

# A Novel Channel Selection Method Based on Partial KL Information Measure for EMG-based Motion Classification

T. Shibanoki<sup>1</sup>, K. Shima<sup>1</sup>, T. Tsuji<sup>1</sup>, A. Otsuka<sup>2</sup> and T. Chin<sup>3</sup>

<sup>1</sup>Graduate School of Engineering, Hiroshima University, Higashi-Hiroshima, Japan

<sup>2</sup>Faculty of Health and Welfare, Prefectural University of Hiroshima, Mihara, Japan

<sup>3</sup>Hyogo Rehabilitation Center Hospital, Kobe, Japan

**Abstract** — To control machines using electromyograms (EMGs), subjects' intentions have to be correctly estimated and classified. However, the accuracy of classification is greatly influenced by individual physical abilities and measuring positions, making it necessary to select optimal channel positions for each subject. This paper proposes a novel online channel selection method using probabilistic neural networks (PNNs). In this method, measured data are regarded as probability variables, and data dimensions are evaluated by a partial KL information measure that is newly defined as a metric of effective dimensions. In the experiments, channels were selected using this method, and EMGs measured from the forearm were classified. The results showed that the number of channels is reduced with  $33.33 \pm 11.8\%$ , and the average classification rate using the selected channels is almost the same as that using all channels. This demonstrates that the method is capable of selecting effective channels for classification.

**Keywords** — Kullback-Leibler information, partial Wilks' lambda, variable selection method, pattern classification, electromyogram

## I. INTRODUCTION

Since electromyograms (EMGs) reflect human intentions and motions, these signals can be used to control various machines estimating these intentions and motions. Recently, neural networks (NNs) to classify EMG patterns have been widely studied, as such networks can represent any nonlinear mapping by learning [1], [2]. In addition, Tsuji *et al.* proposed the Log-Linearized Gaussian Mixture Network (LLGMN) and confirmed its effectiveness in biological signal classification [3].

The number of channels and the measuring positions for EMGs are very important because the accuracy of EMG classification is greatly influenced by the amount of information in measured signals. In the case of EMG classification, effective channel selection methods have been proposed [4], [5] by evaluating all combinations of EMG electrodes. With a very large number of channels, however, computation takes a long time, meaning that real-time applications for human-machine interfaces cannot be achieved. To select effective channels for each subject, it is necessary to evaluate users'

EMG patterns and the impact of each channel on the EMG classification with less computational time.

This paper proposes a novel variable selection method based on Kullback-Leibler (KL) information and its application to channel selection for EMG measurement. In this method, measured signals are regarded as probability variables, and their probabilistic density function (pdf) is estimated using probabilistic neural network (PNN) learning based on Kullback-Leibler (KL) information. In addition, a partial KL information measure based on a partial Wilks' lambda is newly defined as a metric of effective dimensions for classification, and dimensions that are not effective are eliminated. As a result, effective channels for each user can be selected using this method without evaluating all channel combinations.

## II. PARTIAL KL INFORMATION MEASURE

A novel metric of dimensions for pattern classification, called the partial KL information measure, is newly defined. Using the partial KL information measure, the most ineffective dimensions can be eliminated.

### A. Partial Wilks' lambda [6]

Wilks' lambda is a test statistic used in multivariate analysis of variance. There are  $N_k$  samples in the  $k$ th class, and each sample  $\mathbf{x}$  is a vector with  $L$  dimensions ( $\mathbf{x} \in \mathfrak{R}^L$ ).  $\mathbf{x}_n^{(k)}$  ( $n = 1, 2, \dots, N_k$ ) denotes the vector of the  $n$ th sample in the  $k$ th class. Wilks' lambda is defined as follows:

$$\Lambda = |W| / |T|, \quad (1)$$

where  $|\cdot|$  denotes the determinant of the matrix,  $W$  is the within-class sum of squares,

$$W = \sum_{k=1}^K \sum_{n=1}^{N_k} (\mathbf{x}_n^{(k)} - \bar{\mathbf{x}}^{(k)})(\mathbf{x}_n^{(k)} - \bar{\mathbf{x}}^{(k)})^T, \quad (2)$$

and  $T$  is the total sum of squares,

$$T = \sum_{k=1}^K \sum_{n=1}^{N_k} (\mathbf{x}_n^{(k)} - \bar{\mathbf{x}})(\mathbf{x}_n^{(k)} - \bar{\mathbf{x}})^T, \quad (3)$$

where  $\bar{\mathbf{x}}^{(k)}$  is the centroid of class  $k$ , and  $\bar{\mathbf{x}}$  is the overall centroid.  $K$  is the number of classes, and  $(\cdot)^T$  denotes the transpose of the matrix. The smaller the Wilks' lambda, the more classes are separated.

To identify important dimensions for discrimination, a partial Wilks' lambda is then introduced:

$$A(\mathbf{x}_i^* | \mathbf{x}_{[i]}^*) = A(\mathbf{x}_{[i]}^*, \mathbf{x}_i^*) / A(\mathbf{x}_{[i]}^*), \quad (4)$$

where  $\mathbf{x}_i^*$  is the data set of  $i$ th element of  $\mathbf{x}$ , and  $\mathbf{x}_{[i]}^*$  is the set of vectors that are eliminated in the  $i$ th dimension from  $\mathbf{x}$ . Here,

$$F_i = \frac{N-K-(L-1)}{K-1} \cdot \frac{1-A(\mathbf{x}_i^* | \mathbf{x}_{[i]}^*)}{A(\mathbf{x}_i^* | \mathbf{x}_{[i]}^*)} \quad (5)$$

performs the same role as the  $F$ -statistic in one-way analysis of variance, where  $N$  is the number of samples in all classes ( $N = \sum_{k=1}^K N_k$ ). This follows an  $F$  distribution with  $K-1$  and  $N-K-(L-1)$  degrees of freedom [7]. In the backward elimination method, the smallest of all the  $F_i$  values is eliminated if it is lower than the critical value of  $F_{out}$ . This process is repeated on the remaining dimensions until the smallest  $F_i$  value is higher than  $F_{out}$ .

However, the partial Wilks' lambda cannot respond to non-Gaussian data since it is based only on data variance. Therefore, it might be difficult to correctly classify the EMG, which is not always Gaussian, for human-machine interfaces after evaluating the dimensions using the Wilks' lambda. To solve these problems, a new metric of effective dimensions referred to as partial KL information is proposed.

### B. Partial Kullback-Leibler (KL) information

To classify non-Gaussian and/or non-linear data, it is common to regard them as probability variables and estimate their pdf. It is considered possible to evaluate the dimensions using the estimated distribution.

The KL information represents the difference between two probability distributions [8]. Let  $\mathbf{x}_n^{(k)}$  be the vector of the  $n$ th sample in the  $k$ th class, and the probabilities of the  $k$ th class in true probability distribution and estimated distribution are  $\Pr\{\mathbf{x} = \mathbf{x}^{(k)}\} = P_k$  and  $Q_k$ , respectively. The KL information is calculated by

$$I(\mathbf{P}, \mathbf{Q}) = \sum_{k=1}^K P_k \log \frac{P_k}{Q_k}$$

$$= \sum_{k=1}^K P_k \log P_k - \sum_{k=1}^K P_k \log Q_k, \quad (6)$$

where  $\mathbf{P} = [P_1, P_2, \dots, P_K]^T$  and  $\mathbf{Q} = [Q_1, Q_2, \dots, Q_K]^T$ , respectively, and the first term on the right side is reduced to a constant; therefore, if the second term becomes the minimum, the KL information becomes the minimum. This means that the estimated distribution is close to the true distribution.

Based on the above concept, this paper defines the partial KL information to evaluate dimensions:

$$I_i = \sum_{n=1}^N I_n(\mathbf{P}, \mathbf{Q}) / \sum_{n=1}^N I_n(\mathbf{P}_{[i]}, \mathbf{Q}_{[i]}), \quad (7)$$

$I_n(\mathbf{P}_{[i]}, \mathbf{Q}_{[i]})$  is the KL information for the  $n$ th sample  $\mathbf{x}_n$ , and is obtained using the probability variable  $\mathbf{x}_{[i]}^*$  in which the  $i$ th element of each variable included is eliminated. If  $I_n(\mathbf{P}, \mathbf{Q}) \equiv I_n(\mathbf{P}_{[i]}, \mathbf{Q}_{[i]})$ , this means that true distribution can be estimated without the dimension  $x_i$ .

## III. PARTIAL KL INFORMATION-BASED VARIABLE SELECTION USING PNN

A novel variable selection method using the partial KL information measure is newly proposed for pattern classification. In this method, KL information is calculated through PNN learning, and redundant dimensions are eliminated one by one.

### A. Variable selection using the LLGMN

The LLGMN is based on the GMM and the log-linear model of pdf [3]. By applying the log-linear model to a product of a mixture coefficient and the mixture component of the GMM, the semiparametric pdf model is incorporated into a three-layer feedforward NN. This network can calculate the *a posteriori* probability  $Y_k(n)$  for the input pattern  $\mathbf{x}(n)$  after learning [3].

A simple backpropagation algorithm is feasible when the teacher vector  $\mathbf{T}(n)$  for the  $n$ th input vector  $\mathbf{x}(n)$  is given. As the energy function  $J$  for the network,

$$J = - \sum_{n=1}^N \sum_{k=1}^K T_k(n) \log Y_k(n) \quad (8)$$

is used, and learning is performed to minimize the KL information for true distribution. This means that a learned network can approximate the data distribution.

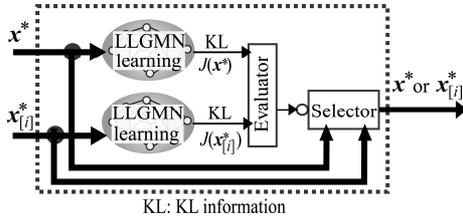


Fig. 1. Variable selector

The variable selection procedure for two input vector sets  $\mathbf{x}^*$  and  $\mathbf{x}_{[i]}^*$  is shown in Fig. 1. First, the KL information  $J(\mathbf{x}^*)$  and  $J(\mathbf{x}_{[i]}^*)$  of two input vector sets are calculated through LLGMN learning. Next, each set of KL information is input to the evaluator. The evaluator outputs the partial KL information calculated based on Eq. (7), where

$$E_i = J(\mathbf{x}^*) / J(\mathbf{x}_{[i]}^*). \quad (9)$$

Based on  $E_i$ , two inputs are evaluated, and the more effective one is selected through the selector. The  $i$ th dimension for which  $E_i$  ( $i = 1, 2, \dots, L$ ) has the largest value is eliminated by the selector, and ineffective dimensions are then eliminated one by one through repetition of this procedure.

### B. Channel selection for EMG classification

Using the partial KL information-based variable selection method, we applied the technique to channel selection for EMG classification. The structure of the channel selection method is shown in Fig. 2.

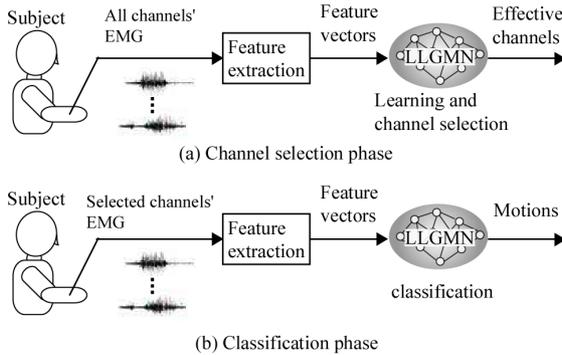


Fig. 2. Structure of the proposed channel selection method

- Feature extraction

First, EMG signals measured from  $L$  pairs of electrodes are digitized by an A/D converter (sampling frequency: 1 [kHz]), and are rectified and filtered out through a second-order Butterworth filter (cut-off frequency: 1 [Hz]). These sampled signals are defined as  $EMG_l(n)$  ( $l = 1, 2, \dots, L$ ). Next, the

$EMG_l(n)$  values are normalized to make the sum of  $L$  channels equal to 1:

$$x_l(n) = \frac{EMG_l(n) - \overline{EMG}_l^{st}}{\sum_{l'=1}^L EMG_{l'}(n) - \overline{EMG}_{l'}^{st}} \quad (l = 1, 2, \dots, L), \quad (10)$$

where  $\overline{EMG}_l^{st}$  is the mean value of  $EMG_l(n)$  while relaxing the muscles. This feature vector  $\mathbf{x}(n) \in \mathfrak{R}^L$  is input to the LLGMN for channel selection.

- Channel selection

For EMG classification, the effective dimension of the feature vector (i.e., the EMG measurement position) is selected through LLGMN learning. In the channel selection algorithm, ineffective channels are eliminated based on the partial KL information when the termination condition is fulfilled. The algorithm details are as follows:

1. EMG signals are measured from  $L$  pairs of electrodes of  $K$  motions for LLGMN learning, and the feature vector  $\mathbf{x}(n)$  is calculated using Eq. (10).
2. The data sets  $\mathbf{x}^*$ ,  $\mathbf{x}_{[1]}^*$ , ...,  $\mathbf{x}_{[L]}^*$  are prepared, and each set is learned by the individual LLGMN.
3.  $J(\mathbf{x}^*)$  and  $J(\mathbf{x}_{[i]}^*)$  are calculated using Eq. (8), and the partial KL information measure  $E_i$  ( $i = 1, 2, \dots, L$ ) is obtained from Eq. (9).
4. Using each obtained value of  $E_i$ , the most ineffective channel is selected. Since the learned network with less partial KL information can approximate the true distribution, therefore, dimension  $j$  (for which the partial KL information  $E_j$  is the largest) is eliminated. Here, if  $E_j \leq E_\theta$  (where  $E_\theta$  is the critical value for the termination condition), the dimension  $x_j$  is not eliminated, and the elimination process ends. Meanwhile, if  $E_j > E_\theta$ , the dimension  $x_j$  is eliminated. Then, let  $\mathbf{x}^* = \mathbf{x}_{[j]}^*$  and  $L = L - 1$ , and return to Step 2.

## IV. CHANNEL SELECTION EXPERIMENTS

To verify the effectiveness of the proposed channel selection method, we performed EMG-based motion classification experiments.

### A. Experimental conditions

The subjects were five normal volunteers (including one person who had experienced EMG control), and six electrodes were attached to each subject's right forearm (see Fig. 3). The subjects performed four motions ( $K = 4$ ; hand grasping, hand opening, wrist flexion and extension) for a few seconds, and

repeated each motion 10 times. The sampling frequency was 1 [kHz]. One set of the measured data was used for learning and channel selection (learning data), and the other sets were used only for classification (evaluation data). The limit value  $E_{\theta} = 0.8$  was used for channel selection, the number of learning data was  $N = 80$ , and the number of evaluation data was 4,000 (20 and 1,000 for each motion respectively).

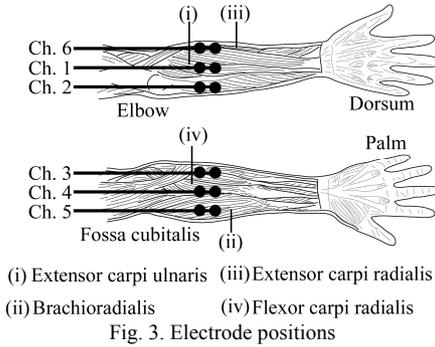


Fig. 3. Electrode positions

**B. Results and discussion**

Figure 4 shows the relationship between classification rates and KL information when the number of channels for classification is changed. The figure outlines the results of reducing the number of channels using learning data measured from Subject A. In this experiment, the initial weights of the LLGMN were changed 10 times, learning deviations were and classification were performed in each trial, and the average classification rates and standard obtained. Since the KL information from these results was almost near zero with three to six channels, it is confirmed that the accuracy of classification was high. At this point, three channels were selected by the termination condition of the proposed method (Ch. 2, Ch. 3 and Ch. 6), and the average classification rate was 100%.

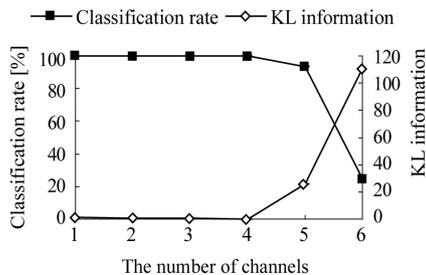


Fig. 4. Relationships between classification rates and KL information (Subject A). Note that the standard deviations are less than 0.27% for the classification rate and 0.06% for the KL information.

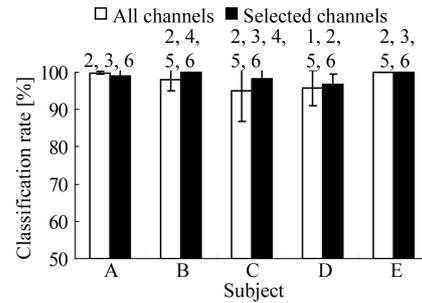


Fig. 5. Classification rates using all channels and selected channels

Next, the average classification rates of the evaluation data using the channels selected with the learning data are shown in Fig. 5. This figure shows the classification results of using all channels' and selected channels' EMG for all subjects. It can be seen that motion could be classified using only the selected channels. The number of channels is reduced with  $33.3 \pm 11.8\%$ . The classification rates of Subject A, which were calculated using three channels of all combinations from six channels' EMG (i.e. the number of combinations was  $6C3 = 20$ ) are shown in Table 1. In this table, the gray zone highlights the outcomes using the proposed method. These results show that the second-highest classification rate was obtained by the proposed method, meaning that semi-optimal channel combination could be selected from all combinations. In the experiments on the other subjects, optimal or semi-optimal channel combinations could also be selected with the highest or second-highest classification rate. These results lead us to the conclusion that effective channel could be selected using the partial KL information measure without evaluating all possible channel combinations.

Table 1. Classification rates using three channels of combinations from all channels (Subject A)

Selected channels	2, 5, 6	2, 3, 6	2, 3, 4	2, 4, 5	1, 2, 3
Classification rate [%]	99.16	98.99	97.52	96.96	95.58
Standard deviation	1.75	2.19	4.4	5.27	7.91
Selected channels	1, 3, 5	1, 4, 5	1, 5, 6	1, 2, 4	3, 4, 5
Classification rate [%]	94.97	94.76	94.13	94.06	92.34
Standard deviation	5.16	6.08	5.64	8.6	7.8
Selected channels	1, 2, 5	2, 3, 5	1, 3, 4	1, 3, 6	3, 5, 6
Classification rate [%]	92.24	91.47	90.83	90.78	90.09
Standard deviation	8.96	5.24	8.37	8.08	7.86
Selected channels	1, 4, 6	4, 5, 6	1, 2, 6	3, 4, 6	2, 4, 6
Classification rate [%]	84.14	81.97	80.91	77.56	75.94
Standard deviation	13.31	19.48	12.5	16.93	17.55

**V. CONCLUSION**

This paper proposes a variable selection method using a new metric to select effective dimensions called the partial

KL information measure. In order to confirm its effectiveness, the method was applied to channel selection for EMG classification. In experiments on forearm motion classification, effective channels for classification could be selected for each subject. In future research, we plan to enhance the channel selection method by considering cross-talk EMG and applying the method to training systems for human-machine interfaces using EMG.

#### ACKNOWLEDGMENT

We would like to cordially acknowledge and express our appreciation to Mr. M. Shigeto for his assistance in the implementation of this research. This work is partially supported by the 21st Century COE Program of JSPS (the Japan Society for the Promotion of Science) on Hyper Human Technology toward the 21<sup>st</sup> Century Industrial Revolution.

#### REFERENCES

1. Hiraiwa, A., Shimohara, K. et al. (1989) EMG pattern analysis and classification by neural network, 1989, Proceedings of IEEE International Conference on Syst., Man and Cybern., pp. 1,113-1,115
2. Kelly, M. F., Parker, P. A. et al. (1990) The Application of Neural Networks to Myoelectric Signal Analysis: A preliminary study, IEEE Trans., Biomedical Eng, Vol. 37, No. 3, 1990, pp. 221-230
3. Tsuji, T., Ichinobe, H. et al., (1995) A Maximum Likelihood Neural Network Based on a Log-Linearized Gaussian Mixture Model, Proc. of IEEE International Conf. Neural Networks, 1995, pp. 2,479-2,484
4. Otsoshi, T., Ushiba, J. et al. (2005) Identification of Forearm Movements using Analysis of Electromyographic Signals, IEICE technical report. ME and bio cybernetics, Vol. 104, No. 756, pp. 13-16, in Japanese
5. Nagata, K., Ando, K. et al. (2006) Recognition of Hand Motions based on Multi-channel SEMG Employing the Monte Carlo Method for Channel Selection, BME, Vol. 44, No. 1, pp. 138-147, in Japanese
6. Nath, R., Pavur, R. (1985) A new statistic in the one-way multivariate analysis of variance, Computational Statistics & Data Analysis 2, 1985, pp. 297-315
7. A. C. Rencher. (1995) Methods of Multivariate Analysis, John Wiley & Sons.
8. Kullback, S., Leibler, R. A., (1951) On information and sufficiency, Annals of Mathematical Statistics, Vol. 22, pp. 79-86

Author: T. Shibanoki  
 Institute: Hiroshima University  
 Street: Kagamiyama 1-4-1  
 City: Higashi-Hiroshima  
 Country: Japan  
 Email: shibanoki@bsys.hiroshima-u.ac.jp