Pattern classification of hand movements using time domain features of electromyography

This item was submitted to Loughborough University's Institutional Repository by the/an author.

Citation: ROBINSON, C.P. ...et al., 2017. Pattern classification of hand movements using time domain features of electromyography. Presented at the 4th International Conference on Movement Computing (MOCO '17), London, 28-30th June, Article no 27.

Additional Information:

- Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. The definitive Version of Record was published in Communications of the ACM: https://doi.org/10.1145/3077981.3078031

Metadata Record: https://dspace.lboro.ac.uk/2134/26360

Version: Published

Publisher: © Association for Computing Machinery (ACM)

Please cite the published version.
Pattern Classification of Hand Movements using Time Domain Features of Electromyography

Carl Peter Robinson  
Loughborough University  
Department of Computer Science  
Loughborough, UK LE11 3TU  
c.p.robinson@lboro.ac.uk

Qinggang Meng  
Loughborough University  
Department of Computer Science  
Loughborough, UK LE11 3TU  
Q.Meng@lboro.ac.uk

Baihua Li  
Loughborough University  
Department of Computer Science  
Loughborough, UK LE11 3TU  
B.Li@lboro.ac.uk

Matthew T.G. Pain  
Loughborough University  
School of Sport, Exercise and Health Sciences  
Loughborough, UK LE11 3TU  
M.T.G.Pain@lboro.ac.uk

ABSTRACT

Myoelectric control of prostheses is a long-established technique, using surface electromyography (sEMG) to detect the electrical signals of muscle activity and perform subsequent mechanical actions. Despite several decades’ research, robust, responsive and intuitive control schemes remain elusive. Current commercial hardware advances offer a variety of movements but the control systems are unnatural, using sequential switching methods triggered by specific sEMG signals. However, recent research with pattern recognition and simultaneous and proportional control shows good promise for natural myoelectric control. This paper investigates several sEMG time domain features using a series of hand movements performed by 11 subjects, taken from a benchmark database, to determine if optimal classification accuracy is dependent on feature set size. The features were extracted from the data using a sliding window process and applied to five machine learning classifiers, of which Random Forest consistently performed best. Results suggest a few simple features such as Root Mean Square and Waveform Length achieve comparable performance to using the entire feature set, when identifying the hand movements, although further work is required for feature optimisation.

CCS CONCEPTS

• Computing methodologies → Machine learning approaches;  
Feature selection; • Human-centered computing → Interaction techniques; • Computer systems organization → Robotics;

KEYWORDS

Electromyography, Myoelectric control, Time domain features, Machine learning

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM Reference format:
https://doi.org/http://dx.doi.org/10.1145/3077981.3078031

1 INTRODUCTION

Wearable sensing technology has provided the capacity for continuous, real-time monitoring of physiological data, enabling numerous human-centred applications in health monitoring [14]. The use of electromyography (EMG) has found particular application in the clinical and commercial use of prostheses for amputees. EMG is a measurement of the cumulative electrical signal generated by the activity of a group of muscles. Termed “myoelectric control”, this method has been established for around fifty years [16] with early controllers monitoring EMG signal amplitude and performing a specific function once a designated threshold was passed [7]. The same principle is used today, driving one or two degrees of freedom using this on/off approach, or proportionally, within a finite interval according to the amplitude levels from opposing muscles [12]. Focusing on upper extremity devices, commercially, great advances have been made in hardware development with products now performing a variety of movements and gestures, including the provision of individual finger manipulation, offering superior potential to past apparatus. Despite such improvements, control strategies for these devices are still not natural, mainly consisting of sequential functionality of a disparate set of movements that require specific sEMG signals or patterns to switch between and select them [2]. There remains a gap between hardware capability and robust, intuitive, responsive control that a user can achieve. If often leads to frustration and abandonment of these complex tools, in favour of a less technical or cosmetic prosthesis [11].

EMG signals are recorded by electrodes, either non-invasively, using surface EMG (sEMG) or invasively, by intramuscular EMG. In the non-invasive method, electrodes are fixed to the skin’s surface to record the voltage potential difference of muscles. This approach can suffer from limitations including muscle crosstalk and identifying adequate electrode location in relation to muscle activity [4, 13].
Intramuscular EMG offers some resilience to this, by requiring the insertion of a needle into the muscle itself. But this incurs its own problems including minor tissue damage, infection risk and a subject’s potential aversion to needles [5]. A more amenable administration and being a usually cheaper alternative, sEMG use has been more extensive in the research field.

This paper uses a benchmark database of sEMG data, representing a series of human hand movements, to evaluate sets of time domain features with five machine learning classifiers, in terms of how accurately the hand movements are identified. Data is normalised and segmented into windows from which the features are extracted and formatted into a series of feature vectors. Classification trials are then performed, splitting the feature vectors into training and testing data and using a workbench application of machine learning algorithms.

2 METHOD

2.1 Benchmark Database

The Ninapro project (http://ninapro.hevs.ch) provides a repository of sEMG data from both intact and amputee subjects. Currently, there are three databases available, each containing results from a series of exercises where subjects performed sets of hand, wrist and finger movements in controlled laboratory conditions.

12 wireless electrodes and a base station (Trigno Wireless System, Delsys Inc.) were used to measure the sEMG signals from each subject. Eight electrodes were located equidistantly around the proximal section of the right forearm, at the height of the radio-humeral joint. Two more were attached to the main activity spots on the anterior and posterior of the forearm and finally, two more placed on the biceps brachii and triceps brachii. Subjects were seated at a desk with their forearm resting comfortably on the surface and guided to perform a series of movements, which they repeated 6 times before moving on to the next. One repetition lasted around 5 seconds, followed by a 3 second rest, where the subject returned to a rest posture. Data were acquired at a 2 KHz sampling rate and recorded via USB cable to a laptop. Prior to making the datasets available online, the sEMG signals were post-processed. This included cleaning the signals from 50Hz (and harmonics) power-line interference and relabelling the movements to correct for mismatches in subject movement execution times. Detailed protocol information regarding the acquisition procedure can be found in [3].

For this experiment, datasets for the first 11 intact subjects from Database 2 were downloaded and the 17 hand and wrist movements of Exercise B were considered, covering a varied range of actions, depicted in Figure 1.

2.2 Preprocessing

Datasets from the Ninapro website were downloaded as formatted MATLAB .mat files. An in-house MATLAB program was written to take the relevant sEMG and movement repetition label detail from these files. This resulted in a movement signal matrix per subject, consisting of 17 rows representing each movement and 6 columns defining the movement repetitions. Each repetition consisted of a time-ordered series of sEMG voltage data from 12 electrodes, in the form of a \( N \times 12 \) matrix. Repetitions 1, 3, 4 and 6 were assigned to a training set leaving repetitions 2 and 5 for the test set and all data were normalised to have zero mean and unit standard deviation, as per [3].

Due to the absence of maximum voluntary contraction (MVC) information in the downloaded datasets, a maximum reference value per electrode was also applied during normalisation, acting as a substitute MVC. For each subject, this consisted of a vector of 12 sEMG values, identified as the subject’s peak sEMG voltages in the training set, measured by each electrode through the entire exercise (all movements). The electrode values in this vector were then used to normalise corresponding electrode data for every movement repetition. This was an attempt to minimise the variance between the subjects’ sEMG data [6] when it was used for inter-subject experimentation. It was deemed necessary after poor initial inter-subject classification results.

2.3 Windowing and Feature Extraction

A sliding window technique with an increment of 10 ms was used to segment the sEMG data into an overlapping sequence of windows of 256 ms in length (a sample rate of 2 KHz equated to 512 samples per window with a 20-sample increment). This ensured the known 300 ms threshold was not crossed, identified as an acceptable delay a prosthesis user is unlikely to detect [16]. It also generated a dense array of windows, increasing the potential for pertinent feature extraction and mitigating the possibility of missing a feature lying between windows. A moving average was however required, to prevent the resulting file sizes from becoming unmanageable, particularly when these files were then combined for inter-subject analysis; an initial feasibility investigation identified this to be an issue for the classification stage. The moving average was applied...
Figure 2: Sliding window technique employed to segment one electrode’s sEMG data. Three windows are presented ($w_1$, $w_2$, $w_3$) each of 256 ms length. A 10 ms window increment is represented by $s$.

by averaging the signal content of 5 consecutive windows into a single window ($AW$), then repeating this process for the next 5 windows and so on, until the end of the sEMG data for that movement repetition was reached:

$$AW_t = \frac{1}{n} \sum_{j=1}^{n} w_{nt-j}$$

where $n$ is the moving average of 5 and $w_{nt-j}$ identifies one segmented window. The segmentation and averaging process was applied to each electrode’s movement repetition sEMG data, resulting in a set of windows for each electrode.

Appropriate feature selection has been identified as the major determining factor to successful pattern recognition performance [9]. A combination of time domain and autoregressive (AR) features offer the best performance, but at an appreciable cost to computational requirements over only using time domain features [16]. This experiment focuses on a selection of time domain features, without the AR component, to evaluate the effect of feature set size and feature type, on hand movement classification accuracy.

The features were chosen based on related work with the Ninapro database [3] and previous studies in time domain feature selection [15]. We start with the time domain statistics (TD), proposed by Hudgins et al. (1993) and used regularly in the research field, consisting of Mean Absolute Value (MAV), Mean Absolute Value Slope (MAVSlope), Waveform Length (WL), Slope Sign Changes (SSC) and Zero Crossings (ZC) — more detailed information can be found in [10]. The Histogram (HIST) feature captures the number of times a range of signal amplitudes are measured within a window by equally dividing the sEMG signal range in a window, into several bins [17]. In this case, 10 bins were used, to reduce computing complexity, and a threshold of 3 standard deviations was applied. The Root Mean Square (RMS) feature of the sEMG signal was used as it relates relatively linearly to the contraction intensity of the associated muscles [16]. It is another indicator of the average signal value, cancelling out negative values by squaring them to obtain a mean value. Finally, the Variance (VAR) feature was included to measure the sample variability of the sEMG signal, calculated as the average of the squared sample values in one window.

These features were extracted from each window, producing a feature vector of 17 scalar values, one each for the 12 electrodes (the five TD features, RMS, VAR and the HIST feature’s 10 bins). This in total produced a 204-value feature vector $f_{t,e} = (f_1, \ldots, f_{204})$, for one window, where $e$ is one electrode, $t$ refers to one time window and $f$ is a single feature scalar value. This yielded a time-ordered feature vector matrix $F_{T,E}$:

$$F_{T,E} = \begin{bmatrix} f_{11} & \ldots & f_{1E} \\ \vdots & \ddots & \vdots \\ f_{T1} & \ldots & f_{TE} \end{bmatrix}$$

which represented the recorded sEMG data for a movement repetition. The outcome was two movement feature matrices for every subject, the first consisting of the 4 feature vector matrices for the training set (corresponding to movement repetitions 1, 3, 4 and 6) and the second containing the 2 feature vector matrices for the test set (corresponding to movement repetitions 2 and 5). Variations of the movement features matrices were created to make seven different feature configurations, as input for classification trials. Table 1 outlines the features in relation to which configurations they were used in (C1 to C7). These configurations were chosen based on results from the attribute evaluation tools as part of the machine learning application described in the next section, and initial exploratory classification tests.

<table>
<thead>
<tr>
<th>Feature</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD Features:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Abs Value</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Abs Value Slope</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Waveform Length</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Slope Sign Changes</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Zero Crossings</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Histogram</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMS</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variance</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.4 Classification

All classification tests were performed using the WEKA application, version 3.8 (www.cs.waikato.ac.nz/ml/weka/), a machine learning platform that provides a variety of ready-to-use algorithms with
Table 2: Classifier parameter settings used in classification tests, identified by manual tuning and WEKA application’s optimisation tools.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Parameter Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>kNN</td>
<td>k=7, Search Algorithm=Manhattan Distance</td>
</tr>
<tr>
<td>MLP</td>
<td>Default Settings: Learning Rate=0.3, Decay=False, Number of Epochs=500</td>
</tr>
<tr>
<td>RF</td>
<td>Default Settings: Number of Features=8</td>
</tr>
<tr>
<td>SVM-POLY</td>
<td>Default Settings: c=1</td>
</tr>
<tr>
<td>SVM-RBF</td>
<td>c=8, gamma=0.1</td>
</tr>
</tbody>
</table>

configurable parameters. The 11 subjects’ training data were combined into a single overall training set and used in a series of inter-subject classification tests. Both the training set and all subject test sets were formatted as per WEKA’s requirements, into comma separated value files and each time window was given a class label, to identify which movement the time window’s feature vector represented. For each classification test, a 10-fold cross validation was performed to provide estimated performance and produce a classifier model to test against. Each subject’s test data was then supplied to the model to test classification precision. Five classifiers were used from the WEKA suite consisting of: k-Nearest Neighbour (kNN), Multilayer Perceptron (MLP), Random Forest (RF), and two variants of a Support Vector Machine, one using a polynomial kernel (SVM-Poly) and another using a radial basis function kernel (SVM-RBF). Parameter optimisation was applied to the classifiers using WEKA’s built-in CVParameterSelection meta-classifier and experimenter tools, although a combination of additional manual tuning and computing power constraints resulted in default parameter settings for the MLP, RF and SVM-POLY classifiers. Table 2 lists the relevant classifier parameter settings.

3 RESULTS

The classification rate (CR) was used to evaluate classifier performance, as a measure of accuracy in correctly identified hand movements. CR is a simple ratio, obtained by dividing the number of correctly classified instances over the total number of instances:

\[
\text{Correctly Classified Instances} \times 100\% = \frac{\text{Correctly Classified Instances}}{\text{Total Number of Instances}} \times 100\%
\]

The first trial used the first configuration (C1) with all five classifiers, to establish a reference and provide an indicator of classifier capability. As can be seen by the chart in Figure 3, the RF classifier exhibited the best performance among all classifiers with an average accuracy of 88.44%. This is 15.69% higher than the 72.75% achieved by the next best classifier, kNN. The SVM-RBF classifier gave similar performance to the kNN classifier, with 71.07%, but both the SVM-Poly and MLP classifiers struggled, producing very poor average accuracies of 50.08% and 36.31%, respectively.

The three classifiers with highest accuracies (RF, kNN and SVM-RBF) were then used in trials with feature subsets, specified in configurations C2 to C7. This was to compare the current results with classifier performance when using smaller numbers of features, to identify an optimum classifier/feature configuration given the utilized data. The results are displayed in Figure 4.

The RF classifier performed best again, for all configurations. But of more significance were the results for smaller feature sets, particularly when used with the kNN and RF classifiers. For RF, configurations C4, C5, C6, and C7 achieved accuracies of 89.59%, 89.84%, 90.53%, and 90.57%, respectively, all outperforming the 88.44% from the configuration with all features (C1). C7 had the greatest improvement of 2.13% followed by C6 with 2.09%. It is worth noting here that C7 consisted of the five TD features while C6 used only three features (WL, RMS and SSC), to accomplish a near identical result. Similarly, C4, with only two features (WL and RMS), was only 0.98% behind C7’s performance. Table 3 provides a comparison of the feature configurations and their average classification accuracies using the RF classifier. For kNN, all feature subset configurations outperformed the 72.75% accuracy achieved by C1. Here interestingly, the three-feature configuration of C6 attained the best result with 85.30%, an improvement of 12.55%. This was followed by the two-feature C4 configuration with 85.10%, improving by 12.35%. Both bettered the TD features of C7, which achieved 84.73%. The SVM-RBF classifier only advanced beyond its C1 result when using the C7 configuration, with an accuracy of 72.37%. 

![Figure 3: Average classification accuracy results for 11 subjects, using five classifiers with feature configuration C1.](image1)

![Figure 4: Average classification accuracy results using the Random Forest classifier with seven feature configurations, C1 to C7.](image2)
Table 3: Comparison of all feature configurations and their average classification accuracy results using the RF classifier. The number of dimensions is calculated as the number of feature scalar values in each configuration multiplied by the number of electrodes, in this case 12.

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Num. Features</th>
<th>Num. Dimensions</th>
<th>Test Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>8 (All)</td>
<td>204</td>
<td>88.44%</td>
</tr>
<tr>
<td>C2</td>
<td>1 (RMS)</td>
<td>12</td>
<td>85.35%</td>
</tr>
<tr>
<td>C3</td>
<td>1 (WL)</td>
<td>12</td>
<td>88.25%</td>
</tr>
<tr>
<td>C4</td>
<td>2 (RMS, WL)</td>
<td>24</td>
<td>89.59%</td>
</tr>
<tr>
<td>C5</td>
<td>3 (RMS, WL, ZC)</td>
<td>36</td>
<td>89.84%</td>
</tr>
<tr>
<td>C6</td>
<td>3 (RMS, WL, SSC)</td>
<td>36</td>
<td>90.53%</td>
</tr>
<tr>
<td>C7</td>
<td>5 (TD)</td>
<td>60</td>
<td>90.57%</td>
</tr>
</tbody>
</table>

Results for all three classifiers show that a small number of time domain features can produce classification accuracy equivalent to, if not better than, a larger set of time domain features, as conveyed in other research [8]. While particularly true of kNN, it is more significant for RF, which maintained its optimum performance despite a decrease in feature configuration size. This is an important factor when considering computational complexity, as a reduction in feature vector size can aid in a more responsive myoelectric control system. In this case, a reduction from C1’s 204 dimensions down to 60 for C7 and 36 for C6 can be seen as a compelling statement for the importance of appropriate feature choice. Figure 5 shows the impact on classification accuracy when using the seven different feature configurations in terms of their size, from smallest to largest. Figure 6 shows the average confusion matrices detailing classification and misclassification of the 17 hand and wrist movements for C6 and C7. They are based on results from all 11 subjects during classification tests using the RF classifier. From this it can be seen that actual movement identification success is highly comparable between the two configurations. C6 identifies movements 1, 2, 4, 7, 9, 10, and 14 more accurately while C7 scores higher with the remaining movements, although the majority of the results are marginal.

The TD features of C7 were the highest performing feature subset in this experiment, achieving higher accuracy than the foremost feature set in comparable work [3], as did all feature subsets used here. In that case, a combination of RMS, TD, HIST and marginal Discrete Wavelet Transform (mDWT) features attained the highest accuracy of 75.27%. (Note that the HIST configuration used had 20 bins as oppose to the 10 bins here.) While this undertaking does not involve frequency domain features, it does present alternative, smaller feature combinations that show almost equivalent performance to the TD features of C7, to within 1% regarding C4 and C5, and only 0.04% in the case of C6.

However, a fair comparison with that work cannot be applied entirely, as only 11 subjects and 17 movements were used here, as oppose to the 40 subjects and 52 movements in [3]. In fact it has been stated that working with a small number of classes can produce accuracy results in the 90% and higher range [1]. Another factor to consider is the effect of the window averaging process applied during segmentation. While required, to maintain feasible
computation during classification trials, there is an argument for possible important feature loss due to its employment. It will be addressed in future work, perhaps by applying a more suitable subsampling method or varying the length of the window and increment size to produce a smaller volume of windows.

4 CONCLUSIONS
The work presented here investigates seven time domain feature configurations with five machine learning classifiers, using a selection of readily-available sEMG datasets from the Ninapro database, representing a variety of hand movements. The results indicate using a few scalar value features can achieve optimum performance with kNN or RF classifiers. The RF classifier performs best out of all five, using a feature configuration of either the five TD features or of WL, RMS and SSC features. Further work is required to evaluate features from other domains and to expand the number of subjects and movements used in trials, widening the variety of sEMG data under test and establishing better comparison with existing research.

ACKNOWLEDGMENTS
The authors would like to thank the Ninapro project (Non-Invasive Adaptive Prosthetics) for freely providing the sEMG datasets, associated metadata and experiment details.

REFERENCES