

A Novel Pattern Classification Method for Multivariate EMG Signals Using Neural Network

Nan Bu, Jun Arita, and Toshio Tsuji

Department of Artificial Complex System Engineering,
Hiroshima University, Higashi-Hiroshima 739-8527, Japan
bu@ieee.org
<http://www.bsyst.hiroshima-u.ac.jp>

Abstract. Feature extraction is an important issue in electromyography (EMG) pattern classification, where feature sets of high dimensionality are always used. This paper proposes a novel classification method to deal with high-dimensional EMG patterns, using a probabilistic neural network, a reduced-dimensional log-linearized Gaussian mixture network (RD-LLGMN) [1]. Since RD-LLGMN merges feature extraction and pattern classification processes into its structure, lower-dimensional feature set consistent with classification purposes can be extracted, so that, better classification performance is possible. To verify feasibility of the proposed method, phoneme classification experiments were conducted using frequency features of EMG signals measured from mimetic and cervical muscles. Filter banks are used to extract frequency features, and dimensionality of the features grows significantly when we increase resolution of frequency. In these experiments, the proposed method achieved considerably high classification rates, and outperformed traditional methods that are based on principle component analysis (PCA).

1 Introduction

Electromyography (EMG) pattern classification has been used to devise elaborate human-machine interfaces for people with physical disabilities [2], [3]. In the recent years, multiple channels of amplitude and/or frequency information of EMG signals have been increasingly used for EMG pattern classification to achieve improved classification performance and to conduct multifunction myoelectric control [3], [4]. Dimensionality of feature grows dramatically when we increase number of electrodes and frequency resolution, and pattern classification is frequently confronted with high-dimensional data. As for high-dimensional pattern classification, feature extraction is usually conducted prior to a classification process, in order to find a compact feature set to avoid exhaustive computation and to reduce statistically redundant/irrelevant attributes to improve classification performance [5], [6].

In feature extraction techniques, original features (d -dimension) are projected into an m -dimensional space, where $m < d$, and the m axes of the reduced feature

space are determined according to some optimal criterion. Principal component analysis (PCA) is one of most successful feature extraction techniques, and a reconstruction error is utilized as the optimal criterion [6]. Up to present, PCA has been widely used in pattern classification processes, such as face recognition and text classification [7]. For EMG pattern classification, Du et al. and Englehart et al. have applied PCA to EMG features for dimensionality reduction, and neural networks (NNs) are used as classifiers [3], [4], [8].

Although PCA shows promising characteristics as a feature extractor in pattern classification tasks, it still suffers from some intrinsic limitations. For instance, optimal criterion of PCA is not directly related to training criteria of its counterparts for pattern classification. Since training of the classification part always aims to realize low error probabilities, it may not always be possible for PCA to extract features in a reduced form, containing high discriminant information [9], [10]. On the other hand, it should be noticed that, in the existing methods, training processes for PCA and the classification part are made separately. Optimization of the whole pattern classification process is almost impossible, and it is hard to gain a high performance of classification, especially in practical applications, such as for EMG patterns.

To deal with these problems, Bu and Tsuji have proposed a probabilistic neural network, a reduced-dimensional log-linearized Gaussian mixture network (RD-LLGMN), for high-dimensional pattern classification [1]. RD-LLGMN uses orthogonal transformation to project the original feature space into a lower-dimensional space, and then calculates posterior probabilities with a Gaussian mixture model (GMM) in the projected lower-dimensional space for classification. Also, parameters in the network is trained with a single criterion, i.e., minimizing an error probability, it is expected that such training algorithm may yield better classification performance [1], [9].

In this paper, a novel EMG pattern classification method is proposed based on RD-LLGMN. With RD-LLGMN, it is expected that discriminative features can be extracted from high-dimensional EMG patterns, and an efficient and consistent classification is possible. Moreover, as an application, phoneme classification experiments are presented, in which frequency information of EMG signals is extracted as feature pattern using filter banks. The rest of this paper is organized as follows. Section 2 briefly introduces the conception and network structure of RD-LLGMN. Then, in Section 3, the EMG pattern classification method is proposed. Phoneme classification and experimental results are presented in Section 4. Finally, Section 5 gives a conclusion of this paper.

2 RD-LLGMN [1]

RD-LLGMN provides a novel feature extraction approach to find discriminant features of a reduced size, and calculates posterior probabilities for classification. There are two basic ideas in this NN: 1) orthogonal transformation, which projects the original feature space into a lower-dimensional space, and 2) GMM, which estimates probability distribution of patterns in the projected space. This

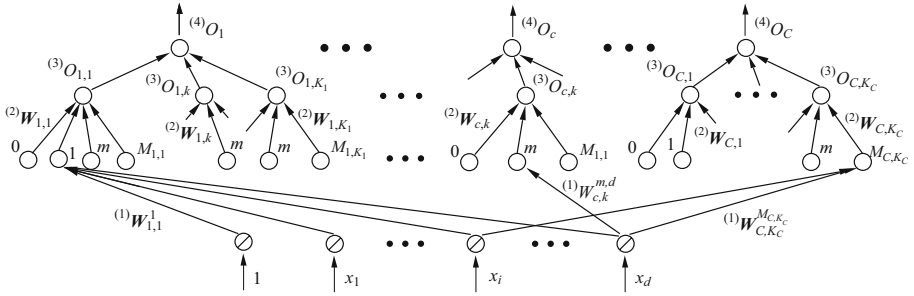


Fig. 1. Structure of RD-LLGMN

network combines the feature extraction process with the classification part, and is trained in a fashion of minimum classification error (MCE) learning [11], which enables the classification part to realize a low error probability.

RD-LLGMN is a four-layer NN, the structure of which is shown in Fig. 1. Given an input vector $\mathbf{x} \in \mathbb{R}^d$, the first layer consists of $d + 1$ units, with one unit has a bias input of 1. Identity function is used for activation of each unit. Let $(1)O_i$ denotes the output of the i th unit in the first layer, we have

$$(1)O_i = \begin{cases} 1, & i = 0 \\ x_i, & i = 1, 2, \dots, d \end{cases} \quad (1)$$

where x_i ($i = 1, 2, \dots, d$) is the element of \mathbf{x} .

In the second layer, the unit $\{c, k, 0\}$, ($c = 1, \dots, C$; $k = 1, \dots, K_c$), is a bias unit, and its output $(2)O_{c,k}^0 = 1$. On the other hand, the unit $\{c, k, m\}$ ($m = 1, \dots, M_{c,k}$) receives the output of the first layer weighted by $(1)W_{c,k}^m$. The input $(2)I_{c,k}^m$ and the output $(2)O_{c,k}^m$, for ($m \neq 0$), are defined as follows:

$$(2)I_{c,k}^m = \sum_{i=1}^d (1)O_i (1)W_{c,k}^{m,i}, \quad (2)$$

$$(2)O_{c,k}^m = ((2)I_{c,k}^m)^2, \quad (3)$$

where C is the number of classes under consideration, K_c is the number of components of the Gaussian mixture distribution corresponding to the class c , $M_{c,k}$ is the number of dimension of component k in class c . Through this layer, vector $\mathbf{x} \in \mathbb{R}^d$ is projected into $M_{c,k}$ -dimension spaces, $M_{c,k} < d$.

The unit $\{c, k\}$ in the third layer sums up outputs of the second layer weighted by coefficients $(2)W_{c,k}^m$. The relationships between the input of unit $\{c, k\}$ in the third layer $(3)I_{c,k}$ and the output $(3)O_{c,k}$ are defined as

$$(3)I_{c,k} = \sum_{m=0}^{M_{c,k}} (2)O_{c,k}^m (2)W_{c,k}^m, \quad (4)$$

$$(3)O_{c,k} = \frac{\exp[(3)I_{c,k}]}{\sum_{c'=1}^C \sum_{k'=1}^{K_{c'}} \exp[(3)I_{c',k'}]}. \quad (5)$$

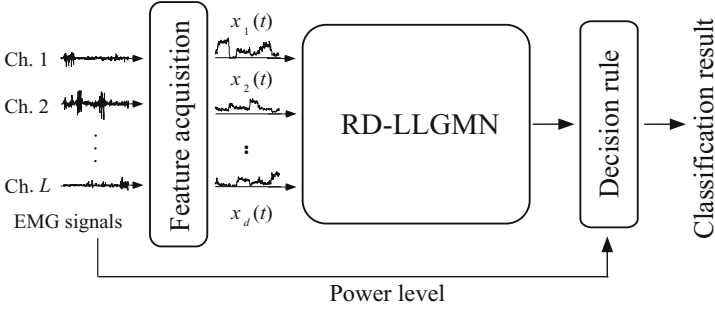


Fig. 2. Schematic view of the proposed method

In the third layer, RD-LLGMN calculates posterior probability of each Gaussian component $\{c, k\}$ using reduced-dimensional features.

The fourth layer consists of C units corresponding to the number of classes. Unit c sums up outputs of K_c components $\{c, k\}$ in the third layer. The function between the input and the output is described as

$${}^{(4)}O_c = {}^{(4)}I_c = \sum_{k=1}^{K_c} {}^{(3)}O_{c,k}. \quad (6)$$

After optimizing the weight coefficients with an MCE-based training algorithm, RD-LLGMN's output, ${}^{(4)}O_c$, can estimate the posterior probability of class c .

3 EMG Pattern Classification Using RD-LLGMN

The proposed EMG pattern classification method, as shown in Fig. 2, consists of three parts: (1) feature acquisition, (2) RD-LLGMN, and (3) decision rule.

L channels of EMG signals are recorded using surface electrodes attached on muscles. From raw EMG signals, we can calculate amplitude and frequency information to represent EMG features. Suppose that, from each channel, Z features are obtained, the dimensionality of EMG features, d , equals $L \times Z$. Also, normalization method can be applied to decrease variability of EMG features.

For pattern classification, RD-LLGMN described in Section 2 is employed. Using samples labeled with the corresponding motions, RD-LLGMN learns the non-linear mapping between the EMG patterns and the motions. The normalized patterns $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_d(t)]^T$ are used as input to RD-LLGMN.

In this method, we assumed that the amplitude level of EMG signals is changed in proportion to muscle force. A power level is defined as

$$F_{\text{EMG}}(t) = \frac{1}{S} \sum_{s=1}^S \frac{E_s(t) - \overline{E}_s^{st}}{E_s^{\text{max}} - \overline{E}_s^{st}}, \quad (7)$$

where S indicates the number of electrodes, $E_s(t)$ is the filtered signal (cut-off frequency: 1 Hz) of rectified raw EMG directly measured from the electrode s

($s = 1, 2, \dots, S$), \overline{E}_s^{st} is the mean value of $E_s(t)$, which is measured while relaxing the muscles, and E_s^{max} is the mean value of $E_s(t)$ measured under the maximum voluntary contraction (MVC). $F_{EMG}(t)$ indicates the force information, and is used to recognize whether the motion has really happened or not, by comparing $F_{EMG}(t)$ with a predefined threshold M_d .

Entropy of RD-LLGMN's output is also calculated to prevent risk of misclassification. The entropy is defined as

$$H(t) = - \sum_{c=1}^C {}^{(4)}O_c(t) \log^{(4)} O_c(t). \quad (8)$$

If the entropy $H(t)$ is less than a threshold H_d , the specific motion with the largest probability is determined according to the Bayes' decision rule. If not, the determination is suspended.

4 Experiments

Phoneme classification based on EMG signals was conducted to examine performance of the proposed method. In the experiments, EMG signals measured from mimetic and cervical muscles were used to classify six Japanese phonemes ($C = 6$: /a/, /i/, /u/, /e/, /o/, and /n/). Experiments were held with four subjects (A, B, and C: healthy; D: a patient with cervical spine injury).

4.1 Phoneme Classification Experiments

In this study, cross-talks between EMG signals [12] are used, and a bank of filters is applied to extract frequency features. The EMG signals are measured with monopolar leads, and cross-talk signals can be derived as the difference between potentials of every two electrodes. Three Ag/AgCl electrodes (SEB120, GE marquette Corp.) with conductive paste were attached to muscles (M. Depressor Labii Inferioris (DLI), M. Zygomaticus Major (ZM), and M. Masseer (MA)). EMG signals were recorded with a sampling frequency of 1 kHz. The cross-talk between DLI and ZM was used as input channel one, the cross-talk between DLI and MA as channel two, and the cross-talk between ZM and MA as channel three. Each channel was then fed into a bank of Z band-pass filters ($BPF_i, i = 0, \dots, Z - 1$). Bandwidths of these filters were set as follows:

$$BPF_i : 20 + \sigma i \text{ [Hz]} \sim 20 + \sigma(i + 1) \text{ [Hz]}, \quad (i = 0, 1, \dots, Z - 1). \quad (9)$$

Here, $\sigma = U/Z$, and U is the range of spectrum under consideration. In the phoneme classification experiments, U was set as 250 Hz, and the number of band-pass filters Z as 6. After the filter-bank stage, the number of EMG features $d = 18$, and these 18 channels of raw signals are rectified and filtered by a low pass filter (cut-off frequency: 1 Hz). The filtered signals are defined as $EMG_i(t)$ ($i = 1, \dots, d$), and normalized to make the sum of d channels equal to 1.0.

$$x_i(t) = \frac{EMG_i(t) - \overline{EMG}_i^{st}}{\sum_{i=1}^d EMG_i(t) - \overline{EMG}_i^{st}}, \quad (i = 1, \dots, d). \quad (10)$$

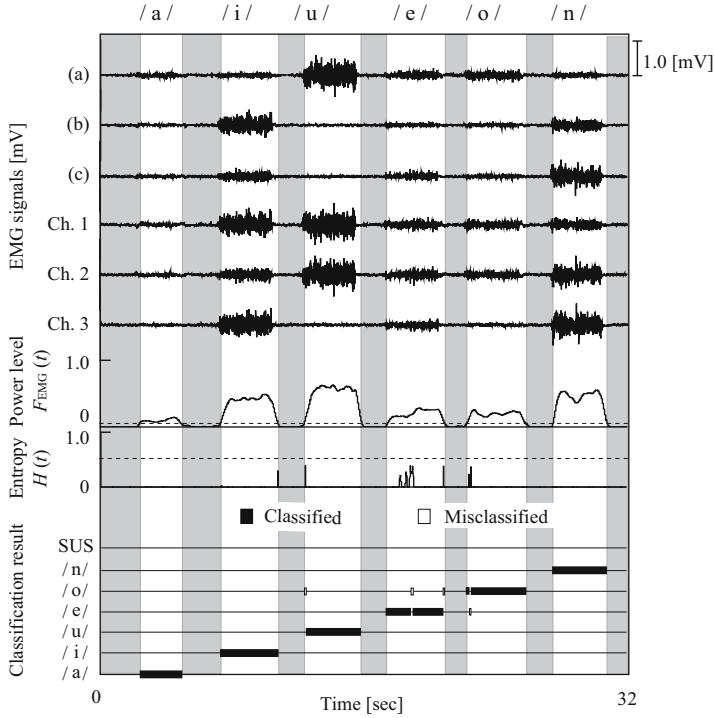


Fig. 3. Examples of the classification results using the proposed method (subject A). (a: *M. depressor labii inferioris*, b: *M. zygomaticus major*, c: *M. masseter*.)

where $\overline{EMG_i^{st}}$ is the mean value of $EMG_i(t)$, which is measured while relaxing the muscles.

Parameters of GMM in RD-LLGMN were set as: $C=6$, $K_c=1$ ($c=1, \dots, 6$). Dimensions of the reduced subspaces $M_{c,k}$, ($c=1, \dots, C$; $k=1$), were set as $M=9$. In the training phase, 50 EMG patterns were extracted from EMG signals of each phoneme, so that teacher signals consisted of $C \times 50$ patterns. The determination thresholds were set as $M_d=0.08$, and $H_d=0.5$.

An example of the classification results (subject A) is shown in Fig. 3. In this figure, three channels of raw EMG signals, three channels of cross-talk EMG signals, the power level $F_{EMG}(t)$, the entropy $H(t)$, and the classification results are plotted. The gray areas indicate that no utterance was conducted because the power level F_{EMG} was less than M_d . Although misclassification can be found in utterance of /u/, /e/ and /o/, the classification results of RD-LLGMN are relatively stable, and a high classification rate about 96.8% was realized in this experiment. Also, for each misclassified utterance, the corresponding entropy is high. It is believed that misclassification can be reduced using an appropriately modulated threshold H_d .

4.2 Comparison Experiments

Comparison experiments were conducted between the proposed method and traditional pattern classification methods based on PCA with neural classifier schemes. In the PCA part, original features are projected into a lower-dimensional space on directions, which correspond to the M highest eigenvalues of the covariance matrix. Feature vectors extracted with these M directions are then fed into neural classifiers. In this paper, a log-linearized Gaussian mixture network (LLGMN) [13] and a multilayer perceptron (MLP) [14], are used. LLGMN is a feedforward probabilistic NN based on GMM. The number of units in the input layer of LLGMN was set equal to M . Units in the hidden layer corresponded to the Gaussian components in GMM, which was set as one. The output layer had six units, and each unit outputs posterior probability for the input vector. On the other hand, the MLP had four layers. Number of units in the first layer was set as M , in both hidden layers unit numbers were $M \times \frac{10}{3}$ (Here, factor $\frac{10}{3}$ was chosen just for ease of setting the MLP), and in the output layer there is six units. Each output of MLP corresponds to one phoneme, and it was normalized to make the sum of all outputs equal 1.0, so that the normalized outputs can be regarded as posterior probabilities of phonemes. Same values of thresholds, M_d and H_d , were used for three classification method.

In the comparison experiments, classification rates of three methods are evaluated by varying the dimensionality of input EMG features d and an extraction rate (denoted as β), which stands for the ratio of M to d . The dimensionality of input EMG features d varies when we changing the number of filters Z from one to six. Five sets of randomly chosen initial weights were used to train each sample data. The EMG signals for six phonemes used for test were about 30 seconds. Figs. 4–6 depict mean values and standard deviations of the classification rates of subject A for different parameter combinations:

$$d \times \beta : \begin{cases} \beta \in [\frac{1}{3}, \frac{2}{3}, 1] & (d \in [3, 9, 15]) \\ \beta \in [\frac{1}{6}, \frac{2}{6}, \frac{3}{6}, \frac{4}{6}, \frac{5}{6}, 1] & (d \in [6, 12, 18]) \end{cases} \quad (11)$$

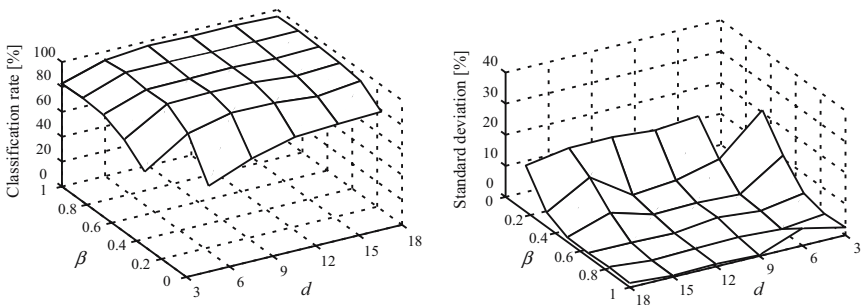


Fig. 4. Classification results using RD-LLGMN (Subject A)

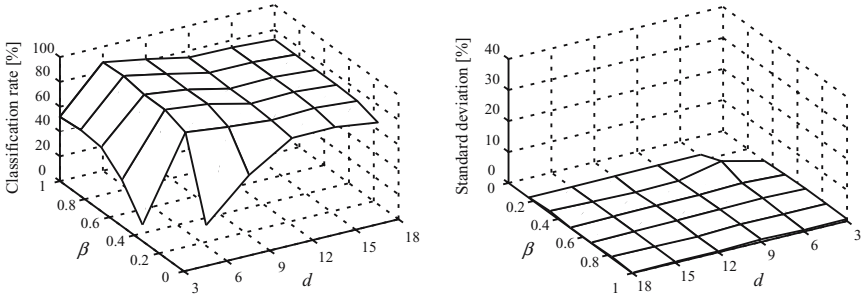


Fig. 5. Classification results using PCA with LLGMN (Subject A)

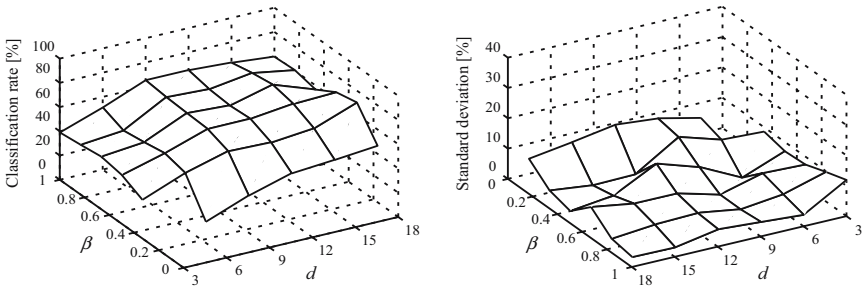


Fig. 6. Classification results using PCA with MLP (Subject A)

Please note that the directions of axes of d and β are reversed in the figures of standard deviations to make them shown clearly. In these figures, it can be found that RD-LLGMN achieved the best classification rates among all methods. Since PCA and the neural classifiers are optimized separately based on different training criteria, the features extracted may not always be consistent with the purpose of classification, and their classification performance was poorer than that of RD-LLGMN. Also, Since RD-LLGMN and LLGMN conduct classification according to probabilistic characteristics of input features, higher classification performance was achieved using RD-LLGMN and PCA with LLGMN than PCA with MLP. Furthermore, we can find that when β increases classification rates of all three method increase slightly. It is due to that more information is used for pattern classification. However, computation complexity and time used for training are significantly increased. On the other hand, when we increase d , similar trend can be observed for RD-LLGMN and PCA with MLP. In contrast, classification rates of PCA with LLGMN method decrease for about 15%. When increasing d , the entropy of LLGMN's output increases at the same time. Since the classification turned more ambiguous, more classification results were suspended, which resulted in decrease in the classification rates of PCA with LLGMN.

Finally, comparison experiments were conducted for four subjects. In these experiments, dimension of the input EMG features (d) was 18, and dimension of

Table 1. Comparison of classification rates for four subjects

Type of methods		PCA with MLP	PCA with LLGMN	RD-LLGMN
Subject A	<i>C.R.</i>	70.0	73.4	88.4
	<i>S.D.</i>	8.1	0.1	8.5
Subject B	<i>C.R.</i>	65.0	31.1	69.6
	<i>S.D.</i>	5.1	0.0	6.2
Subject C	<i>C.R.</i>	49.0	79.8	92.2
	<i>S.D.</i>	4.5	0.8	3.4
Subject D	<i>C.R.</i>	44.3	76.6	90.3
	<i>S.D.</i>	2.8	6.5	5.4

C.R. : Classification rate [%], *S.D.* : Standard deviation [%]

extracted feature set (M) was set as nine. Five sets of randomly chosen initial weights were used to training of three methods. A summary of classification rates for four subjects using three methods is shown in Table 1.

5 Conclusion

This paper proposes a novel classification method for multivariate EMG patterns. This method uses a probabilistic NN, RD-LLGMN, for high-dimensional pattern classification. With RD-LLGMN, discriminative information benefiting classification can be extracted, and an efficient classification of motions is possible.

To examine the classification capability and the accuracy of the proposed method, phoneme classification experiments have been carried out with five subjects. A bank of filters is applied to acquire frequency features from cross-talk EMG signals, and then the feature vectors are input into RD-LLGMN. In the experiments, the proposed method achieved high classification performance. Furthermore, comparison experiments were carried out between the proposed method and two PCA-based traditional methods, and the proposed method outperforms the other methods.

In Fig. 4, it can be found that standard deviations of the proposed method are larger than that of PCA with LLGMN when β is small. A detailed investigation is worthy to study stability of classification results of the proposed method. Also, in the future research, we would like to improve the pre-processing method for EMG signals, such as modulation of the parameters of filter bank and the low-pass filtering for each raw cross-talk EMG signal. Better classification performance would be available using combination of new pre-processing and RD-LLGMN.

References

1. Bu, N., Tsuji, T.: Multivariate Pattern Classification based on Local Discriminant Component Analysis. Proc. of IEEE International Conference on Robotics and Biomimetics (2004) Paper-ID: 290

2. Fukuda, O., Tsuji, T., Kaneko, M., Ohtsuka, A.: A Human-Assisting Manipulator Teleoperated by EMG Signals and Arm Motions. *IEEE Trans. on Robotics and Automation* **19** (2003) 210–222
3. Englehart, K., Hudgins, B., Chen, A.D.C.: Continuous Multifunction Myoelectric Control using Pattern Recognition. *Technology and Disability* **15** (2003) 95–103
4. Du, Y.C., Hu, W.C., Shyu, L.Y.: The Effect of Data Reduction by Independent Component Analysis and Principal Component Analysis in Hand Motion Identification. *Proc. of the 26th Annual International Conference of the Engineering in Medicine and Biology Society* **1** (2004) 84–86
5. Verikas, A., Bacauskiene, M.: Feature Slection with Neural Networks. *Pattern Recognition Letters* **23** (2002) 1323–1335
6. Bishop, C.: *Neural Networks for Pattern Recognition*. New York: Oxford University Press, 1995.
7. Jain, A.K., Duin, R.P.W., Mao, J.C.: Statistical Pattern Recognition: A Review. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **22** (2000) 4–37
8. Englehart, K., Hudgins, B., Parker, P.A., Stevenson, M.: Classification of the Myoelectric Signal using Time-frequency based Representations. *Medical Engineering & Physics* **21** (1999) 431–438
9. Lotlikar, R., Kothari, R.: Bayes-Optimality Motivated Linear and Multilayered Perceptron-Based Dimensionality Reduction. *IEEE Trans. on Neural Networks* **11** (2000) 452–463
10. Tsymbal, A., Puuronen, S., Pechenizkiy, M., Baumgarten, M., Patterson, D.: Eigenvector-based Feature Extraction for Classification. *Proc. of International FLAIRS Conference on Artificial Intelligence* (2002) 354–358
11. Juang, B-H., Katagiri, S.: Discriminative Learning for Minimum Error Classification. *IEEE Trans. on Signal Processing* **40** (1992) 3043–3054
12. Ohga, M., Takeda, M., Matsuba, A., Koike, A., Tsuji, T.: Development of A Five-finger Prosthetic Hand using Ultrasonic Motors Controlled by Two EMG Signals. *Journal of Robotics and Mechatronics* **14** (2002) 565–572
13. Tsuji, T., Fukuda, O., Ichinobe, H., Kaneko, M.: A Log-Linearized Gassian Mixture Network and Its Application to EEG Pattern Classification. *IEEE Trans. on Systems, Man, and Cybernetics-Part C: Application and Reviews* **29** (1999) 60–72
14. Rumelhart, D.E., McClell, J.L., Williams, R.J.: Learning Internal Representations by Error Propagation. *Parallel Distributed Processing* **I** (1986) 318–362