rate keeps in a high level when the sample data has an appropriate length ( $T$ ). However, if  $T > 5$ , it is too long to train R-LLGMN, the discrimination rate tends to decrease because R-LLGMN which is learned by long length sample data failed to discriminate in switching motions.

On the other hand, examinations were carried out by varying the number of components ( $M_{c,k}$ ) and states ( $K_c$ ). Ten sets of randomly chosen initial weights were used to train each sample data. The mean values and the standard deviations of the discrimination rates for different model sizes ( $M_{c,k} \times K_c : M_{c,k} \in [1, 2, 3, 4, 5]; K_c \in [1, 2, 3, 4, 5]$ )

are plotted in Fig. 6. In this figure, the discrimination rate does not increase according to the augmentation of the model size. The reason is considered that even when  $M_{c,k} = K_c = 1$ , the network has enough weights to model the desired non-linear mapping. However, the standard deviation increases as the model size increases, because there are more local minima to learn a larger model.

## 5 Conclusions

In this paper, a new EMG discrimination method based on a recurrent Log-linearized Gaussian mixture network (R-LLGMN) has been proposed for prosthetic control. Because of the recurrent connections between the third and the fourth layer in R-LLGMN, the temporal information in the EMG signal can be used for the pattern discrimination.

To examine the discrimination capability and the accuracy of the proposed method, EMG pattern discrimination experiments have been carried out with five subjects. In the experiments, the proposed method achieved a high discrimination performance for varying EMG signals, and its discrimination results are the best comparing with those of LLGMN and BPNN.

In our future research, we would like to develop a new pre-processing of the EMG signal. Discrimination performance should be improved using a combination of new pre-processing and R-LLGMN.

## Acknowledgments

This work was partly supported by Industrial Technology Research Grant Program in 2001 from New Energy and Industrial Technology Development Organization (NEDO) of Japan.

## References

- [1] D. Graupe, J. Magnussen and A. A. M. Beex, A Microprocessor System for Multifunctional Control of Upper Limb Prostheses via Myo electric Signal Identification, *IEEE Trans. on Automatic Control*, 23(4), 1978, 538-544.
- [2] T. Tsuji, K. Ito and M. Nagamachi, A limb-function discrimination method using EMG signals for the control of multifunctional powered prostheses, *Transactions Institute of Electronics, Information and Communication Engineers*, J70-D(1), 1987, 207-215. (in Japanese)
- [3] A. Hiraiwa, K. Shimohara and Y. Tokunaga, EMG pattern analysis and classification by neural network, *Proc. of IEEE International Conf. on Syst., Man and Cybern.*, 1989, 1113-1115.
- [4] A. Hiraiwa, U. Uchida and K. Shimohara, EMG/EEG pattern recognition by neural networks, *Proc. of the Eleventh European Meeting on Cybernetics and Systems Research*, 1992, 1383-1390.
- [5] M. F. Kelly, P. A. Parker and R. N. Scott, The Application of Neural Networks to Myoelectric Signal Analysis: A Preliminary Study, *IEEE Trans. on Biomedical Engineering*, 37(3), 1990, 221-230.
- [6] Y. Koike and M. Kawato, Estimation of Arm Posture in 3D-Space from Surface EMG Signals Using a Neural Network Model, *Transactions Institute of Electronics, Information and Communication Engineers*, J77-D(4), 1994, 368-375.
- [7] H.P. Huang and C.Y. Chen, Development of a Myoelectric Discrimination System for a Multi-Degree Prosthetic Hand, *Proc. of the 1999 IEEE International Conf. on Robotics and Automation*, Detroit, USA, 1999, 2392-2397.
- [8] T. Tsuji, D. Mori and K. Ito, Motion Discrimination Method from EMG Signals Using Statistically Structured Neural Networks, *The Transactions of The Institute of Electrical Engineers of Japan*, 112-C(8), 1992, 465-473. (in Japanese)
- [9] T. Tsuji, H. Ichinobe, K. Ito and M. Nagamachi, 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*, 29(10), 1993, 1213-1220. (in Japanese)
- [10] T. Tsuji, O. Fukuda, H. Ichinobe and M. Kaneko, A log-linearized Gaussian mixture network and its application to EEG pattern classification, *IEEE Trans. Systems, Man, and Cybernetics-Part C: Applications and Reviews*, 29(1), 1999, 60-72.
- [11] O. Fukuda, T. Tsuji and M. Kaneko, A human supporting manipulator based on manual control using EMG signals, *Journal of the Robotics Society of Japan*, 18(3), 2000, 387-394. (in Japanese)
- [12] T. Tsuji, N. Bu, M. Murakami and M. Kaneko, Pattern Discrimination of Raw EMG Signals Using a New Recurrent Neural Network, *Proc. of IMEKO/SICE/IEEE the Int. Symp. on Measurement, Analysis and Modeling of Human Functions*, Sapporo, Japan, 2001, 96-101.
- [13] L.R. Rabiner, A tutorial on hidden Markov model and selected applications in speech recognition, *Proc. of the IEEE*, 77(2), 1989, 257-286.
- [14] P.J. Werbos, Backpropagation through time: what it does and how to do it, *Proc. of the IEEE*, 78(10), 1990, 1550-1560.