Neural Network Screening of Electromyographic Signals as the First Phase to Design Novel Human-Computer Interaction

Pekka-Henrik Niemenlehto, Martti Juhola, Veikko Surakka

Department of Computer Sciences, University of Tampere, Finland
Tampere University Hospital, Department of Clinical Neurophysiology
P.O. Box 2000, FIN-33521 Tampere, Finland

Abstract

The present aim was to describe the first phase attempts to recognise voluntarily produced changes in electromyographic signals measured from two facial muscles. Thirty subjects voluntarily activated two facial muscles, corrugator supercilii and zygomaticus major. We designed a neural network based recognition system that screened out muscle activations from the electromyographic signals. When several subjects were tested according to the same test protocol, the neural network system was able to correctly recognise more than 95% of all muscle activations. This is a promising result and we shall next proceed to modify the system for real-time functioning and then design its utilisation for various multimodal human-computer interaction techniques. The subsequent phase in the future will be the interaction backwards: when a computer program first recognised the use of the facial muscles, it will then follow the instructions given by the user. For instance, by using the facial muscles the subject could select or activate objects on the computer screen. This would be one of the opportunities that we develop to help, e.g., disabled persons, who are unable to use their hands.

Keywords:
Biomedical Signal Analysis; Neural Networks; Electromyographic Signals; Human-Computer Interaction

1. Introduction

Human bioelectrical signals apply to several diagnostic purposes. They are also useful in other contexts. By recording some intentionally controlled phenomena via signal measurements, we are able to extend means for human-computer interaction (HCI) [1]. For example, moving the cursor on the screen by altering the gaze from one point to another seems a reasonable alternative for fast commands. In the present research we developed a method to recognise voluntary muscle activations from electromyographic (EMG) signals, so that it is possible to use facial muscles for HCI. The facial muscles can be used to accomplish various simple and rapid tasks, such as activating a link on the screen. One of the objectives is to help disabled persons to improve their possibilities to use computers. More generally thinking, in the future it might be possible to use...
multimodal ways to interact with a computer. Let us assume that a head band or eyeglasses would include various sensors to record, e.g., one’s facial muscle activity, eye movements (recorded with electro-oculography or small video cameras), heart activity, and brain activity. A computational challenge is then to analyse these signals and search for significant events from them. Our study is the first phase to design an overall HCI system. We tackled the signal analysis of facial EMG signals and the pattern recognition problem of muscular activity variations that appear during subject’s voluntary muscle activation.

Some studies were recently published in this field. Barreto et al. [2] studied the use of EMG signals for HCI with motor disabilities. Kübler et al. [3] developed a thought translation device that uses slow cortical potentials of a brain for the simple control influence of a computer. Wolpaw et al. [4] explored the same theme. Surakka et al. [5] studied the multimodal use of facial muscles and eye movements in HCI. We presented the idea to employ neural networks for the analysis of facial EMG signals in [6] and implemented it in [7] for two muscles. Earlier we experimented with a wide set of different multilayer perceptron (MLP) neural networks by varying the topology of the networks (for instance by using different number of hidden nodes), by testing with heavy artificial noise, and by testing various signal conditioning methods, including digital bandpass filters and wavelet denoising. There were only slight differences between various networks, except when we reduced the network to a very small form. Thus, at present we proceeded to use only a few suitable architectures and tested their capacity.

2. Signal Measurements and Recognition

Bipolar surface EMG signals were recorded from healthy 22 males and 8 females. The mean age was 24 years in the range of 20–35 years. The electrodes were placed on the left side of the face above corrugator supercilii (activated when frowning) and zygomaticus major (activated when smiling) muscle sites. The measurements were conducted according to common procedures described in [8]. The signals were recorded with Grass Model 15 RXi amplifier. The sampling rate was 3 kHz with a passband of 0.1–1.0 kHz. The average signal-to-noise (SNR) ratio was 12.7 dB. Signal examples from the previous muscle sites are shown in Figure 1.

During recordings every subject was seated comfortably and trained to carefully and calmly contract the two muscles, either one of them individually or both at the same time. Naturally, there is some baseline activation all the time in muscles, but its intensity is lower than during contractions that appear as bursts in EMG signals. Four 10 minutes recordings were accomplished for each subject. These tests followed one by one after a short resting period. The first period included the activation of the corrugator supercilii, the second period consisted of the activation of the zygomaticus major, the third part covered both of the muscles, and finally the last period was pure baseline recording. Due to the large amount of data recorded, we used only a part of it. Rest of the data will be used in our future research.
Figure 1 – (a) A 36 s part of a signal recorded from the corrugator supercillii and (b) another from the zygomaticus major.

Our previous work [7] showed that the use of wavelet denoising yielded slightly better classification results than using the more traditional digital bandpass filtering. While applying wavelets, high frequencies are analysed with a relatively effective time resolution, whereas low frequencies are dealt with a good frequency resolution. The transformation was calculated with the pyramidal algorithm [9], in which filters based on the mother wavelet are used. We utilised the Meyer wavelet as the mother wavelet. The principle of the wavelet denoising is as follows. First, a digital signal is transformed with the pyramidal algorithm up to the given level. In our tests the third level generated sufficiently reliable results. Second, the wavelet coefficients of every level are thresholded properly. We applied SURE threshold selection rule and soft thresholding which are described in more detail in [9]. Finally, the thresholded wavelet coefficients and the signal approximation of the last level are inverse transformed into the time domain representation by using the inverse form of the pyramidal algorithm.

To supply training, validation, and test sets, we randomly selected ten groups of three persons from all the subjects for the cross validation procedure [10]. We shared one group at a time to the test set, another to the validation set, and the other eight to the training set. Ten such shares were performed so that each group of three persons was once a test set and once a validation set. Thus we got ten series of sets. This procedure was performed for the data from both muscles. We used the mean square error measure to assess training of the network and checked the behaviour of the validation set to stop the training in time to avoid overtraining.

We randomly extracted 3 times 16 or 48 signal segments from the recording of a subject. Such a part of 16 signal segments corresponded to one of the three output classes applied to the classification: muscle activity onset, muscle activity offset, and plateau segment that contained...
muscle activity or baseline activity. In other words, there were no significant muscle activity changes in the plateau segments. Consequently, for two muscles of three subjects in a group there were 288 signal segments and 2880 altogether in the whole dataset. The number of 2304 signal segments in a training set was certainly enough for the largest MLP neural network, which consisted of two input nodes, five hidden nodes at two layers, and three output nodes. Thus the largest network contained $2 \times 5 + 5 \times 5 + 5 \times 3 + 13 = 63$ weights to be trained. Usually, the ratio of at least 10 is recommended between the two preceding figures [10]. Each signal segment was comprised of 1024 successive samples (approximately 0.34 s), which was an appropriate length to the wavelet denoising.

After the prior consideration of several different MLP neural network topologies [7], we restricted ourselves only to the three alternatives here. These were the afore-mentioned 2-5-5-3 network as the one of the maximal topology, a minimal network 2-0-3 (no hidden layer), and an average network 2-5-3. The essential aim was to hold the network size very concise to later apply the network for real-time processing. To gain this, it was important to use a minimal amount of input nodes. Therefore, we shared every signal segment of 1024 samples to halves and calculated a normalised root-mean-square (RMS) value for each. We first computed a RMS value for both

$$RMS = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} |x(i)|^2}$$

halves ($N=512$). One more RMS value was computed for the whole signal segment ($N=1024$). The ratio of RMS of either half and that of the whole segment is equal to normalised RMS. The normalisation permits the process to be independent of the magnitude of signal segments and the normalised RMS values are mapped to the real interval $[0, 2^{1/2}]$. This procedure compressed data considerably, but conserved sufficient information after the wavelet denoising for the recognition process of muscular activity in EMG signals. Either half of a signal segment thus yielded an input value into the network. The normalisation property was useful while scaling the input values to a suitable interval and evolving an invariant feature space with respect to signal amplitudes. The three nodes of the output layer expressed different classes which were mentioned previously: muscle activity onsets, plateau segments from the duration of muscular activity or baseline activity, and muscle activity offsets.

3. Results

When we had a 10-fold cross validation and executed 10 runs for each of them, we obtained a total of 100 runs. We then computed the means and standard deviations of recognition accuracy values (also called sensitivity):

$$a = \frac{tp}{tp + fn} \times 100 \% = \frac{tp}{n} \times 100 \%,$$

where $tp$ and $fn$ are the number of true positive and false negative classifications, respectively, and $n$ is the number of all test cases. In addition to the signals themselves, we also experimented with noisy signals by inserting random white noise to the original signal segments. Two variations were formed by incurring signal-to-noise ratios (SNR) of 5 dB and 1 dB. Table 1 encompasses results of
three neural networks tested.

The results in Table 1 show that there were virtually no differences between the three networks. Even the smallest network was able to recognise different types of signal segments very well. It seems that the signal segments were successfully separated in a feature space of two normalised RMS values. Otherwise, such a minimal network could not separate the test cases into the different classes so effectively. We also observed the subtle deterioration in the case of the strongest noise.

Table 1 –Recognition accuracy: means and standard deviations of 100 runs for three neural networks and three signal-to-noise ratios.

<table>
<thead>
<tr>
<th>Network</th>
<th>Number of iterations</th>
<th>Mean square error of training set</th>
<th>Mean square error of validation set</th>
<th>Recognition accuracy [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Original signals</td>
</tr>
<tr>
<td>2-0-3</td>
<td>310±20</td>
<td>0.0084±0.0010</td>
<td>0.0089±0.0056</td>
<td>98.3±1.0</td>
</tr>
<tr>
<td>2-5-3</td>
<td>250±28</td>
<td>0.0083±0.0010</td>
<td>0.0086±0.0055</td>
<td>98.3±1.1</td>
</tr>
<tr>
<td>2-5-5-3</td>
<td>225±21</td>
<td>0.0083±0.0010</td>
<td>0.0088±0.0055</td>
<td>98.3±1.1</td>
</tr>
</tbody>
</table>

The results in Table 2 were computed similarly as those reported in Table 1. In this case the recognition accuracies of different output classes are presented. Only results of the 2-5-3 network are presented, because the variations between the three networks were again slight.

Table 2 –Recognition accuracy: means and standard deviations of 100 runs for 2-5-3 neural network and three signal-to-noise ratios employing three output classes.

<table>
<thead>
<tr>
<th>Output class</th>
<th>Recognition accuracy [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Original signals</td>
</tr>
<tr>
<td>Onset segment</td>
<td>99.3±0.7</td>
</tr>
<tr>
<td>Plateau segment</td>
<td>97.7±1.2</td>
</tr>
<tr>
<td>Offset segment</td>
<td>97.9±2.7</td>
</tr>
</tbody>
</table>

4. Discussion

The developed neural networks functioned properly. On the basis of the current dataset of 30 subjects, it was possible to reliably recognize voluntary muscle activity from the EMG signals
measured from *corrugator supercilii* and *zygomaticus major* facial muscle sites. This probably enables us to expand to other (facial) muscles as well. Because the sizes of the networks were very limited, there is an excellent opportunity to construct a real-time neural network based recognition system. To use such a system, a calibration is typically needed. This task could include the training of the network. Even more simply, we might collect a vast dataset of EMG signals from many subjects and rely on it, thus rendering the calibration phase unnecessary. In the future, we shall perform more measurements with more subjects of various ages, build a real-time system, and test some simple human-computer actions, for instance, selection of objects on the computer screen with volitional muscle activations. One practical viewpoint is to diminish difficulties caused by several electrode wires attached to a subject; we shall use wireless electrode recordings to alleviate this practical issue.

**Acknowledgments**

This research was supported by the Academy of Finland (grants 202185 and 1202183).

**References**


**Address for correspondence**

Pekka Niemenlehto, Department of Computer Sciences, 33014 University of Tampere, Finland; Pekka.Niemenlehto@cs.uta.fi

Section 7: Imaging Informatics