Article

Hand Movement Classification using Burg Reflection Coefficients

Daniel Ramírez-Martínez [1], Mariel Alfaro-Ponce [2], Oleksiy Pogrebnyak [1,‡], Mario Aldape-Pérez[3] and Amadeo-José Argüelles-Cruz [1,*]

1 Centro de Investigación en Computación, Instituto Politécnico Nacional. Av."Juan de Dios Bátiz" s/n esq. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, Del. Gustavo A. Madero, Ciudad de México; dhanielrhamirez@gmail.com; oleksiy@cic.ipn.mx; jamadeo@cic.ipn.mx

2 Departamento de Ciencias e Ingenierías, Universidad Iberoamericana Puebla. Blvrd del Niño Poblano 2901, Reserva Territorial Atlixcáyotl, Centro Comercial Puebla, 72810 San Andrés Cholula, Pue. México; marielalfa@gmail.com

3 Centro de Innovación y Desarrollo Tecnológico en Cómputo, Av."Juan de Dios Bátiz" s/n esq. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, Del. Gustavo A. Madero, Ciudad de México, C.P. 07700.; maldape@ipn.mx

* Correspondence: jamadeo@cic.ipn.mx; Tel.: +52-55-5729-6000 ext.56593

‖ ‡ In Memoriam

Version October 29, 2018 submitted to Preprints

Abstract: Classification of electromyographic signals has a wide range of applications, from clinical diagnosis of different muscular diseases to biomedical engineering, where their use as input control of prosthetic devices has become a hot topic of research. Challenge of classifying this signals relies on the accuracy of the proposed algorithm and the possibility of its implementation on hardware. This paper consider the problem of electromyography signal classification, solved with the proposed signal processing and feature extraction stages, with focus lying on the signal model and time domain characteristics for better classification accuracy. The proposal considers a simple preprocessing technique that produces signals suitable for feature extraction, and the Burg reflection coefficients to form learning and classification patterns. These coefficients yield a competitive classification rate compared to used time domain features. Sometimes, the feature extraction from electromyographic signals showed that procedure can omit less useful traits for machine learning models. Using feature selection algorithms provides a higher classification performance with as fewer traits as possible. Algorithms achieved a high classification rate up to 100% with low pattern dimensionality, with other kinds of uncorrelated attributes for hand movement identification.

Keywords: Electromyography, Hand Movement, Health Monitoring, Maximum Entropy Reflection Coefficients, Classification Algorithms, Machine Learning, Feature Selection.

1. Introduction

Electromyography (EMG) is an electrodiagnostic medical procedure to assess the health of muscles and the nerves cells that controls them, with the detection, recording and analysis of electromyography signals (sEMG). The EMG provide physicians and health experts with the information generated by the muscle contractions, that is the ionic flow through the muscle fibre [1]. Research has considered EMGs as an important field of study due to the diversity of its applications in clinical medicine and biomedical engineering [2]. EMG has applications such as the diagnosis of nervous system disorders and muscular diseases like myopathies detection and neuropathies [3–5]. All these applications need preprocessing of signals and its extraction of features [6]. sEMG are useful as input control signals for prosthetic limbs [7,8], in rehabilitation as a measurement parameter of muscular effort [9], and for the development of muscle machine interfaces [10].

Most EMG applications involves real-time systems, that are needed to run with low-cost computational features [11]. As a matter of the fact, in the development of prosthetic, orthotic and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0647.v1
Peer-reviewed version available at Sensors 2019, 19, 475; doi:10.3390/s19030475

rehabilitation devices the EMG can be employed as a part of the control system. In the results reported by [12] EMG pattern recognition and myoelectric control are compared for prosthetic control, the paper remarks that these signals are suitable for the control with highlights in the implementations of algorithms that are capable to distinguish between signals that have similarities. These similarities are presented in users that have lost a body part, as a consequence of the absence of peripheral structures in the musculoskeletal system, where the classification of EMGs features become a challenge. The most used EMGs features are time domain attributes [13] that can be obtain by root-mean-square value (RMS), mean average value (MAV), variance (VAR), Willison amplitude (WAMP), wavelength (WL), and many others [14] [15]. These characteristics have been used in different classification tasks, and the classification rate increases with the use of a proper signal preprocessing stage; for instance, Chowdhury et al. consider the use of wavelet and empirical mode decomposition, first differentiation or independent component analysis in [15]. Despite the performances achieved at the preprocessing stage, computational complexity might increase adding a delayed response. Several authors use autoregressive models and the characteristics of random processes, such as first and second moments and others, in tasks related with classification of myopathy or neuropathy diseases. For example, Bozkurt et al. report 97% in performance using fifteenth order autoregressive models (AR) Yule-Walker, Burg, Covariance, Modified Covariance and subspace base methods to extract features from 1200 sEMG, applying high resolution and high sampling rate in invasive electrodes implanted in a Bicep brachii muscle [16].

In a different research dedicated to the hand movement, Phinyomark et al. report a high classification rate of 97.76%, achieved by applying a quadratic discriminant analysis and four AR coefficients per channel and including a preprocessing stage whose output is the first differentiation of sEMG [17]. They extracted information from the activity of five forearm muscles: WL, difference absolute mean value (DAMV), difference absolute standard deviation value (DASDV), difference absolute variance (DVARV), difference absolute standard deviation (DASDV), second order moment (M2), WAMP, integrated EMG (IEMG), MAV. They used also these features in their previews works [18]. Simple squared integration SSI, VAR, RMS, myopulse percentage rate (MYOP), cepstral coefficients (CC), log detector (LOG), temporal moment (TK) and v order (V) with another point of view in [17] applying the seventh order Daubechies mother wavelet and the four decomposition levels before sEMG characterisation extracts RMS and MAV. They test the behaviour of these traits to estimate whether they are useful for identification of six daily hand movements monitoring flexor and extensor carpi radialis longus muscles.

Liu et al. describe the use of an assemble classifier of support vector machines (SVM) [19], classifying eight different hand grasps with a precision rate of 93.54%; extracting sEMG from three different forearm muscles and fourth order AR coefficients and HEMG builds the feature vector per channel. In their work the aim is to get significant features and a classification model that permits increasing the classification rate of sEMG. However, assembling SVM models results computationally expensive. Angari et al. consider fifteen channels to digitalise sEMG and to characterise five hand movements, where they extract twenty-one attributes per channel (MAV, WL, ZC, SSC, AR6, and others) to implement feature selection methods and channel discrimination [20]. In this case, the aim of the research was to train the SVM with low dimensionality patterns and the most representative forearm muscles; this work concludes that MAV and WL are appropriative for classification tasks.

The method of Khezri et al. uses an adaptive neuro–fuzzy inference system to test its classification rate in a six hand movement dataset containing four channels [21]. The considered features are MAV, SSC, ZC, and 10 order AR model coefficients. Merging these attributes to create patterns for sEMG representation provokes that classification rates runs from 86% to 100%.

Ruangpaisarn et al. present a feature extraction technique for hand movement classification, considering two pairs of EMG electrodes and the merging and transformation of both channels into a squared matrix to perform factorisation via singular value decomposition [22]. They report the use of singular values in the matrix main diagonal and the training of SVM with fifty feature

 doi:10.20944/preprints201810.0647.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 475; doi:10.3390/s19030475

instances, achieving a performance of 98.22%. The issue in this work comprises taking samples where no muscular activity is looked at, and working with a 2D vector in most cases leads to non-linear computational complexity. With the same dataset, Sapsanis et al. used a preprocessing stage in which signals are 3 level decomposed with empirical mode decomposition, so that noise is reduced [23]. For each decomposition level and raw sEMG, they extract the following attributes: IEMG, ZC, VAR, SSC, WL, WAMP, kurtosis and skewness. With a linear discrimination analysis, a rate of correct classifications reaches 89.21%.

The resumes in [14] and [15] show the variety of features for classifying sEMG and preprocessing techniques that might lead to a classification model performance increasing. However, none of the studies use the reflection coefficients as features for pattern recognition.

The aim of this work is to develop a classification algorithm for sEMG with low computational cost and with a competitive classification rate. The remainder of this paper present the following distribution: Section 2 describe the hand movement database that was employed, also a brief resume of the signal preprocessing techniques and different features useful for classification are described. Then, the proposed classification method is presented. Section 3 shows the results obtained by the classification technique. Section 4 and 5 are the discussion and conclusion of the results achieved by the proposed methodology for sEMG classification.

2. Materials and Methods

2.1. Data selection and preprocessing

We used an EMG dataset from the University of California in Irving (UCI) machine learning repository, the same as in [22] and [23]. The data describes six different hand movements taken from Flexor Capri Ulnaris and Extensor Capri Radialis muscles of five healthy people (three women and two men) who performed with no restrictions thirty times each hand action during 6 seconds each; signal sampling frequency is 500Hz. The dataset contains 1800 time series available to classify 6 hand grasps (Spherical, Tip, Palmar, Lateral, Cylindrical and Hook).

Before feature extraction, there is a simple preprocessing treatment applied to each signal. At the first preprocessing stage, the method eliminate the initial samples, where muscle activation is absent and only noise is present, to avoid the feature extraction lacking the phenomena information. The next step comprises the extract of the signal-mean-value at all data points; this operation is important to comply with the restrictions imposed by the optimal linear filtering theory [24]. Linear prediction model framework requires restrictions such as autoregressive models. Otherwise, performing classification/prediction models might decrease. Here, we used the simple arithmetic mean value computed as

$$\mathbf{x} = \frac{1}{N}\sum_{n=0}^{N} x[n] \tag{1}$$

where $\mathbf{x}$ is the mean value; $x[n]$ is EMG signal and N is the total number of samples. The application of the mean (1) implies a new sample value which is described as $\overline{x}[n] = x[n] - \overline{x}$ for $1 \leq n \leq N$. Figure 1 illustrates the proposed preprocessing stages.

These two conditioning steps have linear complexity and supply feature extraction stage with an appropriated sEMG.

2.2. Standard time domain features

The integrated EMG feature is defined as the cumulative addition of each signal sample absolute value:

$$IEMG = \sum_{n=0}^{N} |x_n| \tag{2}$$

Peer-reviewed version available at Sensors 2019, 19, 475; doi:10.3390/s19030475

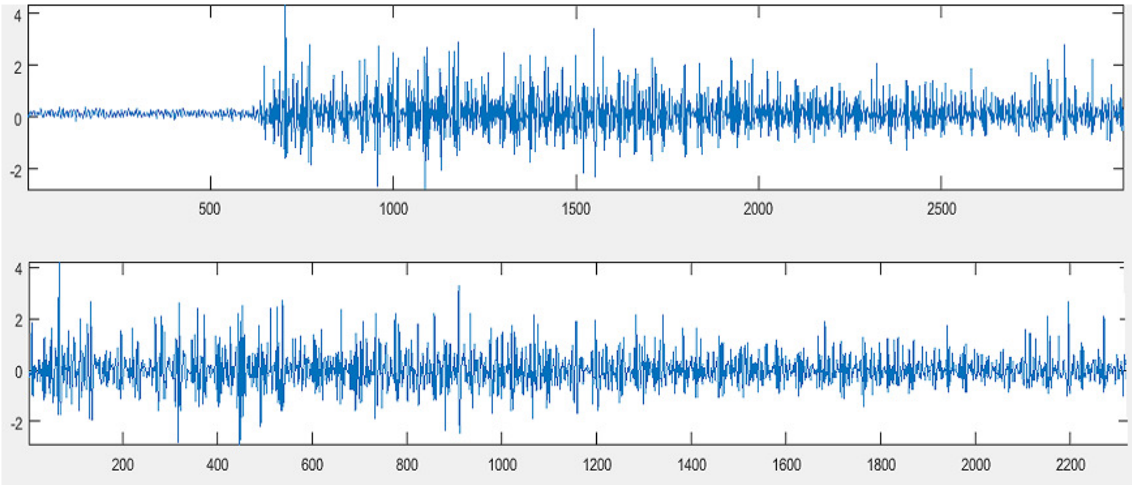

Figure 1. On top original signal, on bottom clipped signal with zero mean value.

where other attribute is mean absolute value; it is one of the most useful attributes in many researches and consists of computing the mean absolute amplitude value of sEMG:

$$MAV = \frac{1}{N}\sum_{n=0}^{N}|x_n| \tag{3}$$

The simple squared integration feature describes the energy of sEMG and is mathematically defined as cumulative addition of absolute squared value of each sample:

$$SSI = \sum_{n=1}^{N}|x_n|^2 \tag{4}$$

A stochastic process as a sEMG can be defined by its first and second order moment, i.e., mean and variance values. Therefore, these features might be part of the pattern. The mathematical definition for the variance takes considers a sEMG is a near to zero mean process, so its definition becomes:

$$VAR = \frac{1}{N-1}\sum_{n=1}^{N}{x_n}^2 \tag{5}$$

The root mean squared value (RMS) reveals the information of the amount of strength yield by a muscle, and is defined as the square root of the mean squared values. In many research works, this attribute is considered important for different tasks:

$$RMS = \sqrt{\frac{1}{N}\sum_{n=1}^{N}{x_n}^2} \tag{6}$$

The wave length is a distance between a pair of adjacent samples along all sEMG:

$$WL = \sum_{n=1}^{N}|x_{n+1} - x_n| \tag{7}$$

The zero crossing feature describes the number of times that the sEMG amplitude becomes positive or negative. Its definition considers a threshold whose aim is to count only the events produced by muscular activity:

$$ZC = \sum_{n=1}^{N-1}\left[sgn(x_n \times x_{n+1}) \bigcap |x_n - x_{n+1}| \geq 0\right], sgn(x) = \begin{cases} 1, & x \geq threshold \\ 0, & otherwise \end{cases} \tag{8}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0647.v1
Peer-reviewed version available at Sensors 2019, 19, 475; doi:10.3390/s19030475

The slope sign attribute considers three adjacent samples to determine the number of times that a slope sign between these sEMG values changes:

$$SSC = \sum_{n=2}^{N} f((x_n - x_{n-1}) \times (x_n - x_{n+1})), f(x) = \begin{cases} 1, & f = th \\ 0, & otherwise \end{cases} \tag{9}$$

where th is a threshold.

The quantity of motor unit action potential is estimated through the Willison amplitude counting the number of times that two adjacent samples overcome a threshold reducing artifacts produced by noise:

$$WAMP = \frac{1}{N} \sum_{n=1}^{N} f(|x_n|), f(x) = \begin{cases} 1, & x \geq th \\ 0, & otherwise \end{cases} \tag{10}$$

The amount of muscular pulses is described by the log detector which uses a threshold to avoid noisy samples.

$$MYOP = \frac{1}{N} \sum_{n=1}^{N} f(|x_n|), f(x) = \begin{cases} 1, & x \geq th \\ 0, & otherwise \end{cases} \tag{11}$$

2.3. Autoregressive model features

A linear autoregressive model describes a random process using p coefficients [24]. The goal consists in extracting p coefficients to construct a representation of each sEMG sample $x[n]$ with the preceding signal values $(x[n-1], x[n-2], \ldots, x[n-p])$ making a linear combination, which carries an error or white noise term:

$$x[n] = \sum_{k=1}^{p} a_k x[n-k] + e[n] \tag{12}$$

where $x[n]$ is the generated sEMG value through k earlier samples $x[n-k]$, p is the order of the model, $e[n]$ expresses an added error or white noise term and a_k is the autoregressive coefficients. The mathematical approach used to derive the autoregressive coefficients defines the regressive model type. The most popular autoregressive model is the Yule-Walker model that uses the estimated values of the correlation function calculated as:

$$\hat{r}_{xx}(n, n-k) = Ex(n)x(n-k)k = 0, \pm 1, \pm 2, \ldots \tag{13}$$

Having $\hat{r}_{xx}$ estimated with (13), a NxN squared Yule-Walker matrix equation is built as follows:

$$\begin{bmatrix} \hat{r}_{xx}(0) & \hat{r}_{xx}(-1) & \cdots & \hat{r}_{xx}(-p) \\ \hat{r}_{xx}(1) & \hat{r}_{xx}(0) & \cdots & \hat{r}_{xx}(-p+1) \\ \vdots & \vdots & \cdots & \vdots \\ \hat{r}_{xx}(p) & \hat{r}_{xx}(p-1) & \vdots & \hat{r}_{xx}(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} \sigma_w^2 \\ \hat{r}_{xx}(1) \\ \vdots \\ \hat{r}_{xx}(p) \end{bmatrix} \tag{14}$$

where σ_w^2 is the variance of the modelled stochastic process. As the correlation matrix describes an equation system and fulfils Toeplitz definition, the method use the recursive Levinson-Durbin algorithm to get the autoregressive coefficients a_p.

Following a different approach, the Burg maximal entropy method, in [24] and [25] proposes the expansion of $\hat{r}_{xx}$, adding $\hat{r}_{xx}(p+1), \hat{r}_{xx}(p+2), \hat{r}_{xx}(p+3)$,... With this consideration in mind, the method extrapolate the new correlation values, maximising the entropy between them, so their randomness is high. The extrapolation of autoregressive series changes the predictions of backward and forward signal values $\hat{x}(n)$ and $\hat{x}(n-m)$:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0647.v1

Peer-reviewed version available at Sensors 2019, 19, 475; doi:10.3390/s19030475

$$\hat{x}(n) = \sum_{k=1}^{m} a_m(k)x[n-k], 0 \leq k \leq m-1, m = 1,2,...p \tag{15}$$

$$\hat{x}(n-m) = -\sum_{k=1}^{m} (a_m)^*(k)x(n+k-m) 0 \leq k \leq m-1, m = 1,2,...,p \tag{16}$$

where $a_m(k)$ is k-th autocorrelation coefficient of the model of order m, which implies a combination of previous values and the reflection coefficients K_m [24]:

$$a_m(k) = a_{m-1}(k) + k_m(a_{m-1})^*(m-k), 1 \leq k \leq m-1, 1 \leq m \leq p \tag{17}$$

The Burg proposal produces good results for different distributions; when the stochastic process has a Gaussian distribution, both autoregressive methods yield the same coefficient values [24].

2.4. Dataset construction

To develop and test the proposed approach for hand movement classification, the features described above are first extracted from different channels $S_{Ch1}[n],\ldots,S_{Chk}[n]$ and placed in a dataset. Each feature extractor (see Figure 2) forms single or multiple features and its output is a vector that represents a pattern of the form $P_{Ch_k}[n] = [feature_1chk, feature_2chk,\ldots,feature_Nchk]$. Next, the derived features from the channels are transferred to the pattern builder that concatenates the instances to generate an object containing the extracted features, and a label assigned for the class instances $P_i[n] = [P_{Ch1}[n], P_{Ch2}[n],\ldots,P_{Chk}[n], class_{label}]$. Figure 2 shows the dataset building block diagram.

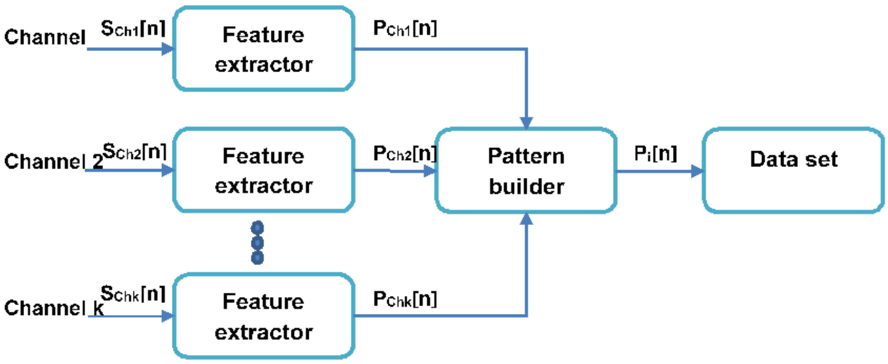

Figure 2. Data set building block diagram.

2.5. Proposed classification methodology

2.5.1. Burg reflection coefficients

As mentioned in Section 2.3, Burg autoregressive model introduces the forward and backward prediction errors

$$f_m(n) = x(n) - \hat{x}(n), b_m(n) = x(n-m) - \hat{x}(n-m) \tag{18}$$

These errors are defined by the following recursive renovation equations of lattice linear prediction filter

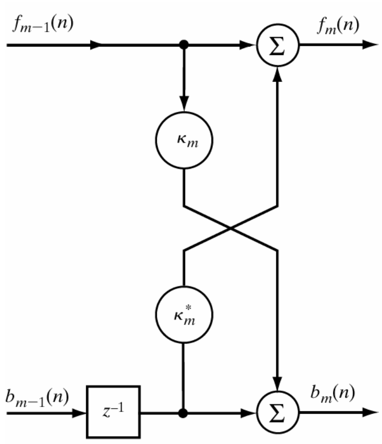

Figure 3. Lattice filter prediction cascade diagram.

$$
\begin{aligned}
f_0(n) &= b_0(n) = x(n), \\
f_m(n) &= f_{m-1}(n) + K_m b_{m-1}(n-1) m = 1, 2, \ldots, p, \\
b_m(n) &= K_m^* f_{m-1}(n) + b_{m-1}(n-1) m = 1, 2, \ldots p
\end{aligned}
\tag{19}
$$

The least squared error is

$$
\varepsilon_m = \sum_{n=m}^{N-1} [|f_m(n)|^2 + |b_m(n)|^2] \tag{20}
$$

Minimizing expression 20, the reflection coefficients are obtained [24]:

$$
K_m^* = \frac{-\sum_{n=m+1}^{N-1} f_{m-1}(n) b_{m-1}^*(n-1)}{\frac{1}{2}\sum_{n=m+1}^{N-1}[|f_m(n)|t^2 + |b_m(n)|^2]}, m = 1, 2, \ldots, p \tag{21}
$$

Reflection coefficients are the harmonic mean value of backward and forward error coefficient cross correlation. The numerator is the cross correlation of the prediction errors and the denominator is the smallest square estimation of these errors, so, $|K_m| \leq 1$. The reflection coefficients (21) are computed iteratively through the signal values; this is the reason they are proposed as features for classification tasks as their complexity is linear, and there is no evidence of their usage in such tasks.

2.5.2. Classification model training

Classification models involved in this research work are Bayesian, K nearest neighbor, multilayer perceptron, decision trees and support vector machines with different kernels. These classifiers are available in machine learning tool WEKA [26] and were taken with the purpose of evaluating their performance using different sEMG features. For the training phase, the following three datasets were generated comprising 900 instances and 10 traits per channel:

- Time domain Data set (1)-(11): TD=[$IEMG\ MAV\ SSI\ VAR\ RMS\ WL\ WAMP\ SSC\ ZC\ MYOP$]
- Burg autoregressive coefficients (17): Arb=[$Arb_1\ Arb_2 \ldots\ Arb_n$]
- Reflection coefficients: K=[$K_1\ K_2 \ldots\ K_n$]

The classification algorithms were trained once, and the performance was obtained by K-fold cross validation with K value of 10 because they widely use it in state-of-the-art related works, and the datasets lack of class unbalance. Moreover, each instance takes part in the training and testing set for a single run of the learning algorithm. Burg autoregressive coefficients (17) were chosen instead of Yule-Walker autoregressive coefficients (14) because different distributions of the Burg model produces a more accurate approximation [24]. After classifying three main datasets, K, Arb and TD features were joined into a new dataset with patterns of the form of $X = [IEMG\ MAV\ SSI\ VAR\ RMS\ WL\ WAMP\ SSC\ ZC\ MYOP\ Arb_1\ Arb_2 \ldots\ Arb_n\ K_1\ K_2 \ldots\ K_n]$; in order to evaluate how the interaction between these different features is reflected in classification model performance.

2.5.3. Feature selection

Under the same validation method (k-fold cross validation), taking the dataset with instances in the form of X, WEKA principal components (PC) and subset evaluation (SE) feature selection models [26] applied to reduce the pattern dimensionality and select the less redundant and correlated attributes for the classification task. Feature selection guarantees a reduction in dimensionality with or without a degradation of the classifier model performance. As observed in Section 2, some time domain features rely on sEMG amplitude such as RMS and MAV, and others depend on counting a certain event according to a threshold value. As a result, these attributes might contain a redundancy; also, the Burg autoregressive coefficients and reflection coefficients are tied according to (17). Sequential forward selection [27] strategy comprising taking just one of the different features used for dataset construction. Another feature selection criterion is plus l – take away r algorithm - [27], based on taking l traits from TD, K and Arb from X and remove the remaining r features, in such a way that the classification performance remains high. If the exclusion of an attribute causes a lower than previous performance, the removed sEMG characteristic is returned to the dataset because it is helpful for the class instance assignation. This process is repeated until the dimensionality cannot be reduced without affecting the classification rate.

3. Results

3.1. Datasets classification

Here, the results of the hand movement classification with separated datasets are presented. The parameter changed in SVM model was kernel function. For the rest of models, the default Weka parameters were not modified. Two values chosen for k in IBk are $k = 1$ and $k = number$ of classes +1; assigning $k = 6$ would cause a tie between six classes; the extra value will establish the majority class. Table 1 describes the results obtained by classifying the hand movements with separated datasets.

Table 1. Classification results of TD, Arb and K datasets separately.

N	features	Bayes		Ik-P 2.0		MLP	Tree		SVM		
–	–	*Naive*	*net*	k_1	k_7	–	*J48*	*Random*	*radial*	*linear*	*P3*
20	*TD*	46.55	68.33	93.22	91.22	86.55	83.33	92.88	27.33	76.33	93.11
20	*Arb*	59.11	63.33	90.77	91.66	86.33	78.00	90.55	47.00	61.11	37.77
20	*k*	57.44	71.11	93.55	93.33	86.44	74.77	92.00	45.55	65.44	29.77

One can observe from Table 1 that TD dataset and the SVM with third order polynomial kernel (P3 column) gives better decision borders than other kernels. The radial kernel yields the lowest performance with Bayesian models; whereas the remaining models (trees and MLP) reached high classification rates, IBk with $k = 1$ obtained the highest. The dataset built only with maximal entropy autoregressive coefficients (Arb column) is more appropriate to classify with the IBk considering seven neighbours more than just one. Other learning algorithms such as MLP and decision trees offer competitive classification rates (78% - 90.77%); Bayesian models overcome linear kernel support vector machines with a performance of 63.33%. The best performance among the different datasets is obtained using the reflection coefficient dataset K of the reflection coefficients, classifying 93.55% of instances using IBk $k = 1$. Despite Bayesian models and SVM still having a low performance, an improvement in Bayesian net is achieved using the reflection coefficients.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 October 2018 doi:10.20944/preprints201810.0647.v1

Peer-reviewed version available at *Sensors* **2019**, *19*, 475; doi:10.3390/s19030475

Table 2. Classification performance of combined datasets.

N	features	Bayes		Ik-P 2.0		MLP		Tree		SVM	
–	–	*Naive*	*net*	k_1	k_7	–	*J48*	*Random*	*radial*	*linear*	*P3*
40	$k+TD$	61.44	83.22	99.88	99.55	98.44	87.11	98.33	18.33	83.11	93.22
40	$k+Arb$	78.22	82.33	99.66	99.33	98.44	84.77	98.66	64.00	90.00	54.66
60	X	76.00	83.33	100.0	99.77	99.11	85.55	95.55	17.77	83.00	93.22

In Table 2, by merging reflection coefficients and TD features for the training phase (K+TD), most of the classifiers reach high performance, excluding naïve Bayes and radial kernel SVM. With a 0.22% classification error, IBk $k = 1$ got the best classification rate above the following models: IBk $k = 7$, MLP, decision trees and 3*th* grade polynomial kernel SVM. The resulting dataset of joining the Burg maximal entropy reflection coefficients, K, and the Burg autoregressive coefficients, Arb, yield patterns that are best classified by the IBk model with k value of one, slightly above the IBk using $k = 7$, MLP and random forest. The $J48$ decision tree and Bayesian models offer high performances and are below 90% accuracy reached by linear kernel SVM; the other kernels have the lowest classification rates. The combination of all features (TD, Arb, and K) results in a sixty-dimension feature vector, useful to classify correctly all 900 dataset instances using the IBk with $k = 1$; increasing the number of neighbors to $k = 7$ decreases the classification rate, but keeps on above the following competitive models, MLP and random forest. Data distribution does not fit to a radial kernel, therefore, the SVM outputs the lowest accuracy.

3.2. Feature selection classification performance

This subsection describes the results of dimensionality reduction trying to reach a higher classification performance with as fewer traits as possible. After running the feature selection algorithms, all instances kept on being classified correctly with 20 and 26 features with the nearest neighbours and support vector machine models, respectively (see Table 3). SE traits exclude amplitude

Table 3. Classification performances using different feature vectors after feature selection: SE=[Ch_1Arb(1,2,5,9,10), ch_1k(1,2,3,4,9), Ch_1WL, Ch_1SSC, Ch_1ZC, Ch_1MYOP, Ch_2Arb(1,2,4,5,10), Ch_2k(1,5,7), Ch_2WL, Ch_2MYOP], FS_1=[ARB(1,2,7,8,10), K(1,2,10),ZCC,MYOP], FS_2=[Arb1, K1, ZCC, RMS], FS_3=[Arb_1, K_1, ZCC, MAV], FS_4=[Arb_1, K_1, MYOP, RMS] FS_5=[Arb_1, K_1, MYOP, MAV].

N	features	Bayes		I-P 2.0		MLP		Tree		SVM	
–	–	*Naive*	*net*	k_1	k_7	–	*J48*	*Random*	*radial*	*linear*	*P3*
26	*SE*	82.55	91.55	99.88	99.88	99.11	87.77	99.55	18.00	74.33	82.00
26	*PC*	88.11	88.66	99.66	98.88	97.55	84.11	98.44	100.0	96.11	99.55
20	FS_1	79.88	86.88	100.0	99.66	97.66	86.33	98.77	24.88	68.88	57.77
8	FS_2	64.88	84.22	99.22	99.22	88.55	88.66	97.88	22.33	55.11	49.88
8	FS_3	65.00	83.22	99.22	98.77	89.22	89.55	97.88	22.33	54.33	45.88
8	FS_4	66.77	84.66	98.22	97.77	87.55	88.77	89.88	66.11	70.77	28.66
8	FS_5	67.33	83.00	98.22	97.77	88.55	88.55	98.33	61.44	68.66	32.55

related values such as RMS, MAV and so on. They provide high classification rates with the exception of the linear kernel SVM. The PC dataset was built with the combination of the less correlated features resulting in a more uniform performance through all tested classifiers, reaching 100% performance using the SVM with a radial kernel. The FS_1 represents the result of the feature selection process; the IBk with k=1 is still the highest and SVM with any kernel, the lowest. Datasets FS_2, FS_3, FS_4, FS_5 were obtained taking the first reflection and autoregressive coefficients in combination with MYOP, ZCC, RMS, MAV. These sets are of low pattern dimensionality and have high performance from 83% up to 99.22% using the IBk with $k = 1$(the highest), MLP, decision trees and Bayes net (the lowest); the remaining models yield a low performance with just 8 attributes.

Peer-reviewed version available at Sensors 2019, 19, 475; doi:10.3390/s19030475

4. Discussion

The classification accuracy rate of different learning algorithms depends on the data distribution. This behaviour is expected according to the "no-free-lunch theorems" [28], which state that the best classification model for all datasets does not exist. The justification of why several models have to be compared using the same dataset concurs with that statement.

While testing classifiers with separated datasets, Burg reflection coefficients K (21) are more appropriate to be used in conjunction with the Bayes net and IBk models. Since K traits are needed to compute the Burg autoregressive traits, Arb (17), their classification performances are similar. Besides, TD features have different values which rely on amplitude or counting events. Therefore, the MLP, decision trees, and linear and polynomial kernel SVM outputs have high accuracy.

Despite the Burg maximal entropy autoregressive and reflection coefficients being closely related, the classification rate increases when these two traits take a part of the dataset. This means that they are not redundant or irrelevant in feature construction tasks; however, as mentioned before, Arb characteristics take more time to be computed because they rely on first computing the reflection coefficients. In addition, as a result of combining three different theoretical frameworks attributes, the highest classification rate is obtained. Hence, a synergy of different attributes is needed for a higher accuracy in the sEMG classification tasks.

In Section 3.2, opposite to the results in Table 2, a high classification performance is reached with 26 or fewer features. The fact that the feature vector X has at least 34 redundant or irrelevant traits, which are necessary to be removed using feature selection tools can explain it. As a result, different data distributions are obtained; for instance, principal component analysis performed by WEKA produces a data distribution suitable for the support vector machine with a radial kernel since all patterns are correctly classified.

Excluding the PC dataset, the best classifier for the remaining datasets is the IBk model with a different number of neighbours used for class assignment, because the performance of the nearest neighbors-based model depends on k value. For instance, TD and K dataset classification rate decreases as a result of increasing k value, contrasting to what happens with Arb patterns (see Table 1). This behaviour is expected due to the classification phase applying the nearest neighbour to different classes causing in the worst case a tie and a misclassified pattern.

The IBk advantage is its simple training phase that bases directly on the dataset compared to Bayesian models that require the computation of probability distributions and cost function. The design of the hidden layer of MLP can be complex and the time of back propagation error reducing might be long. The decision trees as well as multilayer perceptron are hard to design, and they require a pruning process to reduce irrelevant leaves and branches and the SVM has $O(n_3)$ complexity to establish support vectors. The kernel selection and design of the support vectors that best fits the data distribution are needed.

An example of kernel selection can be seen from Table 1 with the TD dataset: using the polynomial kernel, a high performance of 93.11% is reached, and merging the TD with K features an improvement of 0.11% is achieved. This is a sign that most of the support vectors are found in the TD traits using such a kernel. Another example is Arb and K traits: they are better classified with a linear kernel, and by joining these two datasets, a considerable increase in the classification rate is obtained (see Table 2) as a consequence of more appropriate data for the support vector estimation.

The results that are considered for the discussion are those that was obtained using the same signal dataset; otherwise, any comparison concerning the feature extraction and preprocessing stage based on the classification performance would be unfair.

The feature extraction method presented in [22] yields fifty traits and reaches a high performance (98.22%) with no previous signal treatment while the classification rates of 89.21% are obtained with the empirical mode decomposition technique, which denoises the original signal in conjunction with sixty four features proposed in [23]. Our methodology succeeded in classifying all 900 sEMG instances with less than half the features (20 characteristics) than [22] and [23] methods. Their results are exceeded with just 8 traits producing the 99.22% correctly classified signals. Our preprocessing stage consists only in the reduction of noisy samples and the subtraction of the myoelectric signal mean value, which turns out to be effective due to it giving signals suitable to the framework of the developed feature extraction tools, such as Burg reflection coefficients (21).

As shown in [23], decomposing sEMG causes an information loss that is reflected in the classification rate; for this reason, all available signal information for the feature extraction task were taken.

5. Conclusions

The results of hand movement classification of myoelectric signals using different traits have been presented. The Burg linear autoregressive model yields to the reflection coefficients that have been shown to be useful for the sEMG classification; these traits have shown a higher performance that is increased with the standard time domain features and are less complex to compute than the autoregressive coefficients.

A high classification rate has been reached using a simple preprocessing stage, which fits the signals to the theory framework in the feature extraction tools. Low pattern dimensionality results from removing redundant traits and building patterns with uncorrelated features such as Burg maximal entropy reflection and autoregressive coefficients and considering different time domain features related to the signal energy and to event counting. Despite a lower feature dimensionality, the classification rate up to 100% has been achieved separately and with other kinds of uncorrelated attributes for the hand movement identification.

The applications of the classification technique of signals presented in this work can lead to the possible development of state-of-the-art active prosthetic devices, where the myoelectric classification does not represent a challenge and the implementation of the classification algorithm can be performed in an integrated low cost device.

Acknowledgments:

This work was partially supported by the Instituto Politécnico Nacional and CONACyT.

Conflicts of Interest:

The authors declare no conflict of interest.

References

1. De Luca, C.J. Electromyography. In *Encyclopedia of Medical Devices and Instrumentation, Vols 1-4*; John Wiley & Sons, Inc., 2006. doi:10.1002/0471732877.
2. Raez, M.B.I.; Hussain, M.S.; Mohd-Yasin, F.; Reaz, M.; Hussain, M.S.; Mohd-Yasin, F. Techniques of EMG signal analysis: detection, processing, classification and applications. *Biological procedures online* **2006**. doi:10.1251/bpo115.
3. Lahmiri, S.; Boukadoum, M. Improved Electromyography Signal Modeling for Myopathy Detection. 2018 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE, 2018, pp. 1–4. doi:10.1109/ISCAS.2018.8350893.
4. Fuglsang-Frederiksen, A. The role of different EMG methods in evaluating myopathy, 2006. doi:10.1016/j.clinph.2005.12.018.
5. Dostál, O.; Vysata, O.; Pazdera, L.; Procházka, A.; Kopal, J.; Kuchyňka, J.; Vališ, M. Permutation entropy and signal energy increase the accuracy of neuropathic change detection in needle EMG. *Computational Intelligence and Neuroscience* **2018**. doi:10.1155/2018/5276161.

6. Cler, M.J.; Stepp, C.E. Discrete Versus Continuous Mapping of Facial Electromyography for Human-Machine Interface Control: Performance and Training Effects. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **2015**. doi:10.1109/TNSRE.2015.2391054.
7. Brunelli, D.; Tadesse, A.M.; Vodermayer, B.; Nowak, M.; Castellini, C. Low-cost wearable multichannel surface EMG acquisition for prosthetic hand control. 2015 6th International Workshop on Advances in Sensors and Interfaces (IWASI). IEEE, 2015, pp. 94–99. doi:10.1109/IWASI.2015.7184964.
8. Copaci, D.; Serrano, D.; Moreno, L.; Blanco, D.; Copaci, D.; Serrano, D.; Moreno, L.; Blanco, D. A High-Level Control Algorithm Based on sEMG Signalling for an Elbow Joint SMA Exoskeleton. *Sensors* **2018**, *18*, 2522. doi:10.3390/s18082522.
9. Repnik, E.; Puh, U.; Goljar, N.; Munih, M.; Mihelj, M.; Repnik, E.; Puh, U.; Goljar, N.; Munih, M.; Mihelj, M. Using Inertial Measurement Units and Electromyography to Quantify Movement during Action Research Arm Test Execution. *Sensors* **2018**, *18*, 2767. doi:10.3390/s18092767.
10. Merletti, R.; Farina, D. *Surface Electromyography: Physiology, Engineering and Applications*; John Wiley & Sons, 2016. doi:10.1002/9781119082934.
11. Hsueh, Y.H.; Yin, C.; Chen, Y.H. Hardware System for Real-Time EMG Signal Acquisition and Separation Processing during Electrical Stimulation. *Journal of Medical Systems* **2015**, *39*, 88. doi:10.1007/s10916-015-0267-6.
12. Resnik, L.; Huang, H.; Winslow, A.; Crouch, D.L.; Zhang, F.; Wolk, N. Evaluation of EMG pattern recognition for upper limb prosthesis control: a case study in comparison with direct myoelectric control. *Journal of NeuroEngineering and Rehabilitation* **2018**, *15*, 23. doi:10.1186/s12984-018-0361-3.
13. Venugopal, G.; Navaneethakrishna, M.; Ramakrishnan, S. Extraction and analysis of multiple time window features associated with muscle fatigue conditions using sEMG signals. *Expert Systems with Applications* **2014**. doi:10.1016/j.eswa.2013.11.009.
14. Nazmi, N.; Abdul Rahman, M.; Yamamoto, S.I.; Ahmad, S.; Zamzuri, H.; Mazlan, S. A Review of Classification Techniques of EMG Signals during Isotonic and Isometric Contractions. *Sensors* **2016**, *16*, 1304. doi:10.3390/s16081304.
15. Chowdhury, R.H.; Reaz, M.B.I.; Ali, M.A.B.M.; Bakar, A.A.A.; Chellappan, K.; Chang, T.G. Surface electromyography signal processing and classification techniques. *Sensors (Basel, Switzerland)* **2013**. doi:10.3390/s130912431.
16. Bozkurt, M.R.; Subaşi, A.; Köklükaya, E.; Yilmaz, M. Comparison of AR parametric methods with subspace-based methods for EMG signal classification using stand-alone and merged neural network models. *Turkish Journal of Electrical Engineering and Computer Sciences* **2016**. doi:10.3906/elk-1309-1.
17. Phinyomark, A.; Limsakul, C.; Phukpattaranont, P. Application of wavelet analysis in EMG feature extraction for pattern classification. *Measurement Science Review* **2011**. doi:10.2478/v10048-011-0009-y.
18. Veer, K.; Sharma, T. A novel feature extraction for robust EMG pattern recognition. *Journal of Medical Engineering & Technology* **2016**, *40*, 149–154. doi:10.3109/03091902.2016.1153739.
19. Liu, Y.H.; Huang, H.P.; Weng, C.H. Recognition of Electromyographic Signals Using Cascaded Kernel Learning Machine. *IEEE/ASME Transactions on Mechatronics* **2007**. doi:10.1109/TMECH.2007.897253.
20. Al-Angari, H.M.; Kanitz, G.; Tarantino, S.; Cipriani, C. Distance and mutual information methods for EMG feature and channel subset selection for classification of hand movements. *Biomedical Signal Processing and Control* **2016**. doi:10.1016/j.bspc.2016.01.011.
21. Khezri, M.; Jahed, M. A Neuro–Fuzzy Inference System for sEMG-Based Identification of Hand Motion Commands. *IEEE Transactions on Industrial Electronics* **2011**, *58*, 1952–1960. doi:10.1109/TIE.2010.2053334.
22. Ruangpaisarn, Y.; Jaiyen, S. SEMG signal classification using SMO algorithm and singular value decomposition. Proceedings - 2015 7th International Conference on Information Technology and Electrical Engineering: Envisioning the Trend of Computer, Information and Engineering, ICITEE 2015, 2015. doi:10.1109/ICITEED.2015.7408910.
23. Sapsanis, C.; Georgoulas, G.; Tzes, A.; Lymberopoulos, D. Improving EMG based classification of basic hand movements using EMD. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2013. doi:10.1109/EMBC.2013.6610858.
24. Haykin, S.S. *Adaptive filter theory*; Pearson, 2014. doi:0130901261.
25. Burg, J.P. Maximum entropy spectral analysis. PhD thesis, Stanford, 1975. doi:doi.wiley.com/ 10.1002/ 0471667196.ess1570.

26. Witten, I.H.; Frank, E.; Trigg, L.; Hall, M.; Holmes, G.; Cunningham, S.J. Weka : Practical Machine Learning Tools and Techniques with Java Implementations. Proceedings of the ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information Systems, 1999, pp. 192–196. doi:10.1.1.16.949.
27. Ciaccio, E.J.; Dunn, S.M.; Akay, M. Biosignal Pattern Recognition and Interpretation Systems. *IEEE Engineering in Medicine and Biology Magazine* **1994**. doi:10.1109/51.265792.
28. Wolpert, D.H.; Macready, W.G. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* **1997**. doi:10.1109/4235.585893.

Sample Availability: Samples of the compounds ...... are available from the authors.