A Comparison of the SOFM with LVQ, SOFM without LVQ and Statistical Technique

Anjana Bhardwaj, Manish, A. K. Arora

Abstract – The shapes and firing rates of MUAP's (motor unit action potentials) in an EMG (electromyographic) signal provide an important source of information for the diagnosis of neuromuscular disorders. In order to extract this information from EMG signals recorded at low to moderate force levels, it is required: i) to identify the MUAP's composing the EMG signal, ii) to classify MUAP's with similar shape. For the classification of MUAP's two different pattern recognition techniques are presented: i) An artificial neural network (ANN) technique based on unsupervised learning, using a modified version of the self-organizing feature maps (SOFM) algorithm and learning vector quantization (LVQ), and ii) A statistical pattern recognition technique based on Euclidean distance. A total of 521 MUAP's obtained from 2 normal subjects, 4 subjects suffering from myopathy, and 5 subjects suffering from motor neuron disease were analyzed. The success rate for the ANN technique was 97.6%, the success rate for SOFM technique was 94.8%, and for statistical technique it was 95.3%. So SOFM technique along with LVQ is batter technology than the SOFM without LVQ technique and Statistical technique.

Index Terms – Artificial Neural Network, Electromyography, learning vector quantization, Motor unit Action Potentials, Self-organizing feature maps.

I. INTRODUCTION

There are more than 100 neuromuscular disorders that affect the brain and spinal cord, nerves, or muscles. Many of these diseases are hereditary and life expectancy of many sufferers is considerably reduced. Early detection and diagnosis of these diseases by clinical examination and Laboratory tests are essential for their management as well as their prevention. EMG (Electromyography) examination studies the electrical activity of the muscle and forms a valuable neurophysiologic test for the assessment of neuromuscular disorders.

EMG signals recorded at low to moderate force levels are composed of motor unit action potentials (MUAP's) generated by different motor units. The motor unit is the smallest functional unit of the muscle that can be voluntarily activated. The MUAP shape reflects the structural organization of the motor unit. MUAP classification into groups of similar shapes provides important information for the assessment of neuromuscular pathology.

Ms. Anjana Bhardwaj, Electronics & Communication Department, ABES Engineering College, Ghaziabad (U.P.) India,
Mr. Manish, Electronics & Communication Department, ABES Engineering College, Ghaziabad (U.P.) India,
Prof. A.K. Arora, Electronics & Communication Department, ABES Engineering College, Ghaziabad (U.P.) India.

In previous work the classification was done by only the SOFM technique, and statistical technique. But SOFM along with LVQ technique is a new approach to get m ore accurate classification of MUAPs. Recent advances in computer technology have made automated EMG analysis feasible. Although a number of computer-based quantitative EMG analysis algorithms have been developed, some of them are commercially available. Most importantly, there is no uniform international criterion neither for pattern recognition of similar MUAP's nor for MUAP features extraction [1], [2]. A brief survey of quantitative EMG studies carried out during the last two decades follows. LeFever and DeLuca [3][4] used a special three-channel recording electrode and a hybrid visual-computer decomposition scheme based on template matching and firing statistics for MUAP identification. The more recently in their system called multiple motor unit potentials (multi- MUP), they used different shape parameters as input to a template matching technique [5]. Guiheneuc *et al.* [6], classified MUAP's at low levels of voluntary contraction through comparison of shape parameters. In this work, statistical technique, ANN pattern recognition techniques (SOFM along with LVQ) were developed to classify MUAP's. ANN pattern recognition techniques is a modified version of the Kohonen self organizing feature maps (SOFM) algorithm in conjunction with learning vector quantization (LVQ) [11]. The additional use of the LVQ aims to improve the classification performance by slight adaptation of the classification boundaries. The proposed techniques were successfully applied in the classification of EMG signals recorded from normal (NOR) subjects and subjects suffering from motor neuron disease (MND) and myopathy (MYO). The next section will contain all the three techniques and the comparison of the result of all the techniques, which will show that the ANN pattern recognition technique is better than the other two, and its success rate is also higher than the others.

II. METHOD OF STATISTICAL TECHNIQUE

In this procedure the Euclidean distance is used in order to identify the similar waveforms using a constant threshold.

The implementation steps are as follow:

Step 1: Start with the first waveform x as input, being the first member of the class.

Step 2: Calculate the vector length of the input waveform x and the distance between x and all the other segmented waveforms y as

$$l_k = \sum_{i=1}^{N} x^2_{i} \quad \text{where } N{=}120 \qquad (1)$$

and

$$d_{xy} = \sum_{i=1}^{N} (x_i - y_i)^2. \qquad (2)$$

Step 3: Find the waveform y with the minimum distance which is the one with the greatest similarity with x and remove it from the input data set.

Step 4: Sliding and baseline correction. First slide the waveform y with minimum distance up to two points backward and up to two points forward in order to find the best alignment position. Recalculate the distance d_{xy} for each case and assign the smallest as d_{min}. Then, using the beginning and the ending parts of the MUAP waveforms, calculate baseline correction bc as

$$bc = \left(\sum_{i=1}^{10} y_i + \sum_{i=N-9}^{N} y_i - \sum_{i=1}^{10} x_i - \sum_{i=N-9}^{N} x_i \right) \Big/ 20 \qquad (3)$$

Subtract bc from waveform y and recalculate distance d_{xy} with x. If it is smaller than d_{min}, assign it as the new d_{min}.

Step 5:
If $d_{min}/l_x < 0.125$,
then   add to class,
         calculate class average
         and assign class average to input x;
Else, if number of class members >2,
then   form a MUAP class;
Else,   class members are superimposed
         assign waveform to input ;
go to Step 2.

If the minimum distance divided by the vector length of the first waveform is less than a constant threshold, set to 0.125, then the two waveforms form a class. Then the class average is calculated and the procedure is repeated (go to Step 2 with the class average as input) comparing the class average now with all the rest waveforms in order to find the next waveform with the minimum distance. If the condition above is satisfied, then a new waveform is added to the class and a new class average is calculated, and so on. If not, the process stops; if the class members are more than or equal to three, then a MUAP class is formed and its averaged waveform is saved. If they are less than three, they are considered as superimposed waveforms. The process continues where it stopped comparing the last encountered waveform with all the remaining ones until all waveforms are processed. The baseline correction was applied selectively only to the waveform with the greatest similarity to the reference waveform and it was applied only if the distance between and with baseline correction was smaller than the distance without baseline correction. The use of baseline and slide correction improved the performance of the statistical pattern recognition technique by 5% as documented. The value of 0.125 used in Step 5 was also used by Andreassen [7]. This threshold is critical because a smaller value may split a MUAP class with high waveform variability in two or more subclasses, whereas a greater threshold value may merge resembling MUAP classes. Fig.5 illustrates how the segmented signal waveforms of Fig. 2.

III. METHOD OF ANN PATTERN REORGANIZATION TECHNIQUE

The ANN pattern system consists of the following modules: Data acquisition and Preprocessing, Segmentation, and Classification. Figure 1 illustrates the system flowchart. And the next section will contain the procedure for the classification of EMG signal.
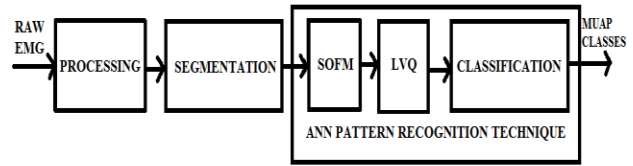

Figure 1 Flow chart of classification

3.1 Data acquisition and preprocessing

The EMG signal was recorded from the muscle at low to moderate force levels up to 30% of maximum voluntary contraction (MVC) under isometric conditions. The signal was acquired for 5 s, using the concentric needle electrode. The signal was analogue band-pass filtered at 3–10 kHz, and sampled at 20 kHz with 12-b resolution. The EMG signal was then low-pass filtered at 8 kHz.

1.2 Segmentation

The next step is to cut the EMG signal into segments of possible MUAP waveforms and eliminate areas of low activity. The segmentation algorithm calculates a threshold depending on the maximum value $\max_i \{x_i\}$ and the mean absolute value $(1/L) \sum_{i=1}^{L} (x_i)$ of the whole EMG signal. Peaks over the calculated threshold 'T' are considered as candidate MUAP's.

The threshold T is calculated as follows:
If $\max_i \{ x_i \} > \frac{30}{L} \sum_{i=1}^{L} (x_i)$,

$$\text{then } T = \frac{5}{L} \sum_{i=1}^{L} (x_i),$$
$$\text{else } T = \max_i \{ x_i \}/5.$$

Where x_i = discrete input values
and L = number of samples in the 5 s EMG signal

The threshold T is allowed between 30 and 100 µV. It is noted that there is no standardized procedure to estimate the threshold level: Dorfman and McGill [2] mentioned an amplitude threshold of 20–50 V, Stalberg *et al.* [5] used a 30-V threshold, whereas Andreassen[7] used an amplitude threshold in excess of 50 V.

1.3 Classification

The segmented EMG signals are processed in order to identify groups of similar MUAP's. In this work a method for MUAP classification is presented: a neural-network-based pattern recognition technique.

Artificial Neural-Network Pattern Recognition Technique

A single layer neural network is used for the identification and grouping of similar MUAP's and separation of superimposed waveforms. The ANN architecture the classification procedure is implemented in three phases.

In the first phase SOFM (self-organizing feature map) (Kohonen) [11] is used. In the second phase, in order to improve classification performance, the learning vector quantization method (LVQ2 by Kohonen) [11], is applied. In the third phase, the actual classification takes place. It should be noted that in each learning phase the input is presented to the network only once, for only one learning epoch. This makes the algorithm fast and suitable for real-time applications.

1.3.1 Self- organizing feature map (SOFM)- Learning phase 1

The objective of this phase is to provide a first "approximate" quantization of the input space by adapting the weight vectors of the neurons in the feature map [9]–[11], [12].
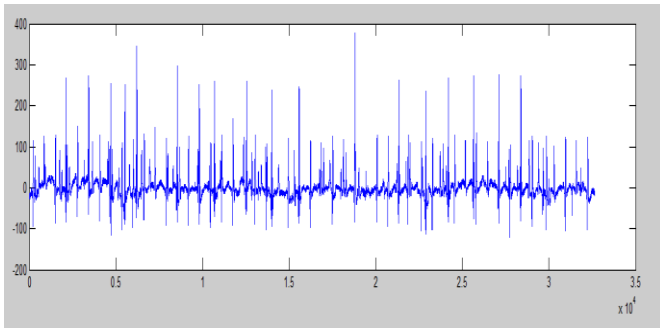
Figure 2 raw EMG signal

A problem with SOFM when the weights are initialized at small random values is that such initialization may give different results at different runs. This is undesired when trying to evaluate and optimize the performance of the algorithm or when the physician wants to review the classification results. In order to avoid this problem, the weights of the output nodes should not initialized at small random values but at 0.0001.
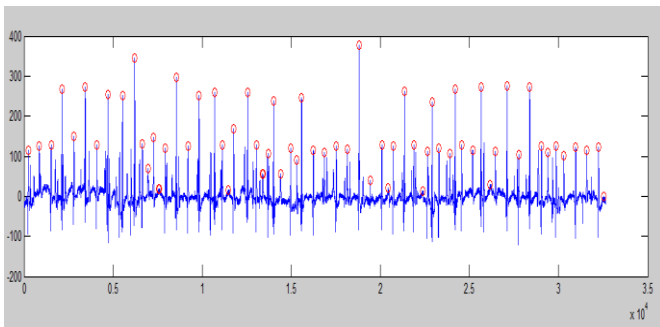
Figure 3 Classified MUAP's of the EMG Signal

The implementation steps are as follow:

Step 1: Initialize weights at 0.0001.

Step 2: Calculate distances between the input vector x_i and weight vectors for each output node k.

$$d_k = \sum_{i=1}^{N}(x_i - w_{ik})^2 \text{ Where } k = 1, 2,\text{----}8 \text{ and}$$
$$N=120 \tag{4}$$

The output node with minimum distance is the winner.

Step 3: Adapt the weights. The weights for each output node k and for each i are adapted with

$$w_{ik}(t + 1) = w_{ik}(t) + h_k (x_i - w_{ik}(t)). \tag{5}$$

Where $h_k=$ learning rate and it is a Gaussian function.

It can be given as:

$$h_k = g \exp ((-(k - k_w)^2 t/2)\sqrt{(t_{kw})} \tag{6}$$

Where,
Value of g can be $0< g \leq 1$,
k_w is the winner node,
t is the number of iterations,
t_{kw} is the number of times the specific node is selected as the winner.

For the initialization $g = 1$ and for the first winner $t_{kw} = 1$ and $k = k_w$, then $h_k = 1$.

If calculated $h_k < 0.005$, then the weights of the specific node are not adapted, since the change in the weights vector will be minimum. This is implemented in order to save computation time.

Step 4: Go to Step 2 and repeat for all segmented inputs. After all inputs are presented to the network, the first adaptation of the weights vector is completed and the system proceeds to the second learning phase.

1.3.2 Learning vector quantization (LVQ)- Learning phase 2

The task of this phase is to adapt the weights vectors slightly in order to improve the classification performance [9], [11]. LVQ demands knowledge of correctly classified inputs. It is assumed that the adaptation carried out during the first learning phase is correct and thus the segmented inputs will be correctly classified Weight adaptation and winner selection is again on-going as described in learning phase 1.

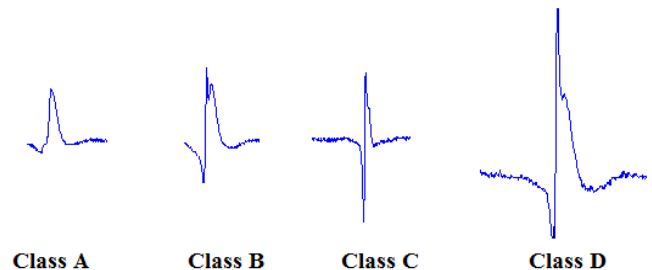Class A   Class B   Class C   Class D

Figure 4 Classes of MUAP's

In the modified version of LVQ2 the implementation steps are as follow:

Step 1: Use the values of the weight vectors as obtained from learning phase 1.

Step 2: Present input and calculate distances d_k between the input vector x_i and weight vectors w_{ik} for each output node as in equation (4). The output node with the minimum distance d_{k1} is the first winner $k1$ and the output node with minimum distance d_{k2} is the second winner $k2$.

Step 3: Adapt weights. The weights for the first winner output node $k1$ is adapted with

$$w_{ik1}(t + 1) = w_{ik1}(t) + h_{k1} (x_i - w_{ik1}(t)) \tag{7}$$

And for the second winner $k2$ with the weight

$$w_{ik2}(t+1) = w_{ik2}(t) - 0.1(d_{k1}/d_{k2})h_{k1}(x_i - w_{ik1}(t)) \quad (8)$$

The learning rate h_{k1} is initialized to 0.2 and decreases linearly with the number of times t_{kw1} the specific node $k1$ is selected as the first winner

$$h_{k1} = 0.2 - 0.01 t_{k1w} \quad (9)$$

Here

w_{ik1} is weight vector with the correct label (first winner),
w_{ik2} is the weight vector with the incorrect label (second winner),

The factor d_{k1}/d_{k2} is used to control the adaptation of the second winner.

Step 4: Go to Step 2 and repeat for all segmented inputs.
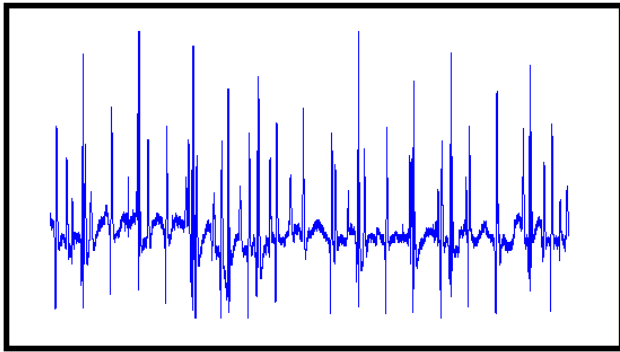After it the actual classification process starts.

Figure 5 Segmented EMG signal

3.3.3. *Classification phase*

In this phase all the input vectors are classified to one of the output nodes.

The implementation steps are the following:

Step 1: Calculate distances d_k between the input vector x_i and the weight vectors w_{ik} as in equation (4).

Step 2: The length l_{kw} of the weight vector of the winner node kw is calculated as the sum of the squares of its vector values

$$l_{kw} = \sum_{i=1}^{N} w^2_{ik} \quad (9)$$

If $d_{kw}/l_{kw} < 0.2$, then the input is assigned to the MUAP class of the winner node.

The physical meaning of d_{kw}/l_{kw} is that the greater its value, the greater the dissimilarity between the waveforms.

Step 3: Go to Step 2 and repeat for all segmented inputs.

Step 4: If the number of members in a class is three or more, then a valid MUAP class is identified. Figure 3 illustrates the classification of MUAP's of the raw EMG signal given in figure 2.

IV. RESULT

Table.1: shows Classification success rate of the three algorithms.

Subject (No.)	SOFM with LVQ	SOFM	STATISTICAL
NOR	150/152 (98.68%)	149/152 (98.02%)	151/152 (99.34%)
MYO	273/277 (98.55%)	264/277 (95.3%)	267/277 (96.38%)
MND	130/136 (95.58%)	126/136 (92.64%)	125/136 (91.91%)
TOT	553/565 (97.87%)	539/565 (95.39%)	543/565 (96.10%)

EMG data collected from 11 subjects were analyzed using the pattern recognition techniques described in Section II and III. Data were recorded from 2 normal (NOR) subjects, 4 subjects suffering from myopathy (MYO) and 5 subjects suffering from motor neuron disease (MND). Diagnostic criteria were based on clinical opinion, biochemical data and muscle biopsy. Only subjects with no history or signs of neuromuscular disorders were considered as normal. Table 1 shows the classification success rate on 565 MUAP's. The classification success rate was defined as the percentage ratio of the correctly identified MUAP classes by the algorithm and the number of true MUAP classes present in the signal as identified by an experienced neurophysiologist.

The average success rate for the SOFM with LVQ algorithm was 97.87%, for the SOFM algorithm alone 95.39%, and for the statistical pattern recognition algorithm 96.10%. The ANN technique also yielded good results without the LVQ learning phase. Examining the classification success rate for each class, the highest success rate was obtained for the NOR group and the lowest for the MND group. This was the case for all three algorithms. The lowest success rate for the MND group is attributed to the more complex and variable waveform shapes. Also, as shown in Table 1, the SOFM with LVQ algorithm improved significantly the success rate for the MND group compared to the other two algorithms. The statistical algorithm gave the highest success rate for the NOR group and the lowest for the MND group compared to the other two algorithms.

The use of slide and baseline correction in the statistical technique improved the classification success rate by about 5%. In general, where all three algorithms failed to identify a MUAP class, it was because of inadequate number of class members in the signal and due to waveform variability.

MATLAB was used for implementing the above algorithms. The processing time on a PC Pentium 233 MHz for a 5-s epoch EMG signal with 77 waveforms was about 0.5 s for the segmentation and about 0.6 s for the classification with SOFM with LVQ, 0.4 s for SOFM, and 1 s for the statistical technique. The processing time for the decomposition of each superimposed waveform with three classes was about 0.02 s. Since MATLAB is an interpreter, all the timings may be significantly improved by the use of a compiled version of the algorithms.

V. CONCLUSION

In this work, two different pattern recognition techniques for the classification of MUAP's were investigated: i) an artificial neural network technique based on unsupervised learning, using modified SOFM and LVQ, and ii) a statistical pattern recognition technique based on the Euclidean distance. Both pattern recognition techniques described are quite simple in their concepts, and gave a high success rate. The ANN technique performed better than the statistical pattern recognition technique and yielded a higher success rate. ANN's seem more appropriate for the classification of MUAP's because of their ability to adapt and to create complex classification boundaries. The additional use of the LVQ algorithm with the SOFM algorithm optimizes the classification boundaries through slight adaptation of the weights vectors.

The improvement of the classification performance is clearly demonstrated in the case of the MND group which contains MUAP's with more complex and variable waveform shapes, where the classification success rate of the SOFM with the LVQ algorithm was considerably higher compared to the statistical one.

Several new ideas were introduced in this work in order to improve the performance of the algorithms.

1. Learning achieved in only one epoch in the SOFM and LVQ algorithms.
2. The factor in the LVQ algorithm in order to optimize the adaptation of the weights vector of the second winner.
3. The threshold in the classification phase of the ANN technique in order to separate the superimposed waveforms.
4. The combination of Euclidean distance and area measures in the decomposition of the superimposed waveforms in order to classify the decomposed waveforms.
5. The learning rate in the SOFM algorithm features the following characteristics:
   a. It is frequency sensitive; i.e., it decreases its value depending on the number of times an output node is selected winner,
   b. It gets narrower as the number of iterations increases.

In conclusion, the pattern recognition techniques as described in this work make possible the development of a fully automated EMG signal analysis system which is more accurate, simple, fast, and reliable enough to be used in routine clinical environment. Future work will evaluate the algorithms developed in this study on EMG data recorded from more muscles and more subjects. In addition, this system may be integrated into a hybrid diagnostic system for neuromuscular diseases based on ANN where EMG [8], muscle biopsy, biochemical and molecular genetics findings, and clinical data may be combined to provide a diagnosis [13].

REFERENCES

[1] E. Stalberg, S. Andreassen, B. Falck, H. Lang, A. Rosenfalck, and W.Trojaborg, "Quantitative analysis of individual motor unit potentials: A proposition for standardized terminology and criteria for measurement," *J. Clin. Neurophysiol.*, vol. 3, no. 4, pp. 313–348, 1986.

[2] L. J. Dorfman and K. C. McGill, "AAEE minimonograph #29: Automatic quantitative electromyography," *Muscle and Nerve*, vol. 11, pp. 804–818, 1988.

[3] R. S. LeFever and C. J. DeLuca, "A procedure for decomposing the myoelectric signal into its constituent action potentials: I. Technique, theory and implementation," *IEEE Trans. Biomed. Eng.*, vol. BME-29,pp. 149–157, Mar. 1982.

[4] "A procedure for decomposing the myoelectric signal into its constituent action potentials: II. Execution and test for accuracy," *IEEE Trans. Biomed. Eng.*, vol. BME-29, pp. 158–164, Mar. 1982.

[5] E. Stalberg, B. Falck, M. Sonoo, S. Stalberg, and M. Astrom, "Multi-MUP EMG analysis—A two year experience in daily clinical work," *Electroencelography and Clinical Neurophysiology 97*. Amsterdam, the Netherlands: Elsevier Science, 1995, pp. 145–154.

[6] P. Guihenec, J. Calamel, C. Doncarli, D. Gitton, and C. Michel, "Automatic detection and pattern recognition of signal motor unit potentials in needle EMG," *Computer-Aided Electromyography—Progress in Clinical Neurophysiology*, vol. 10, J. E. Desmedt, Ed. Amsterdam,the Netherlands: Elsevier Science, 1983, pp. 73–127.

[7] S. Andreassen, "Methods for computer-aided measurement of motor unit parameters," in *Proc. The London Symp.*, R. J. Ellington *et al.*, Eds., 1987, EEG suppl. 39, pp. 13–20.

[8] C.S. Pattichis, C.N. Schizas, and L.T. Mittleton, "Neural network models in EMG diagnosis," *IEEE Trans. Biomed. Eng.*, vol. 42, pp. 486–496, May 1995.

[9] S. Haykin, *Neural Networks—A Comprehensive Foundation*. New York: Macmillan College., 1994.

[10] B. Kroesse and P. Van der Smagt, *An Introduction to Neural Networks*. Amsterdam, the Netherlands: Univ. Amsterdam Press, 1993.

[11] T. Kohonen, "The self-organizing map," *Proc. IEEE*, vol. 78, pp. 1464–1480, Sept. 1990.

[12] R. P. Lippmann, "An introduction to computing with neural nets," *IEEE Acoust., Speech, Signal Processing, Mag.*, vol. 4, pp. 4–22, Apr. 1987.

[13] C. N. Schizas, C. S. Pattichis, and C. A. Bonsett, "Medical diagnostic systems: A case for neural networks," *Technol., Health Care*, vol. 2, pp.1–18, 1994.