

Chapter XVII

A Recurrent Probabilistic Neural Network for EMG Pattern Recognition

Toshio Tsuji

Hiroshima University, Japan

Nan Bu

Hiroshima University, Japan

Osamu Fukuda

National Institute of Advanced Industrial Science and Technology, Japan

ABSTRACT

In the field of pattern recognition, probabilistic neural networks (PNNs) have been proven as an important classifier. For pattern recognition of EMG signals, the characteristics usually used are: (1) amplitude, (2) frequency, and (3) space. However, significant temporal characteristic exists in the transient and non-stationary EMG signals, which cannot be considered by traditional PNNs. In this article, a recurrent PNN, called recurrent log-linearized Gaussian mixture network (R-LLGMN), is introduced for EMG pattern recognition, with the emphasis on utilizing temporal characteristics. The structure of R-LLGMN is based on the algorithm of a hidden Markov model (HMM), which is a routinely used technique for modeling stochastic time series. Since R-LLGMN inherits advantages from both HMM and neural computation, it is expected to have higher representation ability and show better performance when dealing with time series like EMG signals. Experimental results show that R-LLGMN can achieve high discriminant accuracy in EMG pattern recognition.

INTRODUCTION

Electromyographic (EMG) signals provide information about neuromuscular activities and have been recognized as efficient and promising resources for human-machine interface (HMI) used for the rehabilitation of people with mobility limitations and those with severe neuromuscular impairment. Typically, a pattern recognition process is applied to *translate* EMG signals into control commands for the HMIs, such as powered prostheses and functional electrical stimulation devices (Englehart et al., 2001; Fukuda et al., 2003; Hudgins et al., 1993; Lusted & Knapp, 1996). Generally speaking, a successful EMG pattern recognition technique relies on two principle elements: a pattern classifier with reliable discrimination accuracy and efficient representation of EMG feature characteristics.

Probabilistic neural networks (PNNs) developed in the field of pattern recognition make a decision according to the probability density distribution of patterns in the feature space (Specht, 1990; Tsuji et al., 1999). Since PNNs integrate statistical models into the neural networks' architecture as prior knowledge, outstanding performance has been reported. Recently, PNNs have become widely accepted as important classifiers and have been proven to be efficient, especially for complicated problems such as pattern recognition of bioelectric signals.

For EMG pattern recognition using PNNs, the feature characteristics usually used include: (1) amplitude, (2) frequency, and (3) spatial information from multiple channels of EMG signals. However, significant temporal characteristics exist in the transient and non-stationary EMG signals, which cannot be considered by the traditional PNNs based on static stochastic models, and, in some cases, temporal characteristics could be the only clues for reliable recognition.

This chapter introduces a recurrent PNN called recurrent log-linearized Gaussian mixture network (R-LLGMN) (Tsuji et al., 2003) into

EMG pattern recognition, with emphasis on utilizing temporal characteristics. The structure of R-LLGMN is based on the hidden Markov model (HMM) algorithm, which is a routinely used technique for modeling stochastic time series. Since R-LLGMN inherits the advantages from both HMM and neural computation, it is expected to have higher representation ability and exhibit better classification performance when dealing with time series like EMG signals.

After a review of the literature, the structure and algorithm of R-LLGMN are explained. The proposed EMG pattern recognition method using R-LLGMN is then described, and experiments on filtered EMG and raw EMG signals are presented. Based on the experimental results, the possibility of applying the proposed method to practical human interface control is discussed. The final section offers some concluding remarks.

BACKGROUND

Up to now, many techniques have been developed for EMG pattern recognition using statistical methods and neural networks (NNs). Kang et al. (1995) proposed a maximum likelihood method (MLM) based on Mahalanobis distances between input pattern and the prototypes, and the Bayes decision rule is applied in this method. A traditional linear discriminant analysis (LDA) classifier is used in an EMG classification scheme for multifunction myoelectric control (Englehart et al., 2001).

Due to NNs' learning capability of finding near-optimum functional relationships between the class memberships and the EMG patterns, several NN-based EMG pattern recognition methods have been presented. For example, Hiraiwa et al. (1989) used a multilayer perceptron (MLP) NN to perform pattern discrimination of five finger motions. Kelly et al. (1990) applied an MLP to classify four arm functions. Hudgins et al. (1993) devised a control system for powered

upper-limb prostheses using a set of time-domain features extracted from EMG signals and a simple MLP as a classifier. Also, similar studies have been developed using MLPs to classify EMG features, such as autoregressive (AR) parameters (Lamounier et al., 2002) and features of filtered EMG signals (Tsuji et al., 1993). However, several factors have hindered the extension of MLP classifiers for other applications, such as the choice of network structure, slow learning convergence, the need for a large amount of training data, and local minima.

To tackle these problems, numerous attempts have been made by the pattern recognition community to integrate statistical models, as prior knowledge, into the classifier's architecture, to take advantage of both statistical classification methods and neural computation. Consequently, probabilistic neural networks (PNNs) have been developed for pattern recognition (Specht, 1990; Zhang, 2000). In particular, Tsuji et al. (1999) proposed a feedforward PNN, a log-linearized Gaussian mixture network (LLGMN), which is based on the Gaussian mixture model (GMM) and a log-linear model. Although weights of the LLGMN correspond to a non-linear combination of the GMM parameters, such as mixture coefficients, mean vectors, and covariance matrices, constraints on the parameters in the statistical model are relieved in the LLGMN. Therefore, a simple backpropagation-like learning algorithm can be derived, and the parameters of LLGMN are trained according to a criterion of maximum likelihood (ML). The LLGMN has been successfully applied to EMG pattern recognition, where eight motions of the forearm have been classified using EMG signals measured by several pairs of electrodes (Fukuda et al., 2003). Also, the LLGMN has been further used to develop interface applications like prosthetic devices and EMG-based pointing devices (Fukuda et al., 1997, 1999; Fukuda et al., 2003).

However, since the GMM is a *static* stochastic model, it cannot make efficient use of temporal

(time-varying) characteristics in EMG signals. Generally, pattern recognition using LLGMN is made under the assumption that feature patterns are stationary or change very slowly. EMG signals, in fact, are non-stationary and vary significantly in amplitude and frequency, even in the space domain. Due to the complicated nature of EMG signals, it is widely accepted that the temporal characteristic contains information important for pattern recognition (Englehart et al., 1999).

In order to cope with the time-varying characteristics of EMG signals, a pattern recognition method using an MLP classifier and a neural filter (NF) was applied (Tsuji et al., 2000). Continuous motions by the operators can be discriminated with sufficient accuracy even using the non-stationary time series of EMG signals. In addition to improving the classifiers, time-frequency representations of EMG signals have been adopted to gain a high level of discrimination accuracy (Englehart et al., 1999, 2001; Hussein & Granat, 2002). Although these methods can generate sufficient discrimination accuracy, there may be some criticism due to more complicated signal processing required or more intricate structure of classifiers. Also more parameters in the algorithm(s) of the signal processing and/or the classifier need to be determined by the user. Optimization of the whole pattern recognition method is almost impossible, and it is hard to gain a high performance of discrimination, especially in practical applications.

The present study focuses on the classifier aspect of EMG pattern recognition and introduces a recurrent PNN to improve discrimination accuracy when dealing with non-stationary EMG signals.

A RECURRENT PROBABILISTIC NEURAL NETWORK

The recurrent PNN, R-LLGMN (Tsuji et al., 2003), is based on the algorithm of continuous density hidden Markov model (CDHMM), which is a

combination of the GMM and the HMM (Rabiner, 1989). The probability density function (pdf) of input patterns is estimated using GMM; HMM is used simultaneously to model the time-varying characteristics in stochastic time series. In the R-LLGMN, recurrent connections are incorporated into the network structure to make efficient use of the time-varying characteristics of EMG signals. With the weight coefficients well trained using a learning scheme of the backpropagation through time (BPTT) algorithm, R-LLGMN can calculate posterior probabilities of the discriminating classes.

HMM-Based Dynamic Probabilistic Model

First, let us consider a dynamic probabilistic model, as shown in Figure 1. There are C classes in this model, and each class c ($c \in \{1, \dots, C\}$) is composed of K_c states. Suppose that, for the given time series $\tilde{\mathbf{x}} = \mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(T)$ ($\mathbf{x}(t) \in \mathfrak{R}^d$), at any time $\mathbf{x}(t)$ must occur from one state k of class c in the model. With this model, the posterior probability for class c , $P(c | \tilde{\mathbf{x}})$, is calculated as

$$P(c | \tilde{\mathbf{x}}) = \sum_{k=1}^{K_c} P(c, k | \tilde{\mathbf{x}}) = \sum_{k=1}^{K_c} \frac{\alpha_k^c(T)}{\sum_{c'=1}^C \sum_{k'=1}^{K_{c'}} \alpha_{k'}^{c'}(T)}. \quad (1)$$

Here, $\alpha_k^c(T)$ is the forward variable, which is defined as the probability for time series $(\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(T))$ to be generated from class c , and vector $\mathbf{x}(T)$ occurs from state k in class c . According to the forward algorithm (Rabiner, 1989), it can be derived as:

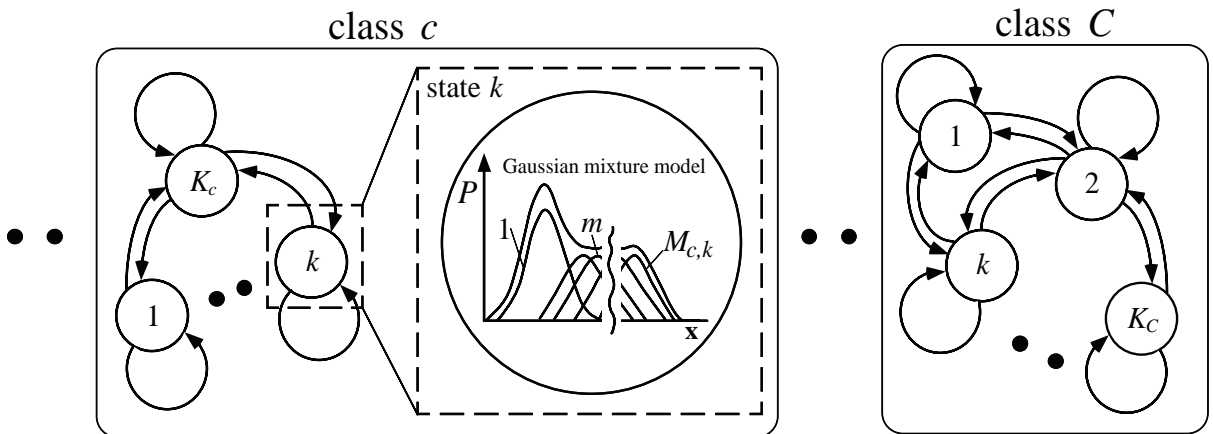
$$\alpha_k^c(1) = \pi_k^c b_k^c(\mathbf{x}(1)), \quad (2)$$

$$\alpha_k^c(t) = \sum_{k'=1}^{K_c} \alpha_{k'}^c(t-1) \gamma_{k',k}^c b_k^c(\mathbf{x}(t)) \quad (1 < t \leq T), \quad (3)$$

where $\gamma_{k',k}^c$ is the probability of the state changing from k' to k in class c , and $b_k^c(\mathbf{x}(t))$ is defined as the posterior probability for state k in class c corresponding to $\mathbf{x}(t)$. Also, the prior probability π_k^c is equal to $P(c, k) |_{t=0}$.

In this model, the posterior probability $b_k^c(\mathbf{x}(t))$ is approximated by summing up $M_{c,k}$ components of a Gaussian mixture distribution, and $\gamma_{k',k}^c b_k^c(\mathbf{x}(t))$ on the right side of (3) is derived in the form shown in Box 1, where $r^{(c,k,m)}$, $\boldsymbol{\mu}^{(c,k,m)} = (\mu_1^{(c,k,m)}, \dots, \mu_d^{(c,k,m)})^T$, $\Sigma^{(c,k,m)} \in \mathfrak{R}^{d \times d}$, $s_{ji}^{(c,k,m)}$ and $x_j(t)$ stands for the mixing proportion, the mean vector, the covariance matrix of each component $\{c, k, m\}$, the element of the inverse of covariance matrix $\Sigma^{(c,k,m)^{-1}}$, and the element of $\mathbf{x}(t)$.

Figure 1. HMM-based dynamic probabilistic model with C classes and K_c states in class c



The R-LLGMN is developed from the model defined above. For an input time series $\tilde{\mathbf{X}}$, the posterior probability for each class can be estimated with a well-trained R-LLGMN. The R-LLGMN network structure and learning algorithm are explained in the following.

Network Architecture

R-LLGMN is a five-layer recurrent NN with feedback connections between the fourth and the third layers, the structure of which is shown in Figure 2. First, the input vector series $\mathbf{x}(t) \in \mathfrak{R}^d$ ($t=1, \dots, T$) is preprocessed into the modified input series $\mathbf{X}(t) \in \mathfrak{R}^H$ as follows:

$$\mathbf{X}(t) = (1, \mathbf{x}(t)^T, x_1(t)^2, x_1(t)x_2(t), \dots, x_1(t)x_d(t), x_2(t)^2, x_2(t)x_3(t), \dots, x_2(t)x_d(t), \dots, x_d(t)^2)^T, \quad (5)$$

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 to activate each unit. The output of the h th ($h = 1, \dots, H$) unit in the first layer is defined as $^{(1)}O_h(t)$.

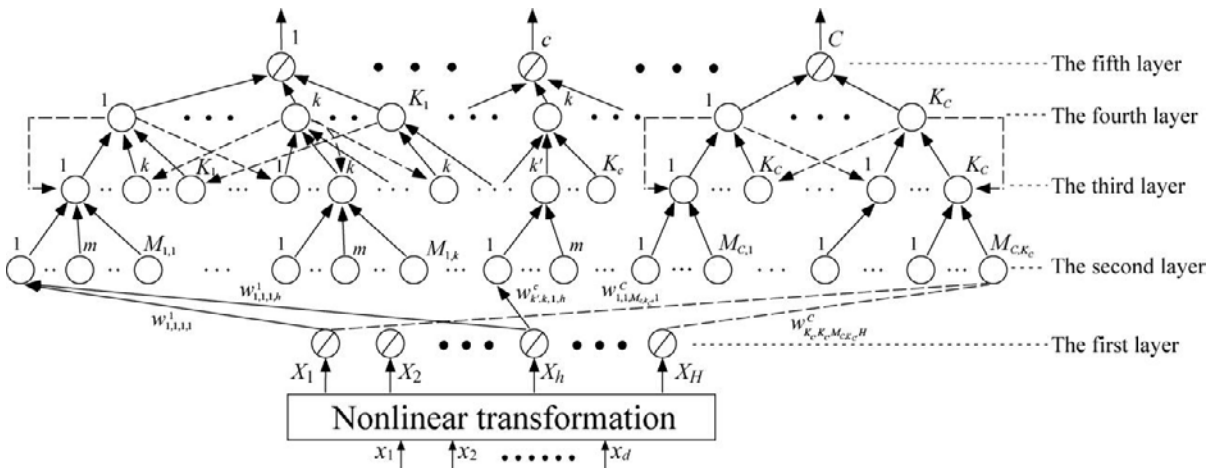
Unit $\{c, k, k', m\}$ ($c = 1, \dots, C; k', k = 1, \dots, K_c; m = 1, \dots, M_{c,k}$) in the second layer receives the output of the first layer, weighted by the coefficient $w_{k',k,m,h}^c$. The input $^{(2)}I_{k',k,m}^c(t)$ and the output $^{(2)}O_{k',k,m}^c(t)$ are defined as:

$$^{(2)}I_{k',k,m}^c(t) = \sum_{h=1}^H ^{(1)}O_h(t)w_{k',k,m,h}^c, \quad (6)$$

Box 1.

$$\begin{aligned} \gamma_{k',k}^c b_k^c(\mathbf{x}(t)) &= \sum_{m=1}^{M_{c,k}} \gamma_{k',k}^c r_{c,k,m} g(\mathbf{x}(t); \boldsymbol{\mu}^{(c,k,m)}, \Sigma^{(c,k,m)}) \\ &= \sum_{m=1}^{M_{c,k}} \gamma_{k',k}^c r_{c,k,m} (2\pi)^{-\frac{d}{2}} |\Sigma^{(c,k,m)}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \sum_{j=1}^d \sum_{l=1}^j (2 - \delta_{jl}) s_{jl}^{(c,k,m)} x_j(t) x_l(t) \right. \\ &\quad \left. + \sum_{j=1}^d \sum_{l=1}^d s_{jl}^{(c,k,m)} \mu_j^{(c,k,m)} x_l(t) - \frac{1}{2} \sum_{j=1}^d \sum_{l=1}^d s_{jl}^{(c,k,m)} \mu_j^{(c,k,m)} \mu_l^{(c,k,m)} \right], \end{aligned} \quad (4)$$

Figure 2. The structure of R-LLGMN



$${}^{(2)}O_{k',k,m}^c(t) = \exp({}^{(2)}I_{k',k,m}^c(t)), \quad (7)$$

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 . In (7), the exponential function is used in order to calculate the probability of the input pattern.

The outputs of units $\{c,k,k',m\}$ in the second layer are summed 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:

$${}^{(3)}I_{k',k}^c(t) = \sum_{m=1}^{M_{c,k}} {}^{(2)}O_{k',k,m}^c(t), \quad (8)$$

$${}^{(3)}O_{k',k}^c(t) = {}^{(4)}O_{k'}(t-1) {}^{(3)}I_{k',k}^c(t), \quad (9)$$

where ${}^{(4)}O_{k'}(0) = 1.0$ is for the initial phase. The recurrent connections between the fourth and the third layers play an important role in the process, which corresponds to the forward computation; see Equation (3).

The activation function in the fourth layer is described as:

$${}^{(4)}I_k^c(t) = \sum_{k'=1}^{K_c} {}^{(3)}O_{k',k}^c(t), \quad (10)$$

$${}^{(4)}O_k^c(t) = \frac{{}^{(4)}I_k^c(t)}{\sum_{c'=1}^C \sum_{k'=1}^{K_{c'}} {}^{(4)}I_{k'}^{c'}(t)}. \quad (11)$$

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:

$${}^{(5)}I^c(t) = \sum_{k=1}^{K_c} {}^{(4)}O_k^c(t), \quad (12)$$

$${}^{(5)}O^c(t) = {}^{(5)}I^c(t). \quad (13)$$

In R-LLGMN, the posterior probability of each class is defined as the output of the last layer.

After optimizing the weight coefficients $w_{k',k,m,h}^c$ between the first layer and the second layer, the NN can estimate the posterior probability of each class. Obviously, the R-LLGMN's structure corresponds well with the HMM algorithm. R-LLGMN, however, is not just a copy of HMM. The essential point of R-LLGMN is that the parameters in HMM are replaced by the weight coefficients $w_{k',k,m,h}^c$, and this replacement removes restrictions of the statistical parameter in HMM (e.g., $0 \leq$ the transition probability ≤ 1 , and standard deviations > 0). Therefore, the learning algorithm of R-LLGMN is simplified and can be expected to have higher generalization ability than that of HMMs. That is one of the major advantages of R-LLGMN.

A Maximum Likelihood Training 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. We assume 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 ${}^{(5)}O^c(T_n)$, is close enough to the teacher signal $\mathbf{T}^{(n)}$. The objective function for the network is defined as:

$$J = \sum_{n=1}^N J_n = - \sum_{n=1}^N \sum_{c=1}^C T_c^{(n)} \log({}^{(5)}O^c(T_n)). \quad (14)$$

The learning process attempts to minimize J , that is, to maximize the likelihood that each teacher vector $\mathbf{T}^{(n)}$ is obtained for the input stream $\tilde{\mathbf{x}}^{(n)}$.

The weight modification $\Delta w_{k',k,m,h}^c$ for $w_{k',k,m,h}^c$ is defined as:

$$\Delta w_{k',k,m,h}^c = \eta \sum_{n=1}^N \frac{\partial J_n}{\partial w_{k',k,m,h}^c}, \quad (15)$$

in a collective learning scheme, where $\eta > 0$ is the learning rate. Due to the recurrent connections in R-LLGMN, a learning algorithm based on the BPTT algorithm has been applied. It is supposed that the error gradient within a stream is accumulated, and weight modifications are only computed at the end of each stream; the error is then propagated backward to the beginning of the stream. The term $\frac{\partial J_n}{\partial w_{k',k,m,h}^c}$ in (15) can be defined as seen in Box 2, where $\Gamma_{(c',k''),(c,k)}$ is defined as:

$$\Gamma_{(c',k''),(c,k)} = \begin{cases} 1 & (c' = c; k'' = k) \\ 0 & (\text{otherwise}) \end{cases}, \quad (17)$$

and ${}^{(n)}\Delta_{k''}^{c'}(t)$ is the partial differentiation of J_n to ${}^{(4)}O_{k''}^{c'}(T_n - t)$ (see Box 3 for Equation (19)):

$${}^{(n)}\Delta_{k''}^{c'}(0) = \frac{T_n^{(n)}}{{}^{(5)}O_{k''}^{c'}(T_n)}, \quad (18)$$

It should be mentioned that all intermediate values of the R-LLGMN's feedforward computation are used in the calculation of Equations (16)-(19).

EMG PATTERN RECOGNITION USING R-LLGMN

The structure of the proposed EMG pattern recognition system is shown in Figure 3. This system consists of three parts in sequence: (1) EMG signal processing, (2) recurrent probabilistic neural network, and (3) discrimination rule.

1. EMG signal processing

The EMG signals are processed to extract the feature patterns. In this study, feature patterns extracted from filtered EMG signals

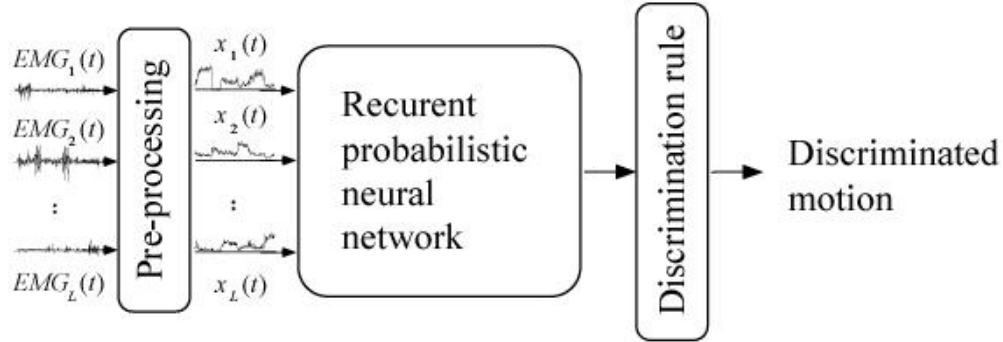
Box 2.

$$\begin{aligned} \frac{\partial J_n}{\partial w_{k',k,m,h}^c} &= \sum_{t=0}^{T_n-1} \sum_{c'=1}^C \sum_{k''=1}^{K_{c'}} {}^{(n)}\Delta_{k''}^{c'}(t) \frac{\partial^{(4)}O_{k''}^{c'}(T_n - t)}{\partial^{(4)}I_{k''}^c(T_n - t)} \frac{\partial^{(4)}I_{k''}^c(T_n - t)}{\partial^{(3)}O_{k',k}^c(T_n - t)} \\ &\quad \times \frac{\partial^{(3)}O_{k',k}^c(T_n - t)}{\partial^{(3)}I_{k',k}^c(T_n - t)} \frac{\partial^{(3)}I_{k',k}^c(T_n - t)}{\partial^{(2)}O_{k',k,m}^c(T_n - t)} \\ &\quad \times \frac{\partial^{(2)}O_{k',k,m}^c(T_n - t)}{\partial^{(2)}I_{k',k,m}^c(T_n - t)} \frac{\partial^{(2)}I_{k',k,m}^c(T_n - t)}{\partial w_{k',k,m,h}^c} \\ &= \sum_{t=0}^{T_n-1} \sum_{c'=1}^C \sum_{k''=1}^{K_{c'}} {}^{(n)}\Delta_{k''}^{c'}(t) (\Gamma_{(c',k''),(c,k)} - {}^{(4)}O_{k''}^{c'}(T_n - t)) \\ &\quad \left(\frac{{}^{(4)}O_{k''}^{c'}(T_n - t)}{{}^{(4)}I_{k''}^{c'}(T_n - t)} {}^{(4)}O_{k',k}^c(T_n - t - 1) {}^{(2)}O_{k',k,m}^c(T_n - t) X_h(T_n - t) \right), \end{aligned} \quad (16)$$

Box 3.

$$\begin{aligned} {}^{(n)}\Delta_{k''}^{c'}(t+1) &= \sum_{c''=1}^C \sum_{k'''=1}^{K_{c''}} {}^{(n)}\Delta_{k''' }^{c''}(t) \sum_{k''''=1}^{K_{c''}} (\Gamma_{(c'',k'''),(c',k''')} - {}^{(4)}O_{k''''}^{c''}(T_n - t)) \\ &\quad \left(\frac{{}^{(4)}O_{k''''}^{c''}(T_n - t)}{{}^{(4)}I_{k''''}^{c''}(T_n - t)} {}^{(3)}I_{k'',k'''}^{c'}(T_n - t) \right). \end{aligned} \quad (19)$$

Figure 3. Structure of the proposed EMG pattern recognition system



and raw EMG signals are used for motion discrimination. Also, the force information is extracted for motion onset detection and to determine the speed of the motion classified.

2. *Recurrent probabilistic neural network*

The R-LLGMN described in the previous section is employed for motion discrimination. 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 with length T , the output ${}^{(5)}O^c(T)$ ($c = 1, \dots, C$) presents the posterior probability of each discriminating motion.

3. *Discrimination rule*

In order to recognize whether the motion has really occurred or not, the force information $\sigma(t)$ is compared with a prefixed motion appearance threshold M_d . The motion is considered to have occurred if $\sigma(t)$ exceeds M_d . The entropy of R-LLGMN's outputs is also calculated to prevent the risk of mis-discrimination. The entropy is defined as:

$$H(t) = - \sum_{c=1}^C {}^{(5)}O^c(t) \log_2 {}^{(5)}O^c(t). \quad (20)$$

If the entropy $H(t)$ is less than the discrimination threshold H_d , the specific motion with the largest probability is determined

according to the Bayes decision rule. If not, the determination is suspended.

The discriminated motion can be used as *control commands* for HMIs, for example, powered prosthetic limbs.

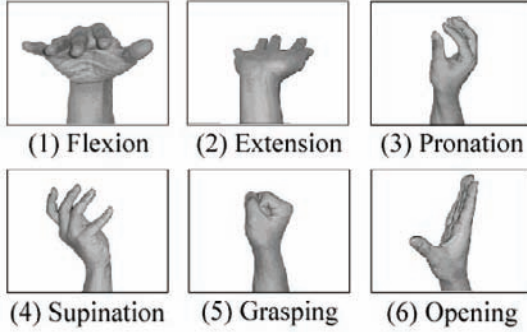
Experimental Conditions

Five subjects (amputee subjects A and B, and normal subjects C, D, and E) participated in this study. Six pairs of Ag/AgCl electrodes (NT-511G: NIHON KOHDEN Corp.) with conductive paste were attached to the 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 subjects were asked to continuously perform six motions ($C = 6$): flexion, extension, supination, pronation, hand grasping, and hand opening. The motions are shown in Figure 4.

The differential EMG signals were amplified (70 [dB]) and filtered out with a low-pass filter (cut-off frequency: 100 [Hz]) by a multi-telemeter (Web5000: NIHON KOHDEN Corp.), as shown in Figure 5, then digitized by an A/D converter (sampling frequency: 200 [Hz]; quantization: 12 [bits]).

In the experiments, the network structure of R-LLGMN is set as ($C = 6$), $K_c = 1$ ($c = 1, \dots, C$),

Figure 4. Six motions used in the experiments



and the component for each unit in the third layer is one. The parameters used are chosen to make conditions of comparison experiments as equal as possible. The lengths of training sample streams, T_n ($n = 1, \dots, N$), are set as T , which was determined with respect to the EMG features. In accordance with previous researches on EMG pattern classification (Tsuji et al., 1993; Fukuda et al., 2003), the determination threshold H_d was set to 0.5, and the motion appearance threshold M_d to 0.2. All pattern recognition experiments were conducted off-line.

Pattern Recognition of Filtered EMG Signals

First, motion discrimination experiments using filtered EMG signals were conducted to examine the performance of the proposed method. In the experiments, the training sample consists of 20 EMG patterns extracted from the filtered EMG signals for each motion.

Six channels of EMG signals ($L = 6$) are rectified and filtered by a second-order Butterworth filter (cut-off frequency: 1 [Hz]). The filtered EMG signals are defined as $FEMG_l(t)$ ($l = 1, \dots, L$) and are normalized to make the sum of L channels equal to 1:

Figure 5. The multi-telemeter (Web5000) and electrodes (NT-511G) used in the experiments



$$x_l(t) = \frac{FEMG_l(t) - FEMG_l^{st}}{\sum_{l'=1}^L FEMG_{l'}(t) - FEMG_{l'}^{st}} \quad (l = 1, \dots, L), \quad (21)$$

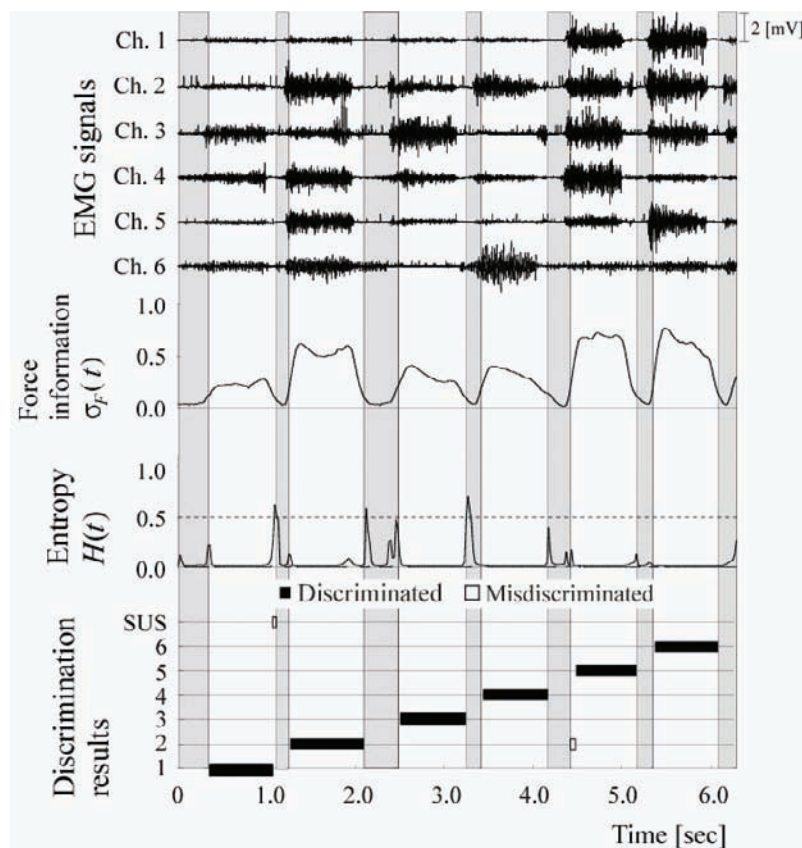
where $FEMG_l^{st}$ is the mean value of $FEMG_l(t)$ measured while the arm is relaxed. The feature vector $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_L(t)]$ is used for the input of the neural classifier, R-LLGMN, where the dimension of R-LLGMN's input, d , is set as $d = L$. In this study, it is assumed that the amplitude level of EMG signals varies in proportion to muscle force. Force information $\sigma_F(t)$ for the input vector $\mathbf{x}(t)$ is defined as follows:

$$\sigma_F(t) = \frac{1}{L} \sum_{l=1}^L \frac{FEMG_l(t) - FEMG_l^{st}}{FEMG_l^{\max} - FEMG_l^{st}}, \quad (22)$$

where $FEMG_l^{\max}$ is the mean value of $FEMG_l(t)$ measured while maintaining the maximum arm voluntary contraction.

An example of the discrimination results of subject A is shown in Figure 6. The subject was an amputee (51-year-old male), whose forearm, three cm from the left wrist joint, was amputated when he was 18 years old as the result of an accident. He has never used EMG controlled prosthetic limbs and usually uses a cosmetic hand. In the experiments, he was asked to perform six motions in the order continuously for six seconds. Figure 6

Figure 6. Example of the discrimination results for filtered EMG signals (subject A)



plots six channels of the input EMG signals, the force information $\sigma_f(t)$, the entropy $H(t)$, and the discrimination results. The labels of the vertical axis in the discrimination results correspond to the motions shown in Figure 4, and SUS means that the determination was suspended. The gray areas indicate that no motion was determined because the force information was less than M_d . Incorrect determination was eliminated using the entropy. Figure 6 demonstrates that the proposed method achieves high discrimination accuracy with filtered EMG signals during continuous motion.

The discrimination accuracy for five subjects was then investigated, and LLGMN and an MLP classifier were used for comparison. The same preprocessing method and discrimination rule

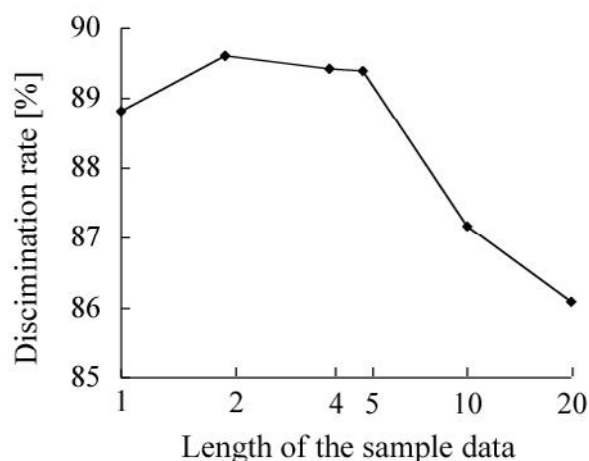
were applied to the experiments using LLGMN and MLP. The number of units in the input layer of LLGMN was equal to the dimension of input signal (L). Units in the hidden layer corresponded to the Gaussian components in GMM, the number of which was set in the same manner as for the R-LLGMN. The output layer included C units, and each unit gave the posterior probability for the input pattern. In contrast, MLP had four layers (two hidden layers), and the units of the layers were set at 6, 10, 10, and 6. Each output of MLP corresponded to a motion, and all six outputs were normalized to make the sum of all outputs equal 1.0 for comparison with R-LLGMN and LLGMN. The learning procedure of MLP continued until the sum of the square error was less than 0.01, where the learning rate was 0.01. However, if the

Table 1. Discrimination results of five subjects with filtered EMG signals

Type of the methods		MLP	LLGMN	R-LLGMN
Subject A (Amputee)	DR	73.4	94.0	99.1
	SD	7.9	5.5	0.0
Subject B (Amputee)	DR	46.5	82.8	89.3
	SD	12.3	0.0	0.4
Subject C (Normal)	DR	44.2	88.5	93.0
	SD	10.4	0.0	0.1
Subject D (Normal)	DR	69.8	88.7	93.5
	SD	10.0	0.2	0.0
Subject E (Normal)	DR	69.2	89.3	92.8
	SD	7.0	0.1	0.0

DR : Discrimination rate [%], SD : Standard deviation [%]

Figure 7. Discrimination rates for various data lengths (subject B)



sum of the square error after 50,000 iterations was still not less than 0.01, the learning procedure was stopped. In all three methods, ten different sets of initial weights (all randomized between [0, 1]) were used.

Discrimination rate, which is defined as the ratio of correctly classified data to the total test set, is used to evaluate discrimination accuracy of three methods. The mean values and the standard deviations of the discrimination rates are shown in Table 1. It can be seen that R-LLGMN achieved the best discrimination rate among all three methods and had the smallest standard deviation.

Also, the classification results were examined by altering the experiment conditions, such as the length of sample data. Experiments were performed using various lengths of sample data. For each sample data, R-LLGMN was trained with ten different sets of initial weights, which were randomly chosen in the range [0, 1]. The mean values of the discrimination rates for each length are shown in Figure 7, where the standard deviations are all very small, close to 0. It can be seen from Figure 7 that the discrimination rate maintains a high level when the sample data is of an appropriate length (T). However, if $T > 5$, it is

too long to train R-LLGMN using filtered EMG signals. The discrimination rate tends to deteriorate because R-LLGMN, which was trained using the long-length sample data, failed to discriminate the switching of motions.

Pattern Recognition of Raw EMG Signals

This subsection presents pattern recognition experiments of time series of raw EMG signals. In the previously proposed methods for classifying the intended motion of an operator, the filtered or smoothed EMG signals (Fukuda et al., 1997, 2003; Kelly et al., 1990; Tsuji et al., 1993, 2000) or the extracted characteristics in a fixed time window (Hiraiwa et al., 1989; Hudgins et al., 1993) have been used as the input vector to the NN classifier. However, these signal-processing steps result in considerable phase delay and time delay caused by the low-pass filtering and the time window operation. To avoid such delay, raw EMG signals without any preprocessing are used as the input to R-LLGMN. The experiments were performed with the subjects (A, B, C, D, and E) who had experience in manipulating the EMG signals.

As raw EMG signals, six channels of EMG signals ($L=6$) sampled from the input of multi-telemeter are denoted by $REMG_l(t)$ ($l = 1, \dots, L$). For the case of raw EMG signals, force information $\sigma_R(t)$ is obtained calculating moving average within the length T :

$$\sigma_R(t) = \frac{1}{L} \sum_{l=1}^L \frac{\overline{REMG_l(t)}}{\overline{REMG_l}^{\max}}, \quad (23)$$

$$\overline{REMG_l(t)} = \frac{1}{T} \sum_{j=0}^{T-1} |REMG_l(t-j)|, \quad (24)$$

where $\overline{REMG_l}^{\max}$ is the premeasured integral EMG of each channel under the maximum voluntary contraction. Also, it should be noted that $REMG_l(t-j) = 0$ when $t-j < 0$.

The input vector $\mathbf{x}(t)$ ($t = 1, \dots, T$) of R-LLGMN is normalized $REMG_l(t)$ with $\sigma_R(t)$ as

$$x_i(t) = \sigma_R^{-1}(T) REMG_i(t). \quad (25)$$

Here, the normalization enables R-LLGMN to discriminate motions from a pattern of all channels as well as from the amplitude of the raw EMG signals.

In pattern recognition experiments of raw EMG signals, the length of training sample stream T is set as 20. Eight sample streams are used for each motion. The threshold for motion onset detection M_d is 0.155.

Figure 8 provides an example of the classification results of subject A. The figure shows six channels of the raw EMG signals, the force information $\sigma_R(t)$, the entropy $H(t)$ calculated from the output probability of R-LLGMN, and the classification results of the R-LLGMN. The discrimination rate was about 95.5% in this experiment. It can be seen that R-LLGMN generates acceptable classification results during continuous motion, and the entropy is low during motions except for the motion one (Flexion). It indicates that R-LLGMN can discriminate the hand and forearm motions from the raw EMG signals, even for control purposes.

Comparisons were conducted with discrimination results of MLP, LLGMN, and R-LLGMN using filtered EMG signals. It should be noted that due to the stochastic nature of raw EMG signals, MLP and LLGMN could not learn motion patterns of raw EMG signals. The network structures of MLP and LLGMN were set to the same as those in pattern recognition experiments of filtered EMG, and the beginning and ending of motions were recognized according to the force information $\sigma_R(t)$. The discrimination threshold H_d was not used in the comparison, so that all classification results were used for comparison. Each experiment was repeated ten times with different randomly chosen initial weights. Table 2 depicts the mean values and the standard deviations of the discrimina-

Figure 8. Example of the discrimination result for raw EMG signals (subject A)

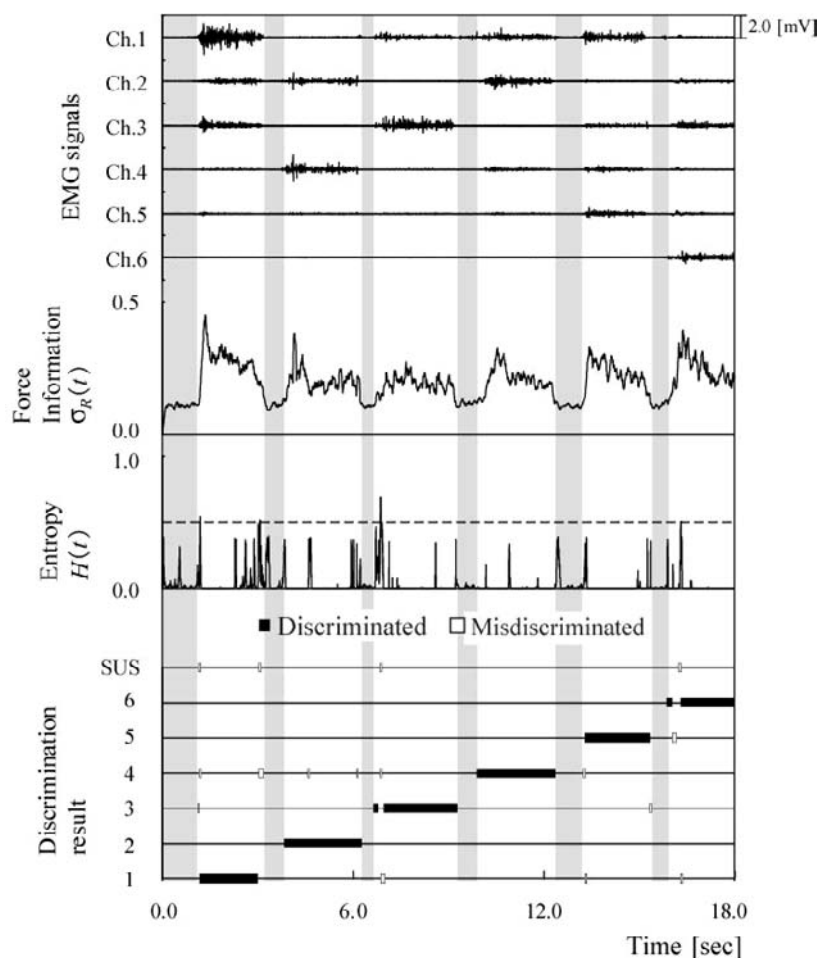


Table 2. Motion discrimination results for raw EMG signals, comparing with methods using filtered EMG signals

Type of the methods		MLP (Filtered EMG)	LLGMN (Filtered EMG)	R-LLGMN (Filtered EMG)	R-LLGMN (Raw EMG)
Subject A (Amputee)	DR	66.1	89.8	96.1	93.8
	SD	14.0	0.0	0.0	0.0
Subject B (Amputee)	DR	70.1	89.3	92.5	91.2
	SD	10.8	0.0	0.0	1.3
Subject C (Normal)	DR	80.5	82.9	94.2	94.1
	SD	8.1	0.0	0.0	0.4
Subject D (Normal)	DR	78.9	88.3	97.4	90.4
	SD	4.1	0.0	0.0	0.9
Subject E (Normal)	DR	75.8	85.9	90.7	91.0
	SD	4.5	0.0	0.0	1.8

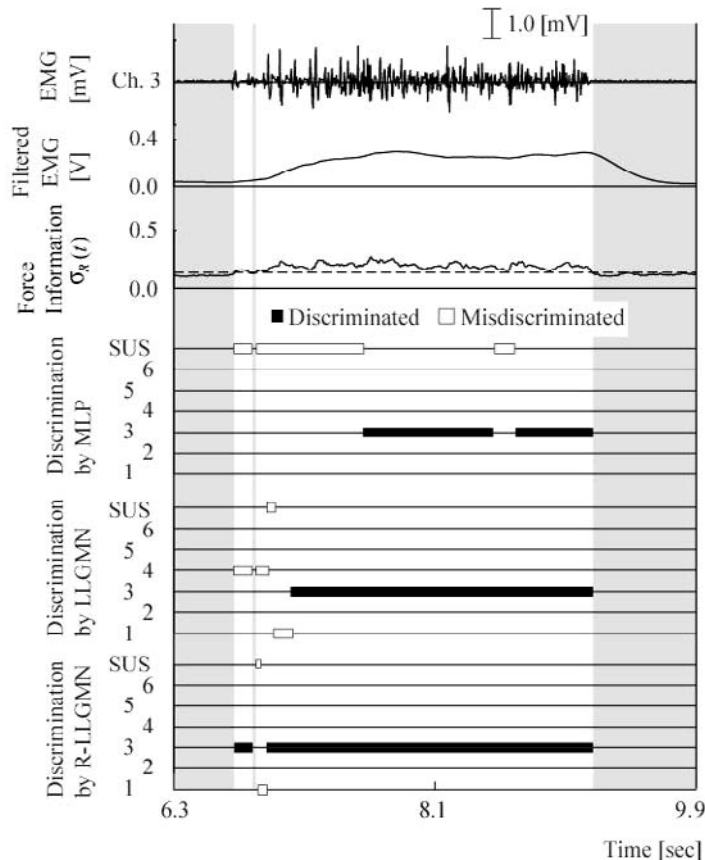
DR : Discrimination rate [%], SD : Standard deviation [%]

tion rates for five subjects. Due to the filtering processes, onsets of the filtered EMG signals are delayed, and the EMG patterns vary significantly in time domain during the transient phase. Since MLP cannot deal well with time-varying patterns, MLP's discrimination result is the worst among these methods. Although LLGMN shows better discrimination accuracy than MLP due to the statistical model incorporated in its structure, it still provides poor discrimination accuracy since the model is static. Consequently, it can be concluded that phase delay due to the filtering processes is one of the major causes of degradation in the discrimination results in cases of MLP and LLGMN. In contrast, R-LLGMN provides superior discrimination results for both the filtered EMG signals and the raw EMG. Also, we found

that patterns of raw EMG signals are much more complicated than that of filtered EMG signals, and training and estimation of R-LLGMN using raw EMG signals are more difficult. Therefore, the classification performance of R-LLGMN with filtered EMG signals is a little higher than that using raw EMG signals. However, since no signal processing is used, the latter has a faster response. There is thus a trade-off between discrimination accuracy and response speed.

The response time of raw EMG-based motion discrimination was further investigated, and the proposed method and traditional classifiers (MLP and LLGMN) were compared. Figure 9 illustrates the signals magnified from 6.3 s to 9.9 s in Figure 8 during the wrist extension motion. This figure depicts the EMG signal of the channel 3, the fil-

Figure 9. Changes of the discrimination results by three types of neural networks (subject A)



tered EMG signal that is rectified and filtered out by the second-order Butterworth low-pass filter (cut-off frequency: 1.0 [Hz]), the force information $\sigma_k(t)$, and the discrimination results of three comparison methods. The MLP and LLGMN used the features extracted from filtered EMG signals as the input, while the R-LLGMN achieved motion discrimination based on the raw EMG signals. It can be seen from the figure that there is a considerable phase delay between the raw EMG and the filtered EMG signals, which causes the misdiscrimination in the results of MLP and LLGMN. In contrast, using the raw EMG signals, R-LLGMN achieves higher discrimination accuracy than the others, and a correct classification is made just after the beginning of motion. It was also found that the discrimination rates of both MLP and LLGMN decreased considerably when the cut-off frequency of the low-pass filter increased. The increase of the cut-off frequency results in filtered EMG signals containing high frequency components, so that the learning of the NNs becomes very difficult.

Then, discrimination accuracy during the beginning and ending of motions was investigated. In these experiments, EMG signals during 100 msec from onset and 100 msec before ending of each motion were used. Similarly, MLP and

LLGMN using filtered EMG signals were used for comparison. Table 3 presents the discrimination results for five subjects using three different methods. The mean values and the standard deviations of the discrimination rates are computed for ten randomly chosen initial weights. From this table, it can be seen that R-LLGMN attained the best discrimination rates during the beginning and ending of motions; therefore, the R-LLGMN provides superior response performance.

DISCUSSIONS

A new EMG pattern recognition method using R-LLGMN is proposed to improve discrimination accuracy when dealing with non-stationary EMG signals. R-LLGMN performs both the filtering process and pattern classification within the same network architecture, so the proposed method outperforms the previous methods with filtered EMG and raw EMG patterns. What is even more encouraging is that the response time of discrimination results can also be shortened by using raw EMG signals.

In the studies on human-machine interfaces (HMIs), it is widely believed that the response time is an important aspect, especially for practi-

Table 3. Discrimination results for five subjects

Type of the methods		BPNN	LLGMN	R-LLGMN
Subject A (Amputee)	DR	30.4	58.3	89.6
	SD	13.6	4.1	0.3
Subject B (Amputee)	DR	67.2	63.7	75.0
	SD	11.0	0.2	0.0
Subject C (Normal)	DR	49.4	50.0	91.7
	SD	11.5	0.0	0.0
Subject D (Normal)	DR	52.8	58.3	73.8
	SD	3.8	0.0	0.0
Subject E (Normal)	DR	64.4	63.3	66.7
	SD	3.0	0.0	0.0

DR : Discrimination rate [%], SD : Standard deviation [%]

cal application systems. For HMIs used in daily activities, it has been mentioned that techniques for real-time classification are needed in order to decrease global time delay of response, which would reduce the operator's mental burden and increase the range of applications and number of potential users (Chang et al., 1996; Vuskovic & Du, 2002). A classification system based on digital signal processors (DSP) was used to realize the pattern classification algorithm for fast processing (Chang et al., 1996). Vuskovic and Du (2002) attempted to simplify a fuzzy ARTMAP network used for EMG classification, which resulted in overall smaller computational times. On the other hand, since the EMG signals include high-frequency components, adequate signal processes such as low-pass filtering are necessary in order to extract meaningful information for HMIs. Actually, this low-pass filtering process increases the time delay.

In contrast to these previous studies, which focus on reduction of the computational time (complexity) of classifiers, a pattern recognition technique that directly uses raw EMG signals is an interesting choice. Given the experimental results in the previous section, it is expected that improved response performance is possible by adopting the proposed raw EMG pattern recognition scheme into traditional HMIs, which use filtered EMG patterns (Fukuda et al., 1997, 2003; Kelly et al., 1990; Tsuji et al., 1993, 2000). Further studies should focus on this idea.

This chapter introduced R-LLGMN in order to make effective use of the non-stationary (time-varying) characteristics in EMG signals. In recent years, time-frequency analysis has attracted increasing attention for representing the non-stationary essence of frequency domain (Englehart et al., 1999, 2001; Hussein & Granat, 2002). Since the wavelet transform results in a good time-frequency resolution, it has become a very popular feature extraction method for time-frequency representation of EMG signals. Based on the idea of building prior information

into neural network design, the algorithm of wavelet transform can be incorporated into the probabilistic neural network, so that the PNNs could process frequency information of EMG signals more effectively.

SUMMARY

This chapter proposes a new EMG pattern recognition method based on a recurrent log-linearized Gaussian mixture network (R-LLGMN). Because of the recurrent connections in the R-LLGMN's structure, the temporal information of EMG signals can be used to improve discrimination accuracy.

To examine the discrimination capability and accuracy of the proposed method, EMG pattern recognition experiments were conducted with five subjects. In the experiments, the proposed method achieved high discrimination accuracy for varying EMG signals, and its discrimination results are superior to those of the LLGMN and MLP classifiers. We found that the discrimination results change when different lengths of sample stream T are used. The length T should be well modulated according to the input signals. For example, to discriminate filtered EMG signals, T should be less than five.

Even more encouraging is the outcome of EMG pattern recognition experiments using the non-stationary time series of raw EMG signals. Results of these experiments demonstrate that R-LLGMN performs both the filtering process and pattern recognition within the same network architecture and can realize a relatively high discrimination rate that is good enough for control purposes. It should be noted that there is a trade-off between discrimination accuracy and response speed when using R-LLGMN as a classifier. In practical applications, such as prosthetic control, the latter may be preferred.

REFERENCES

- Chang, G. C., Kang, W. J., Luh, J. J., Cheng, C. K., Lai, J. S., Chen, J. J., & Kuo, T. S. (1996). Real-time implementation of electromyogram pattern recognition as a control command of man-machine interface. *Medical Engineering & Physics*, 18, 529-537.
- Englehart, E., Hudgins, B., & Parker, A. (2001). A wavelet-based continuous classification scheme for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, 48, 302-311.
- Englehart, E., Hudgins, B., Parker, A., & Stevenson, M. (1999). Classification of the myoelectric signal using time-frequency based representations. *Medical Engineering & Physics*, 21, 431-438.
- Fukuda, O., Tsuji, T., & Kaneko, M. (1997). An EMG controlled robotic manipulator using neural networks. *Proceedings of the IEEE International Workshop on Robot and Human Communication*, 442-447.
- Fukuda, O., Tsuji, T., & Kaneko, M. (1999). An EMG-controlled pointing device using a neural network. *Proceedings of the 1999 IEEE International Conference on Systems, Man, and Cybernetics*, 4, 63-68.
- Fukuda, O., Tsuji, T., Kaneko, M., & Ohtsuka, A. (2003). A human-assisting manipulator teleoperated by EMG signals and arm motions. *IEEE Transactions on Robotics and Automation*, 19, 210-222.
- Hiraiwa, A., Shimohara, K., & Tokunaga, Y. (1989). EMG pattern analysis and classification by neural network. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, 3, 1113-1115.
- Hudgins, B., Parker, P., & Scott, R.N. (1993). A new strategy for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, 40, 82-94.
- Hussein, S. E., & Granat, M. H. (2002). Intention detection using a neuro-fuzzy EMG classifier. *IEEE Engineering in Medicine and Biology Magazine*, 21(6), 123-129.
- Kang, W. J., Shiu, J. R., Cheng, C. K., Lai, J. S., Tsao, H. W., & Kuo, T. S. (1995). The application of cepstral coefficients and maximum likelihood method in EMG pattern recognition. *IEEE Transactions on Biomedical Engineering*, 42, 777-785.
- Kelly, M. F., Parker, P. A., & Scott, R. N. (1990). The application of neural networks to myoelectric signal analysis: A preliminary study. *IEEE Transactions on Biomedical Engineering*, 37, 221-230.
- Lamounier, E., Soares, A., Andrade, A., & Carrijo, R. (2002). A virtual prosthesis control based on neural networks for EMG pattern classification. *Proceedings of the 6th IASTED International Conference on Artificial Intelligence and Soft Computing*, 456-461.
- Lusted, H. S., & Knapp, R. B. (1996, October). Controlling computers with neural signals. *Scientific American*, 82-87.
- Rabiner, L.R. (1989). A tutorial on hidden Markov model and selected applications in speech recognition. *Proceedings of the IEEE*, 77, 257-286.
- Specht, D.F. (1990). Probabilistic neural networks. *Neural Networks*, 3, 109-118.
- Tsuji, T., Bu, N., Fukuda, O., & Kaneko, M. (2003). A recurrent log-linearized Gaussian mixture network. *IEEE Transactions on Neural Networks*, 14, 304-316
- Tsuji, T., Fukuda, O., Ichinobe, H., & Kaneko, M. (1999). A log-linearized Gaussian mixture network and its application to EEG pattern classification. *IEEE Transactions on Systems, Man,*

and Cybernetics, Part C: Applications and Reviews, 29, 60-72.

Tsuji, T., Fukuda, O., Kaneko, M., & Ito, K. (2000). Pattern classification of time-series EMG signals using neural networks. *International Journal of Adaptive Control and Signal Processing*, 14, 829-848.

Tsuji, T., Ichinobe, H., Ito, K., & Nagamachi, M. (1993). Discrimination of forearm motions from EMG signals by error back propagation typed neural network using entropy (in Japanese).

Transactions of the Society of Instrument and Control Engineers, 29, 1213-1220.

Vuskovic, M., & Du, S. (2002). Classification of prehensile EMG patterns with simplified fuzzy ARTMAP networks. *Proceedings of the 2002 International Joint Conference on Neural Networks*, 2539-2545.

Zhang, G.D. (2000). Neural network for classification: A survey. *IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews*, 30, 451-462.