REDUCING THE NUMBER OF CHANNELS AND SIGNAL-FEATURES FOR AN ACCURATE CLASSIFICATION IN AN EMG PATTERN RECOGNITION TASK

Iker Mesa, Angel Rubio, Javier Diaz, Jon Legarda and Beatriz Sedano
CEIT, Paseo Mikeletegi, N 48, 20018 San Sebastian, Spain

Keywords: EMG, sEMG, mRMR, SVM, Pattern recognition, Variable selection, Feature selection.

Abstract: In this work 32 surface Electromyography (sEMG) electrode locations and 41 signal-features are evaluated in order to achieve an accurate classification rate in a static-hand gesture classification task. A novel implementation of the minimal Redundancy Maximal Relevance (mRMR) Variable Selection algorithm is proposed with the aim of selecting the most informative and least redundant combination of sEMG channels and signal features. The performance of the new algorithm and of the selected set of channels and signal-features are tested with a Support Vector Machine classifier.

1 INTRODUCTION

Surface Electromyography (sEMG) is a noninvasive technique to measure the electrical activity on the skin produced by the muscles beneath it. With disposable adhesive electrodes (channels) the sum of multiple Motor Unit Action Potentials (MUAP) can be easily captured. The interpretation of this electrophysiological activity leads to diverse applications. The most common one is to diagnose neuromuscular disorders like dystrophy, tremor or nocturnal bruxism (Merletti and Parker, 2004) (Fuglsang-Frederiksen and Pugdahl, 2010). In addition, sEMG is being largely used in many other different research areas such as exoskeletons (Khokhar et al., 2010) or powered upper-limb prostheses (Tenore et al., 2007), (Kuiken et al., 2007).

The aforementioned applications are based on the following scheme: A) recording the sEMG signals from several channels; B) extraction of features from the signals to represent the recordings using a vector of variables; C) application of a dimensionality reduction technique to the vector; and finally D) classification of the vector of variables in one of K possible classes. The meaning of each of these classes depends on the application. For example, the classes could be the different gestures that an amputee wants to do with his prosthesis (Tenore et al., 2009); or the kind of tremor that a patient suffers (Palmes et al., 2010).

The classification of step (D) is done by machine learning algorithms such as Neural Networks (Tenore et al., 2009) or Support Vector Machines (Palmes et al., 2010). Naturally, supervised training is needed; using training samples (examples of vectors) the classifier learns how to classify them into the different classes. Next, the generalization properties of the classifier can be tested with validation samples (cross-validation).

In this work, the classes represent different hand static-gestures among K different ones. In addition, we use the term *trial* for the set of sEMG signals recorded during a period of time in which the user performed a static-gesture k; the term *feature* for the parameter obtained after a given signal-processing is applied to a signal; and the term *variable*, denoted as z, for the pair (channel, signal-feature). This means that a variable refers to a specific signal-feature extracted from a specific channel. In addition, we denote to the set of all possible variables as Z. After the feature extraction step (B), a trial is represented by a vector of V variables and it is denoted as $\{z_j\}_{j=1}^{V}$.

The dimensionality reduction step (C) is recommended since the higher the dimension V of the vector of variables $\{z_j\}_{j=1}^{V}$, the more training samples are needed in order to sufficiently train the classifier. This is known as *the curse of dimensionality* (Bellman, 1961) or as *the Hughes phenomenon* (Hughes, 1968). However, diminishing the number of variables may lead to a loss in the discrimination power, and hence a worse classification performance (Jain and

Mesa I., Rubio A., Diaz J., Legarda J. and Sedano B..
REDUCING THE NUMBER OF CHANNELS AND SIGNAL-FEATURES FOR AN ACCURATE CLASSIFICATION IN AN EMG PATTERN RECOGNITION TASK.
DOI: 10.5220/0003767300380048
In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing* (BIOSIGNALS-2012), pages 38-48
ISBN: 978-989-8425-89-8

Duin, 2000). In practice, if we have a limited training set it is better to select only a low number of powerful variables. Variable z_j is more powerful than variable z_h, with $z_j, z_h \in Z$, if the patterns from the different classes can be classified easier with z_j than with z_h.

There are two main dimension reduction techniques: Variable Selection and Variable Projection. On the one hand, the first method aims to determine a subset of variables that brings together the most important information of the whole set. On the other hand, the projection techniques transforms the data to a space of fewer dimensions by linear or non-linear combinations of the original variables. As the Variable Selection techniques do not alter the original representation of the variables, but merely select a subset of them, they provide a direct interpretation possibility. Specially, the Variable Selection filter methods rank the variables according to an evaluation function that relies solely on properties of the data. There is active research in this area related to the sEMG classification problem; (Zardoshti-Kermani et al., 1995) and (Boostani and Moradi, 2003) rank several signal-features according to Davies-Bouldin or Scattering criteria without taking into consideration the relation among the signal-features or the channels. However, it is known in the Variable Selection field that the *m* best independent variables are not the best *m* variables (Cover, 1974), (Jain and Duin, 2000). The reason for this is the redundancy. The combination of two uncorrelated variables could be more informative than the combination of two correlated variables with the best individual properties.

The objective of this research is to reduce the necessary number of the sEMG channels and signal-features present in an sEMG pattern recognition system. There are two main reasons for this reduction. Firstly, if we reduce the number of physical sEMG channels we make the sEMG-recording device simpler, thus cheaper. Secondly, if we diminish the number of features and channels we reduce the dimension V of the vector $\{z_j\}_{j=1}^{V}$ presented to the classifier, hence, less training examples are needed and, furthermore, the classifier requires less memory and computational power.

In this work we propose a novel multivariate filter method, based on the *minimal Redundancy maximal Relevance* algorithm (mRMR) (Peng et al., 2005), to rank the variables with the aim of finding the most informative channels and signal-features and, at the same time, the least redundant. In addition to the aforementioned advantages, the reduction of the required number of sEMG channels can also be seen as an analysis of the best electrode placement for this kind of applications.

For the accomplishment of the objectives, the following procedure is used. Firstly, sEMG data is obtained from a static-hand gesture experiment performed by 6 volunteers. After the corresponding data-preprocessing and feature-extraction steps, the variables are ranked with three methods: A) a simple univariate filter method, the F-statistic, B) the regular mRMR algorithm, C) a novel implementation of the mRMR to penalize even more the redundancy. Therefore, three different ranking lists of variables are obtained for each user.

Secondly, a Support Vector Machine (SVM) classifier is used to classify the hand gestures with the top-ranked variables of each list. This way, the performance of each ranking method can be analyzed. An SVM is used since it performs well in many different fields and it is insensitive to overtraining (Jain and Duin, 2000) (i.e. a degradation in generalization properties due to an excessive number of training samples).

Thirdly, a study of the top-ranked variables is done to select the best combinations of features and channels. The aim, as mentioned before, is to reduce the number of variables. Finally, the selected variables are tested with the SVM classifier and the results are discussed.

2 METHODS

2.1 Data Acquisition

Six healthy normal-limbed subjects volunteered for the experiment. None of them reported any muscular or skin disorders. The following protocol was used during each session.

After a careful skin preparation, 32 Ag/AgCl disposable electrodes (diameter 10 mm) are equidistantly placed on each subject's right forearm in 4 rows of 8 electrodes each. The reference and ground electrodes are placed on the shoulder and wrist respectively. The layout of the electrodes is described in Figure 1. To ensure an adequate position, a flexible armband with 32 small guide points is used.

The user sits comfortably in front of a computer screen and a webcam. With the elbow resting on a table the subject can perform hand movements without constraints. For the experiment, $K = 15$ static-gestures were selected from the Spanish Sign Language Alphabet (letters A, B, C, D, E, F, I, K, L, M N, O, P, Q and U) as they involve a wide variability of wrist and finger positions.

There are two computers involved in the acquisition, namely, one for guiding the user and watching

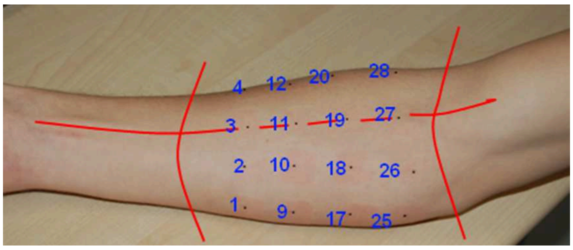


Figure 1: The electrode naming and positioning scheme.

the correct execution of the gestures; and another one for recording the sEMG signals. Each participant performs between 3 and 5 runs of 50 gestures each. A computer program in the first machine is in charge of: A) asking the subject to perform a specific gesture by displaying an aleatory gesture on the screen, B) ensuring with the webcam that the user is maintaining a steady gesture for ~3 seconds, C) informing the second PC about the gesture that the user is asked to do and if the subject has performed it correctly (steady) or not (not steady).

The signals are recorded with a sampling frequency of 1000 Hz by a BrainAmp Standard amplifier connected to the second PC.

2.2 Data Preprocessing

The signals are high-pass filtered (10 Hz) to remove movement artifacts, and a notch filter is used to remove the 50 Hz band (power-line noise). These operations are performed by 4^{th} order Butterworth filters.

Next, the correctly performed gestures of each subject are arbitrarily divided into 3 datasets: Training-dataset, Validation-dataset and Test-dataset. The first dataset is used for training the classifier as it is explained in Section 2.5, the Validation-dataset is used in the same section to compare the different Variable Selection techniques, and finally, the Test-dataset is used in Section 4 to check the performance of the final selection of variables.

Each of these performed gestures has a duration of ~3 seconds. To increase the size of the datasets they are segmented into trials of 256 ms with an overlapping of 50% (Englehart et al., 2003). Consequently, the number of available trials in each dataset is described in Table 1. Each of these trials contains the 32 sEMG signals (one per channel) recorded during a 256 ms period. The static-gesture that the user was performing during the recording of each trial is known.

2.3 Feature Extraction

Let x_i be the i-th sample of sEMG signal x, and let N be the signal length. From the sEMG signals of each of the trials obtained in Section 2.2, the following 41 features are extracted. Note that each trial has 32 sEMG signals, hence, a vector of $V = 1312$ variables, $\{z_j\}_{j=1}^{1312}$, is obtained for each trial.

Table 1: The number of trials in each dataset. Each trial contains the 32 sEMG signals of 256ms long. The static-gesture corresponding to each trial is known.

Subject	Train-dataset	Validation-dataset	Test-dataset
S1	4929	1232	0
S2	6382	1595	0
S3	4787	1115	655
S4	5625	1446	964
S5	2218	0	555
S6	3871	968	0

Mean Absolute Value (MAV). It is a very common sEMG amplitude indicator which provides information about the muscle contraction level.

$$\mathrm{MAV} = \frac{1}{N}\sum_{i=1}^{N}|x_i| \tag{1}$$

Median Absolute Value (MedAV). Like MAV, the MedAV is an indicator of the contraction level. However, it is more robust against outliers.

$$\mathrm{MedAV} = \operatorname*{median}_{i}|x_i| \tag{2}$$

Variance (VAR). As the mean value of the sEMG signal is 0 the variance of the signal can also be seen as a measurement of the average signal power (Zardoshti-Kermani et al., 1995), (Boostani and Moradi, 2003).

$$\mathrm{VAR} = \frac{1}{N-1}\sum_{i=1}^{N}x_i^2 \tag{3}$$

Waveform Length (WL). It is the cumulative length of the waveform over the segment. The resultant value is an indicator of the waveform's amplitude, frequency, and duration (Hudgins et al., 1993).

$$\mathrm{WL} = \sum_{i=1}^{N-1}\Delta_i \tag{4}$$

Where $\Delta_i = |x_i - x_{i+1}|$

Mean Absolute Difference Value (MADV). The mean value of the difference in amplitude between adjacent samples.

$$\mathrm{MADV} = \frac{1}{N-1}\sum_{i=1}^{N-1}\Delta_i \tag{5}$$

Note that, as N is constant the MADV provides exactly the same information as WL. We compute both features to test the different Variable Selection methods, since a good method has to be able to find the correlation between both features and discard one of them. As it is explained in Section 3.1, there are some methods that are unable to do this.

Zero Crossing (ZC). It counts the number of times that the signal crosses the zero amplitude axis. To avoid signal crossing counts due to low-level noise, a threshold ε is included. In our case it is set to $\varepsilon = 10\mu V$ according to (Hudgins et al., 1993).

$$\mathrm{ZC} = \sum_{i=1}^{N-1} \phi(\Delta_i, x_i, x_{i+1}) \tag{6}$$

Where:

$$\phi(\Delta_i, x_i, x_{i+1}) = \begin{cases} 1 & x_i \cdot x_{i+1} < 0, \Delta_i > \varepsilon \\ 0 & \text{otherwise} \end{cases}$$

Number of Turns (NT). This feature counts the number of times that the slope of the waveform changes in sign. This means that it counts the number of local maxima and minima in the sEMG signal. To reduce noise effects, a slope change is only taken into account if the difference in amplitude with adjacent slope changes is at least $\varepsilon = 30\mu V$ (Hudgins et al., 1993).

Wilson Amplitude (WAMP). The Wilson amplitude counts the number of times that the difference in amplitude for any two consecutive samples exceeds a certain threshold ε. This ε is set to $50\mu V$ according to (Philipson, L, Larsson, P, 1988).

$$\mathrm{WAMP} = \sum_{i=1}^{N-1} \psi(\Delta_i) \tag{7}$$

Where $\psi(\Delta_i) = 1$ if $\Delta_i > \varepsilon$; and 0 otherwise.

Amplitude Histogram (A). This feature counts the number of times that the sEMG signal reaches 9 different amplitude levels. In other words, it is the voltage histogram of the signal into 9 equal bins (Zardoshti-Kermani et al., 1995). Therefore, 9 features are obtained per signal, named A1-A9.

Auto-Regressive Coefficients (AR). The AR model of the signal approximates the sample x_i as a linear combination of earlier samples plus an independent error term. The number of samples taken into account defines the order of the model. As well as (Tkach et al., 2010) and (Boostani and Moradi, 2003) an order of 4 is set.

$$x_i = \sum_{r=1}^{4} a_r x_{i-r} + \varepsilon_i \tag{8}$$

Where the a_r are the auto-regressive coefficients and ε_i the error term. The a_r coefficients are computed minimizing the squared error between the original signal and its approximation. Each coefficient a_r is used as a different feature, hence, 4 different features are obtained, denoted as AR1-AR4, for each signal.

Cepstral Coefficients (C). Cepstral analysis has been mainly applied to speech recognition but it has been also shown to be suitable for sEMG classification (Kang et al., 1995). The cepstral coefficients c_r are obtained from the Auto-Regressive coefficients a_r.

$$\begin{aligned} c_1 &= -a_1 \\ c_r &= -a_r - \textstyle\sum_{n=1}^{r-1} \left(1 - \frac{n}{r}\right) a_n c_{r-n} \end{aligned} \tag{9}$$

Where a_r is the r-th auto-regressive coefficient. Consequently, there are 4 cepstral features per signal and they are denoted as C1-C4.

As in the case of WL and MADV, there is no difference between the cepstral coefficient c_1 and the Auto-Regressive one a_1. Likewise, this fact plays an important role in the comparison of Variable Selection techniques as it is explained in Section 3.1.

Mean Frequency (Fmean). This feature measures the mean frequency of the signal's power spectrum. Assuming the frequency spectrum is divided into M frequencies, let f_b be the b-th frequency of the signal's spectrum and $P(f_b)$ the power of f_b. The mean frequency is computed as follows:

$$\mathrm{Fmean} = \frac{\sum_{b=1}^{M} f_b P(f_b)}{\sum_{b=1}^{M} P(f_b)} \tag{10}$$

Quantiles (Q). The quantiles mark the boundaries between specific consecutive subsets of the signal spectrum, e.g. the Qy quantile is the frequency that marks the upper boundary of the lower y% of the spectrum's power. The following frequency spectrum quantiles are considered: Q10, Q30, Q50, Q60, Q75, Q90. Therefore, 6 features are obtained per signal.

Frequency Histogram (F). The frequency spectrum is divided into 9 equal-size segments and the percentages of power in each of them are taken as features. Hence, 9 features are obtained per signal and are denoted as F1-F9.

2.4 Variable Ranking: mRMR Algorithm

After the feature extraction step (Section 2.3), each trial is represented by a vector of variables $\{z_j\}_{j=1}^{1312}$, and it belongs to a known class k (static-gesture). We can measure the discriminant power of each variable z_j using the F-statistic as follows:

$$F(z_j) = \frac{\left[\sum_{k=1}^{K} n_j^{(k)} (\overline{z_j^{(k)}} - \overline{z_j})^2\right] / (K-1)}{\left[\sum_{k=1}^{K} \sum_{t=1}^{n_j^{(k)}} (z_{j_t}^{(k)} - \overline{z_j^{(k)}})^2\right] / (n_j - K)} \tag{11}$$

Where:

- K : Number of possible classes.
- n_j : Number of samples of variable z_j.
- $n_j^{(k)}$: Number of samples of z_j within the k-th class.
- $\overline{z_j}$: Mean value of z_j.
- $\overline{z_j^{(k)}}$: Mean value of z_j within the k-th class.
- $z_{j_t}^{(k)}$: Sample t-th of z_j within the k-th class.

F-statistic only gives a measure of the classification power of each variable z_j by itself. However, the aim is to select a group S of m variables z_j which jointly have the largest relevance with the classification task. This is called *maximal relevance* criterion (Ding and Peng, 2005), and it has the following form:

$$\max D_F(S), \quad D_F = \frac{1}{|S|} \sum_{z_j \in S} F(z_j) \tag{12}$$

However, we also aim to ensure that the redundancy among the variables is as low as possible. Therefore, it is necessary to introduce the *minimal redundancy* condition. The minimal redundancy criterion based on the Pearson correlation coefficient (Ding and Peng, 2005) has the following form:

$$\min R_c(S), \quad R_c = \frac{1}{|S|^2} \sum_{z_j, z_h \in S} |c(z_j, z_h)| \tag{13}$$

The Pearson correlation coefficient $c(z_j, z_h)$ is computed as:

$$c(z_j, z_h) = \frac{\sum_{t=1}^{n_j} (z_{j_t} - \overline{z_j})(z_{h_t} - \overline{z_h})}{\sqrt{\sum_{t=1}^{n_j} (z_{j_t} - \overline{z_j})^2} \sqrt{\sum_{t=1}^{n_j} (z_{h_t} - \overline{z_h})^2}} \tag{14}$$

where z_{j_t} is the sample t-th of z_j.

The minimal redundancy - maximal relevance criterion (mRMR) (Ding and Peng, 2005) (Peng et al., 2005) combines (12) and (13). The optimization of these two conditions can be done in several different ways. One commonly used combination criteria is the following:

$$\max\,(D_F / R_c) \tag{15}$$

A near optimal solution of the mRMR is obtained with the following iterative method (Ding and Peng, 2005). The first variable is selected according to (12), i.e. the variable with the highest $F(z_j)$. Next, the rest of variables are added to the set S in the following incremental way. Assuming that we have already chosen $m-1$ variables, we select the m^{th} variable from the set $\{Z - S_{m-1}\}$ (i.e. all the variables except those already selected) that combines the following conditions:

$$\max_{z_j \in Z - S_{m-1}} F(z_j) \tag{16}$$

$$\min_{z_j \in Z - S_{m-1}} \frac{1}{m-1} \sum_{z_h \in S_{m-1}} |c(z_j, z_h)| \tag{17}$$

The equation (16) is equivalent to (12), while (17) is an approximation of (13). Therefore, if we use the combination criteria (15) the desired m^{th} variable satisfies the following condition known as FCQ: F-test Correlation Quotient (Ding and Peng, 2005):

$$\max_{z_j \in Z - S_{m-1}} \frac{F(z_j)}{\frac{1}{m-1} \sum_{z_h \in S_{m-1}} |c(z_j, z_h)|} \tag{18}$$

Moreover, in this work we introduce a new combination criterion that gives more weight to the redundancy factor. We call it FCO: F-test Correlation Out:

$$\max_{z_j \in Z - S_{m-1}} F(z_j) \cdot \left(1 - \max_{z_h \in S_{m-1}} |c(z_j, z_h)|\right) \tag{19}$$

The reason for this new scheme is that we discovered that the FCQ criterion is unable to separate highly correlated variables in the rank lists. The FCQ (18) selects a variable z_j for the set S if its F-value is high and if the averaged correlation with previously selected variables (set S_{m-1}) is low. However, if there is a variable in S_{m-1} that has a high correlation with z_j it may go unnoticed. On the other hand, the FCO (19) selects a variable z_j whose individual F-value is high, but whose maximum individual correlation with the variables of S_{m-1} is the lowest possible. As it is shown in Section 3.1, FCO outperforms FCQ in the sEMG classification task.

2.4.1 Application

For each subject, three different ranking lists of variables are computed from the subject's Training-dataset. The first ranking list is obtained using only a F-statistic (11) to rank the variables; this method is called Base ranking. The second and third ranking lists are obtained with the FCQ and FCO algorithms respectively.

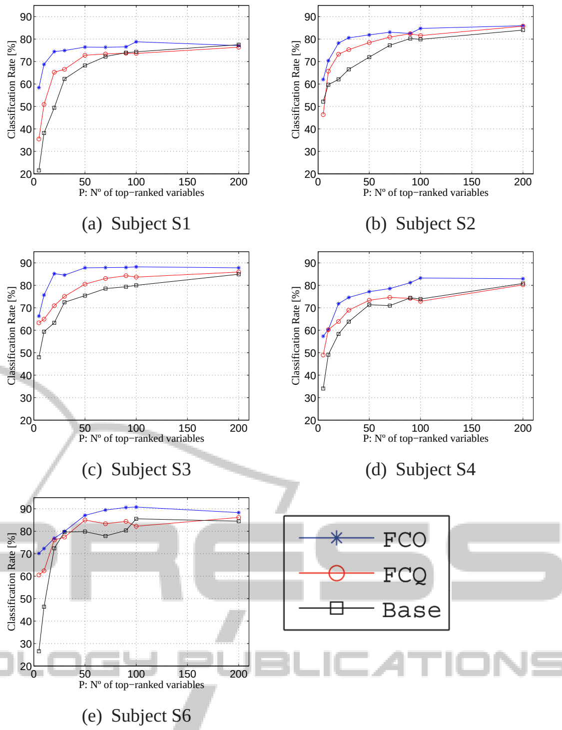
2.5 Performance of the Variable Selection Algorithms

In order to evaluate the performance of a Variable Selection's ranking list we have to compute the classification rate that is obtained when we only maintain in the vector $\{z_j\}_{j=1}^{1312}$ variables that are top-ranked in the ranking list. Given a Variable Selection method MET (i.e. a ranking list of variables) with $MET \in \{Base, FCQ, FCO\}$, we denote as z_p^{MET} to a variable that is top-ranked in the ranking list MET. If we only maintain in $\{z_j\}_{j=1}^{1312}$ the top-P variables of the ranking list MET the new vector is denoted as $\{z_p^{MET}\}_{p=1}^{P}$ (e.g. if we want to represent the trials with the Top-5 variables of the FCQ list the vector is denoted as $\left\{z_p^{FCQ}\right\}_{p=1}^{5}$).

The following procedure is followed with a Support Vector Machine (SVM) classifier and the data of subjects {S1, S2, S3, S4, S6} to evaluate the performance of the Base, FCQ and FCO algorithms. Subject S5 is excluded as according to Table 1 no trials are present in that subject's Validation-dataset. Given a subject and a Variable Selection method MET we have: A) the user's Training-dataset and Validation-dataset, in which the trials are represented by vectors of 1312 variables (Section 2.3); and B) a ranking list of the variables (Section 2.4.1). The evaluation procedure of the ranking list is as follows:

1. The vectors of variables from the Training-dataset and Validation-dataset are reduced to the Top-5 variables z_p of the ranking list MET (i.e. now each trial is represented by a vector of 5 variables, $\{z_p^{MET}\}_{p=1}^{5}$).
2. A grid search using a 5-fold cross-validation with the Training-dataset is used to find the best kernel's parameters of the SVM (radial basis function as kernel)(Chang and Lin, 2011).
3. The SVM classifier is trained with the vectors $\{z_p^{MET}\}_{p=1}^{5}$ of the Training-dataset.
4. The vectors $\{z_p^{MET}\}_{p=1}^{5}$ of the Validation-dataset are classified, hence, the classification rate is obtained.

The results are shown in Figure 2 for subjects {S1, S2, S3, S4, S6} and Variable Selection methods {Base, FCQ, FCO}. In addition to the Top-5 variables z_p of the ranking lists (step 1 of the above procedure), the Top-{10, 20, 30, 50, 90, 100 and 200} variables of each ranking list are also evaluated (i.e. P goes from $P = 5$ to $P = 200$ for each method).



(a) Subject S1 (b) Subject S2 (c) Subject S3 (d) Subject S4 (e) Subject S6

Figure 2: Classification rate for the Validation-dataset of each user when the vector $\{z_j\}_{j=1}^{1312}$ is reduced to $\{z_p^{MET}\}_{p=1}^{P}$, where z_p^{MET} are top-ranked variables of the ranking list MET, with $MET \in \{Base, FCQ, FCO\}$ and $P = \{5, 10, 20, 30, 50, 90, 100, 200\}$.

3 RESULTS

3.1 Analysis of the Performance of the Variable Selection Algorithms

Figure 2 shows that the more variables the vector $\{z_p^{MET}\}_{p=1}^{P}$ has, the better the classification rate. This is due to the fact that in our experiment the number of available training samples is much higher than the dimension P of the vector $\{z_p^{MET}\}_{p=1}^{P}$ (e.g. for subject S1 there are 4929 training samples according to Table 1, but the number of variables of the vector $\{z_p^{MET}\}_{p=1}^{P}$ is between 5 and 200). Therefore, we could infer that maintaining the whole vector $\{z_j\}_{j=1}^{1312}$ would be the best option. Nevertheless, as it is explained in Section 1, it would require more training samples and the classifier would be much more complex (Jain and Duin, 2000). In addition, note that the advantage of having more variables becomes insignificant at some point depending on what variables we are maintaining. For example, for subject S3 the difference in performance between using $\{z_p^{FCO}\}_{p=1}^{50}$ or $\{z_p^{FCO}\}_{p=1}^{200}$ is insignificant, whereas if we are using the ranking list Base the difference between $\{z_p^{Base}\}_{p=1}^{50}$ or $\{z_p^{Base}\}_{p=1}^{200}$ is high. This highlights the importance of the adequate selection of variables to have a set of variables as reduced as possible but obtaining, at the same time, a high classification rate with it; the set of variables capable of this is defined as a compact-set and the following facts reveal FCO as an excellent algorithm to search for this compact-set.

- The mRMR Variable Selection algorithms (FCO, FCQ) perform better than the classic univariate method of ranking by F-statistic. Figure 2 shows that when the variables are selected according to the Base rank the number of variables has to be higher to achieve an acceptable classification rate. It is true that the top-ranked variables of Base have large discriminant power individually, but they are redundant and when too few are used the classifier is unable to extract enough information from them. However, when the mRMR lists (FCO, FCQ) are used the classifier achieves better classification rate with less variables. Moreover, the fact that the FCO's classification rates are better than the FCQ's ones illustrates even more that reducing redundancy is a key factor.
- FCO algorithm shows better performance dealing with highly correlated variables. On the one hand, if we count the number of times that each feature appears in the ranking lists (Table 2), we can see that there are groups of features that appear more than others. It is clearly visible that the first AR coefficient (AR1) is a strong candidate for the compact-set as it appears many times (i.e. it extracts important information from many channels) and the three methods rank it very high. On the other hand, the first cepstral coefficient (Ceps1) also appears many times in the FCQ and Base ranks, but never in the FCO one. The reason is that if we look at (9) we see that if features AR1 and C1 are computed in the same channel, the coefficients are identical (with opposite signs). Unsurprisingly, we verified that the FCQ and Base top-ranked variables whose feature-term was AR1 and C1 had the same channel-term; this means that while FCO finds the correlation, FCQ and Base do not. In addition, the same behaviour was found in the WL (4) and MADV (5) features. Therefore, we can state that the F-test and the FCQ criteria fail to rank appropriately highly correlated variables.
- FCO algorithm finds key variables. Figure 2 shows that when the Base or FCQ lists are used, the respective vectors $\{z_p^{Base}\}_{p=1}^{P}$ or $\left\{z_p^{FCQ}\right\}_{p=1}^{P}$ have to have a high amount of variables to approximate the rate achieved with $\{z_p^{FCO}\}_{p=1}^{P}$. Based on Table 2 we could make the assumption that the classifier reaches the desired rate when features like MAV, MedAV or MADV are computed. This is supported by the fact that these variables are not in the first tops (Top-$\{5, 10, 20\}$) but they increase their presence as we take larger tops (Top-$\{30, 50, 100\}$). However, two facts reject this hypothesis. Firstly, these features are utterly discarded by the FCO algorithm (Table 2), they are not even in the FCO-Top-200. Secondly, at the end, the FCQ and Base lists include the best FCO variables. This is explained if the number of coincidences among the lists is analyzed. Table 3 shows how many variables from the FCO-Top-*P* are in the Top-$\{5,10,...,200\}$ of the FCQ list. The following examples are highlighted in bold. Table 3 shows that only 8 variables from the FCO-Top-30 are in the FCQ-Top-30, but if we look at the FCQ-Top-200 we see that it has 26 variables from the FCO-Top-30. On the other hand, only 14 variables from the FCQ-Top-30 appear in the FCO-Top-200. This is even more impressive if we check higher tops of the lists; while the whole FCO-Top-5 and Top-10 appear in the the FCQ-Top-200 list, just 3 and 6 variables of the FCQ-Top-$\{5, 10\}$ respectively are in the FCO-Top-200. That means that even the most important variables of the FCQ are not that informative according to the FCO method. The same trends were found for the case of Base-FCO.

From these facts and linked with the performance data, we can infer that there is a group of key variables that hold the important information, and they are well ranked by the FCO algorithm. The FCQ and Base lists do not achieve good classification rates due to the high number of variables, but rather at the end these lists include the best variables of the FCO list; the key variables. As explained in Section 1, a variable is a combination of a channel and the signal-feature extracted from it, hence, the task now is to analyze these variables and detect which are the best features and channels.

3.2 Selection of Features and Channels

The search for the variables of the compact-set entails a selection of features and channels among all the possible combinations. As it has been shown in Section 2.5 the FCO algorithm ranks these combinations and it highlights some of them as the most important

Table 2: The number of times that a feature appears in the Top-{30, 50, 100} of the FCO, FCQ and Base ranking lists (e.g. AR1 appears 7 times in the FCO-Top-30, that means that the information that AR1 extracts from 7 different channels is very important). The median across subjects is given. In gray the values greater than zero.

	Top-30			Top-50			Top-100		
	FCO	*FCQ*	*Base*	*FCO*	*FCQ*	*Base*	*FCO*	*FCQ*	*Base*
MAV	0	1	1	0	2	2	0	4	5
MedAV	0	0	0	0	1	1	0	3	3
VAR	0	0	0	0	0	0	0	1	1
MADV	0	2	2	0	3	4	0	7	8
WL	0	2	2	0	3	4	0	7	8
ZC	1	0	0	2	1	1	6	1	2
NT	3	2	4	4	3	6	7	6	12
WAMP	3	4	3	4	6	4	5	9	8
A1	1	0	0	2	1	0	3	2	0
A2	1	0	0	1	0	0	3	1	0
A3	0	0	0	0	0	0	1	0	0
A4	0	0	0	0	0	0	1	0	0
A5	0	0	0	1	0	0	2	1	0
A6	0	0	0	0	0	0	2	1	0
A7	0	0	0	0	0	0	2	0	0
A8	0	0	0	1	0	0	1	0	0
A9	1	1	0	2	2	1	3	3	2
AR1	**7**	3	4	8	6	6	10	9	11
AR2	1	0	0	3	1	1	6	3	2
AR3	0	0	0	1	0	0	2	0	0
AR4	0	0	0	0	0	0	0	0	0
C1	0	3	4	0	4	6	0	9	11
C2	1	0	0	1	0	0	2	1	0
C3	1	0	0	1	0	0	1	1	0
C4	0	0	0	0	0	0	2	0	0
Fmean	2	1	1	2	2	2	3	4	3
Q10	0	0	0	0	0	0	1	0	0
Q30	0	0	0	0	0	0	1	0	0
Q50	0	0	0	0	1	0	2	1	1
Q60	0	1	0	0	1	1	1	3	2
Q75	0	1	1	1	2	1	3	4	3
Q90	2	1	1	2	2	2	3	5	4
F1	0	0	0	0	0	0	0	0	0
F2	0	0	0	0	0	0	1	0	0
F3	0	0	0	0	0	0	1	0	0
F4	0	0	0	0	0	0	2	0	0
F5	0	1	0	1	1	0	3	2	1
F6	1	0	0	1	1	0	3	3	1
F7	0	1	0	1	2	0	2	3	2
F8	2	1	0	3	2	0	6	2	2
F9	1	0	0	2	1	0	3	1	0

Table 3: The number of coincidences between the FCO-Top-*P* and the FCQ-Top-*P* (e.g. 26 variables from the FCO-Top-30 are in the FCQ-Top-200). The median across subjects is given. In bold the examples explained in the text.

FCQ-Top-*P* variables	FCO-Top-*P* variables: *5*	*10*	*20*	*30*	*50*	*70*	*90*	*100*	*200*
5	1	2	2	2	2	2	2	2	**3**
10	3	3	4	4	4	4	4	4	**6**
20	4	5	5	6	7	8	8	8	10
30	4	6	7	**8**	10	11	11	12	**14**
50	5	7	9	13	16	18	19	20	26
70	5	8	11	16	19	22	25	27	34
90	5	9	12	17	22	26	31	32	41
100	5	10	14	18	22	26	32	33	46
200	**5**	**10**	18	**26**	37	47	53	56	85

ones for achieving a high classification rate. However, not only the best variables are the ones that maximize the classification rate, but also the ones that minimize the number of different channels and signal-features needed. Therefore, the search has to be focused on: A) features that appear many times in the ranking list and that share the same channels, since this reveals features with complementary information; and B) channels that appear many times in the highest positions of the FCO ranking list, because the algorithm has found them to be the most informative ones.

3.2.1 Features

We analyze the variables of the FCO-Top-100 by looking for variables whose channel-term is the same one. In other words, we count the number of times that any two features share the same channel in the FCO-Top-100 list. Table 4 shows the best features in the horizontal axis. We can see that the group of features formed by the first and second Auto-Regressive coefficients (AR1,AR2), the Zero Crossing (ZC), the Number of Turns (NT), and the Wilson Amplitude (WAMP) appear together many times. Furthermore, they also appear many times in the best positions of the FCO ranking list (Table 2), hence we can state that they interact perfectly well among themselves and that they are of great importance for identifying the different kinds of EMG spike burst activity. In addition, they seem to be able to share a channel with almost any other feature of the list. Note that there is no interaction with the features {MAV, MedAV, Var, MADV, WL} because these are not present in the FCO-Top-100 (Table 2). This is due to the fact that some of these features do not provide any valuable information or are highly correlated with the first group. Therefore, we can discard the features {MAV, MedAV, Var, MADV, WL}; and consider the {AR1, AR2, WAMP, ZC, NT} as the initial members of the compact-set.

Another group of features that shows great interaction is the frequency histogram features (F1-F9). However, a selection becomes necessary as there are important differences among them:

- On the one hand, we select for the compact-set the coefficients corresponding to high parts of the spectrum ({F5-F9}). These coefficients represent the percentage of power that resides in the 222-500 Hz band (divided in 5 segments). It is very interesting to see in Table 4 that {F5-F9} fit with NT, ZC and the AR coefficients AR1 and AR2. The reasons are the following. With the number of turns (NT) we measure the number of spikes that are generated in the muscle fibers; therefore, NT grows with this number. However, if the frequency of these spikes is too high they are hardly distinguishable in the resulting sEMG wave; the NT does not count those slope turns and the coefficient does not grow. Therefore, the {F5-F9} coefficients combine perfectly with NT as they provide the missing information. Similarly, ZC can be complemented by {F5-F9} as a high number of spikes may hinder the crossings in the zero amplitude axis. Finally, the FCO algorithm does not find high correlations between the {F5-F9} and {AR1, AR2}, hence both groups of coefficients can share the same channels. The interaction of {F5-F9} with the WAMP feature is not so strong as some redundancy is found, but given their high positions in the FCO ranking we decide to maintain both {F5-F9} and WAMP in the compact-set.

- On the other hand, the information of the {F1-F4} coefficients is not relevant as it is embedded in features like NT, ZC and WAMP.

The amplitude histogram features (A1-A9) tend to appear together and with the {AR1, AR2, WAMP, ZC, NT } group. However, their interaction with other features is insignificant and their presence in the different FCO-Tops (Table 2) is not large enough. Moreover, establishing the size of the different amplitude levels is not straightforward, as it is different for each user and requires some tuning. Based on these facts, we can reject them for the compact-set.

The auto-regressive coefficients AR3 and AR4, and the four Cepstral coefficients (C1-C4) are discarded because they are are not top-ranked by FCO. For identical reasons we discard also the quantiles {Q10-Q90}. The information of {Q10, Q30} is already provided by the group {AR1, AR2, WAMP, ZC, NT }, and {Q50-Q90} are not necessary as the frequency histogram features {F5-F9} are found to be more informative.

Finally, the mean frequency (Fmean) is discarded because even though it is well ranked in the FCO-Top-30 (it appears twice), its presence in FCO-Top-50 and FCO-Top-100 is almost the same. This means that its contribution is not determinant for the classification task, the FCO algorithm finds that any of the features {AR1, AR2, WAMP, ZC, NT, F5-F9} provides better information that Fmean.

In summary, the selected signal-features for the compact-set are {AR1, AR2, WAMP, ZC, NT, F5-F9}.

3.2.2 Channels

The selection of the best channels is straightforward. We search in the different FCO-tops of each user the most repeated channels and their evolution. Thanks to the mRMR-FCO algorithm we know that those channels will carry useful and non-redundant information.

Table 5 shows the number of times that each channel appears in the FCO-Top-{50,100} (the median across subjects is shown). There are two important groups clearly visible: {Ch05, Ch06, Ch13, Ch14, Ch21, Ch22, Ch29, Ch30} which are above the extensor carpi radialis and extensor digitorum communis muscles; and {Ch25, Ch26} which are above the flexor digitorum sublimis and flexor carpi ulnaris muscles.

Table 4: The number of times that, in the FCO-Top-100 list, two given features share the same channel, e.g. the ZC shares two channels with NT, 2 with WAMP, 1 with A1, etc. The median across subjects is given. We only show in the horizontal axis the best features. The values greater than zero are in gray.

	ZC	NT	WAMP	AR1	AR2	F5	F6	F7	F8	F9
MAV	0	0	0	0	0	0	0	0	0	0
MedAV	0	0	0	0	0	0	0	0	0	0
VAR	0	0	0	0	0	0	0	0	0	0
MADV	0	0	0	0	0	0	0	0	0	0
WL	0	0	0	0	0	0	0	0	0	0
ZC		2	2	2	2	1	1	1	3	2
NT	2		2	4	1	1	1	1	2	1
WAMP	2	2		2	2	1	1	1	0	0
A1	1	2	1	3	0	0	1	0	0	0
A2	0	0	1	1	0	0	0	0	1	0
A3	0	0	0	0	0	0	0	0	0	0
A4	0	1	0	0	0	0	0	0	0	0
A5	1	1	1	0	1	0	0	0	0	0
A6	1	1	1	1	0	0	0	0	0	0
A7	0	1	0	0	1	0	0	0	0	0
A8	0	1	1	0	0	0	0	0	1	0
A9	1	1	0	1	1	1	0	0	1	1
AR1	2	4	2		0	1	0	0	1	1
AR2	2	1	2	0		1	1	2	2	1
AR3	1	0	1	1	0	0	0	0	1	0
AR4	0	0	0	0	0	0	0	0	0	0
C1	0	0	0	0	0	0	0	0	0	0
C2	1	1	0	2	0	0	0	0	1	1
C3	1	1	1	0	1	0	1	1	0	0
C4	2	1	1	0	1	0	0	0	1	0
Fmean	0	1	1	0	1	0	0	1	1	1
Q10	0	0	1	1	0	0	0	0	0	0
Q30	0	0	0	1	0	0	0	0	0	0
Q50	1	1	1	1	1	1	1	1	1	0
Q60	1	1	0	1	0	0	0	0	1	0
Q75	1	0	0	1	1	0	1	0	1	0
Q90	2	1	0	1	1	0	1	1	1	0
F1	0	0	0	0	0	0	0	0	0	0
F2	1	1	0	0	0	1	0	0	1	1
F3	1	0	0	1	0	0	0	0	0	0
F4	1	0	0	1	1	1	0	0	1	1
F5	1	1	1	1	1		1	1	2	1
F6	1	1	1	0	1	1		1	2	0
F7	1	1	1	0	2	1	1		2	1
F8	3	2	0	1	2	2	2	2		2
F9	2	1	0	1	1	1	0	1	2	

In (Hargrove et al., 2007) they found the same results with an EMG pattern recognition task with similar gestures and a wrapper Variable Selection method. The results in the present study are further confirmation of the importance of these muscles and the fact that the mRMR-FCO algorithm is an excellent tool for Variable Selection.

4 FINAL TEST

The selected 10 features {AR1, AR2, WAMP, ZC,

Table 5: The number of times that a channel appears in the FCO-Top-{50,100}.The median across subjects is given. The selected channels are in bold.

	FCO-Top-50							
Ch01-08	0	0	1	1	**3**	**2**	1	1
Ch09-16	1	2	0	0	**3**	**3**	1	0
Ch17-24	1	1	0	0	**4**	**5**	2	0
Ch25-32	**3**	**3**	0	0	**3**	**6**	1	0

	FCO-Top-100							
Ch01-08	1	1	2	2	**5**	**5**	4	1
Ch09-16	1	2	0	1	**4**	**8**	2	1
Ch17-24	1	3	0	2	**6**	**10**	2	0
Ch25-32	**4**	**4**	1	1	**6**	**9**	2	2

NT, F5-F9} and the selected 10 channels {Ch05, Ch06, Ch13, Ch14, Ch21, Ch22, Ch25, Ch26, Ch29, Ch30} form 100 combinations (i.e. a compact-set of 100 variables). To test the hypothesis that this compact-set holds most of the key information, the following procedure is applied to the data of subjects S3, S4 and S5. Firstly, from the trials of the Train, Validation and Test datasets the selected variables from the compact-set are extracted, which means using only the selected channels and computing the selected features from them. Secondly, an SVM classifier is trained with the Train-dataset. Finally the Validation and Test datasets are passed to the classifier for classification. The results are shown in Table 6. Note that as subject S5 has no Validation-dataset (Table 1) no classification rate is computed in that case.

On the one hand, with the classification rate of the Validation-datasets we can make a comparison of the performance between using the Top-100 variables of the FCO ranking list or using the 100 variables of the compact-set. As it is shown in Table 6, the differences in performance are very low. For example, for user S3, using the best 100 combinations of channels and features of that specific user (according to FCO) we only achieve a 4% extra classification rate than when using the selected 100 combinations of the compact-set. On the other hand, Table 7 shows that the FCO-Top-100 list of user S3 demands 25 different sEMG channels in the system, whereas the compact-set list only demands 10 channels. Moreover, the compact-set has the same channels and signal-features for every user, unlike the FCO-Top-100 list, which has the best ones depending on the user.

Finally, as the channels and features of the compact-set are selected knowing that they ensure a high classification rate of the Validation-dataset's trials (Figure 2), it is imperative to compute the classification rate for non-tested trials; the Test-dataset. As it is shown in Table 6 the information extracted from the features {AR1, AR2, WAMP, ZC, NT, F5-F9} and the channels {Ch05, Ch06, Ch13, Ch14, Ch21, Ch22, Ch25, Ch26, Ch29, Ch30} is enough for an accurate classification in any user.

Table 6: Classification rate of the Validation and Test datasets when the compact-set of variables is used.

	Validation-dataset		Test-dataset
Subject	FCO-Top-100	Compact-set	Compact-set
S3	88%	84%	77%
S4	83%	80%	84%
S5	-	-	80%

Table 7: Comparison of the number of different channels and signal features in the Top-100 variables of the FCO, FCQ and Base ranking lists. The compact-set has also 100 variables as all the combinations among the selected channels and the selected features are included.

Subject S3	Channels	Features
FCO-Top-100	25	26
FCQ-Top-100	21	18
Base-Top-100	24	16
Compact-set	10	10

Subject S4	Channels	Features
FCO-Top-100	22	31
FCQ-Top-100	20	32
Base-Top-100	16	24
Compact-set	10	10

Subject S5	Channels	Features
FCO-Top-100	25	24
FCQ-Top-100	26	22
Base-Top-100	28	15
Compact-set	10	10

5 CONCLUSIONS

In the present study 32 sEMG electrode locations and 41 signal-features are tested with the aim of reducing the necessary number to obtain an accurate classification rate in a 15 static-hand gesture classification task. A novel implementation of the mRMR Variable Selection algorithm is introduced to highlight the most informative but least redundant combinations of sEMG channels and signal-features.

The results show that the electrodes above the extensor carpi radialis, extensor digitorum, flexor digitorum sublimis and flexor carpi ulnaris muscles are the best ones for this kind of application. Moreover, there is a group of signal-features that have high discriminant power individually and that can extract non-redundant information from each of these channels.

The signal-features are the first and second Auto-Regressive coefficients, the Zero Crossing, the Number of Turns, the Wilson Amplitude and the frequency histogram coefficients of the 225-500 Hz band (divided in 5 segments).

We believe that the proposed methodology for channel and feature selection makes a significant improvement with respect to the current ones. This novel methodology may not only be appropriate for the particular application presented in this paper (i.e. channel and feature selection for hand gesture detection based on sEMG signals) but also in other case scenarios, such as gene selection for classification of phenotypes based on microarray data.

ACKNOWLEDGEMENTS

The authors would like to thank the department of Education of the Basque Government for their support.

REFERENCES

Bellman, R. (1961). *Adaptive Control Processes: A Guided Tour*, volume 19 of *A Rand Corporation Research Study Series*. Princeton University Press.

Boostani, R. and Moradi, M. H. (2003). Evaluation of the forearm EMG signal features for the control of a prosthetic hand. *Physiological measurement*, 24(2).

Chang, C.-c. and Lin, C.-j. (2011). LIBSVM : a Library for Support Vector Machines.

Cover, T. T. (1974). The Best Two Independent Measurements are not the Two Best. *Systems, Man and Cybernetics, IEEE Transactions*, 1(JANUARY):116–117.

Ding, C. and Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*.

Englehart, K., Hudgins, B., and Member, S. (2003). A Robust , Real-Time Control Scheme for Multifunction Myoelectric Control. *Engineering In Medicine And Biology*, 50(7):848–854.

Fuglsang-Frederiksen, A. and Pugdahl, K. (2010). Current status on electrodiagnostic standards and guidelines in neuromuscular disorders. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, 122(3):440–455.

Hargrove, L. J., Englehart, K., and Hudgins, B. (2007). A comparison of surface and intramuscular myoelectric signal classification. *IEEE transactions on biomedical engineering*, 54(5):847–53.

Hudgins, B., Parker, P., and Scott, R. N. (1993). A new strategy for multifunction myoelectric control. *IEEE transactions on bio-medical engineering*, 40.

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63.

Jain, A. and Duin, R. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4–37.

Kang, W. J., Shiu, J. R., Cheng, C. K., Lai, J. S., Tsao, H. W., and Kuo, T. S. (1995). The application of cepstral coefficients and maximum likelihood method in EMG pattern recognition. *IEEE transactions on biomedical engineering*, 42(8):777–85.

Khokhar, Z. O., Xiao, Z. G., and Menon, C. (2010). Surface EMG pattern recognition for real-time control of a wrist exoskeleton. *Biomedical engineering online*.

Kuiken, T. a., Miller, L. a., Lipschutz, R. D., Lock, B. a., Stubblefield, K., Marasco, P. D., Zhou, P., and Dumanian, G. a. (2007). Targeted reinnervation for enhanced prosthetic arm function in a woman with a proximal amputation: a case study. *Lancet*.

Merletti, R. and Parker, P. (2004). Electromyography: Physiology, engineering, and noninvasive applications. *Engineering In Medicine And Biology*.

Palmes, P., Ang, W. T., Widjaja, F., Tan, L. C. S., and Au, W. L. (2010). Pattern mining of multichannel sEMG for tremor classification. *IEEE transactions on biomedical engineering*, 57(12):2795–805.

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE transactions on pattern analysis and machine intelligence*, 27(8):1226–38.

Philipson, L, Larsson, P, G. (1988). The electromyographic signal as a measure of muscular force: a comparison of detection and quantification techniques. *Electromyography and clinical neurophysiology*.

Tenore, F., Ramos, A., Fahmy, A., Acharya, S., Etienne-Cummings, R., and Thakor, N. V. (2007). Towards the control of individual fingers of a prosthetic hand using surface EMG signals. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*

Tenore, F. V. G., Ramos, A., Fahmy, A., Acharya, S., Etienne-Cummings, R., and Thakor, N. V. (2009). Decoding of individuated finger movements using surface electromyography. *IEEE transactions on biomedical engineering*, 56(5):1427–34.

Tkach, D., Huang, H., and Kuiken, T. a. (2010). Study of stability of time-domain features for electromyographic pattern recognition. *Journal of neuroengineering and rehabilitation*, 7:21.

Zardoshti-Kermani, M., Wheeler, B., Badie, K., and Hashemi, R. (1995). EMG feature evaluation for movement control of upper extremity prostheses. *IEEE Transactions on Rehabilitation Engineering*.