EMG Signal Processing with Clustering Algorithms for motor gesture Tasks

Víctor Asanza
Facultad de Ingeniería en Electricidad y Computación
Escuela Superior Politécnica del Litoral, ESPOL
Guayaquil, Ecuador
vasanza@espol.edu.ec

Enrique Peláez
Facultad de Ingeniería en Electricidad y Computación
Escuela Superior Politécnica del Litoral, ESPOL
Guayaquil, Ecuador
epelaez@espol.edu.ec

Francis Loayza
Facultad de Ingeniería en Electricidad y Computación
Escuela Superior Politécnica del Litoral, ESPOL
Guayaquil, Ecuador
floayza @espol.edu.ec

Iker Mesa
CEIT, Parque Tecnológico de San Sebastian
Universidad de Navarra
San Sebastian, Spain
imesa@ceit.es

Javier Díaz
CEIT, Parque Tecnológico de San Sebastian
Universidad de Navarra
San Sebastian, Spain
jdiaz@ceit.es

Edwin Valarezo
Department of Biomedical Engineering, College of Electronic and Information
Kyung Hee University
Yongin, Republic of Korea
Facultad de Ingeniería en Electricidad y Computación
Escuela Superior Politécnica del Litoral, ESPOL
Guayaquil, Ecuador
edgivala@khu.ac.kr

***Abstract*— Recent research shows the possibility of using electromyography (EMG) electrical signals to control devices or prosthesis. The EMG signals are measured in muscles, such as the forearm. These signals can lead to determine the intentionality of the patient when performing any motor tasks, however the signals are susceptible to noise due to the voltage sensed, which is in the microvolts scale. In this work, the preprocessing of the EMG signals includes the design and test of a filter. Our designed filter allows eliminating any signal components from the electrical network or any other sources that are not EMG signals. To validate the preprocessing efficiency, we analyze the frequency components and the distribution of the filtered EMG signals. Later, the filtered data was processed with K-means, DBSCAN and Hierarchical Clustering algorithms to determine a subject's intention when performing a task. The results show that the K-means clustering algorithm was able to group the nine gestures made by the subjects, as compared to the DBSCAN and Hierarchical algorithms, which were not able to perform the clustering as expected. However, they match the performance of clustering two groups of combining gestures.**

Keywords— Electromyography, Fast Fourier Transform, Butterworth Filter, K-means, DBSCAN, Hierarchical-Clustering.

I. INTRODUCTION

The biomedical signals of Electromyography (EMG) allow measuring the potential difference generated in the muscles during contraction and relaxation. These signals represent neuro-muscular activities controlled by the motor cortex of the brain [1]. In the motor cortex, a train of electrical impulses are generated; then the cerebral nerves carry out this signal across the midbrain. These signals descend through the spinal cord; via the synapse the receptor excites the muscle, causing the muscle to shorten [2].

There are two methods for measuring EMG signals: intramuscular and superficial. The intramuscular measurement is performed with a hypodermic needle that is inserted through the skin to the muscle tissue; this method is considered invasive and unnecessary in some cases. In contrast the superficial method is non-invasive and represent less risk for the patient, consequently it is most popular. In this method an electrode measures the muscle activation within few muscle fibers [3].

The superficial method has interference because of the sensors used in the skin to collect the electrical signals coming from the muscles. While they travel through different tissues, the amplitude of the EMG signals lies in the range of microvolts (uV) to millivolts (mV); however, the electrical level typically is less than 10mV. The properties of the skin (eg. skin thickness, adipose tissue, among others), duration and intensity of muscle contraction define the characteristics of the EMG signals in time and frequency domains; as well as, the properties of the amplifier, type of electrodes, and contact skin-electrode [4, 5].

The development of technologies applied to rehabilitation of patients with motor disabilities is needed, especially for those who are not able to control their limbs because diseases such as: Cerebral palsy, stroke or muscular spasticity. One could determine the patient movements intention by measuring and processing bioelectrical signals. Then an assistive device could help them to perform an action. The noise susceptibility of EMG signals has motivated several methodologies to analyze and preprocess EMG signals. Our contribution is to use preprocessed EMG signals with clustering algorithms to detect the intention while developing motor tasks. [6].

In this work, we used statistical analysis of the EMG signals to detect the presence of noise. Then, the EMG signals are preprocessed to eliminate the noise components. Finally, the preprocessed signals were processed using clustering algorithms.

This paper is organized as follows: Section II describes the methodology used to collect the data with healthy subjects as well as the pre and processing of the collected data. Section III shows our results. In section IV we present a discussion of our results and conclusions.

978-1-5386-6657-9/18/$31.00 ©2018 IEEE

II. METHODOLOGY

The data set used in this work was collected by colleagues, researchers at Parque Tecnológico de San Sebastian – Spain. The EMG signals were measured from six healthy subjects, right-handed and without muscular disorder. The subjects provided a written consent before the experiments were performed.

The experiments consisted in recording electrical EMG signals obtained from the forearms of the subjects through thirty superficial electrodes (10mm diameter), while the subjects responded to a visual stimulus. The stimulus is related to the sign language, and a computer program randomly generates the gestures. The subjects' tasks consisted in reproduce the random gestures displayed on the computer monitor [4].

Fig. 1 shows the set of gestures from which a set of nine letters in sign language were selected. The projection time of each visual stimulus, corresponding to nine tasks, was 6.6 seconds. This time was established considering previous tests, as well as to give enough time to the subject to reproduce the task and capture the EMG signals [4].

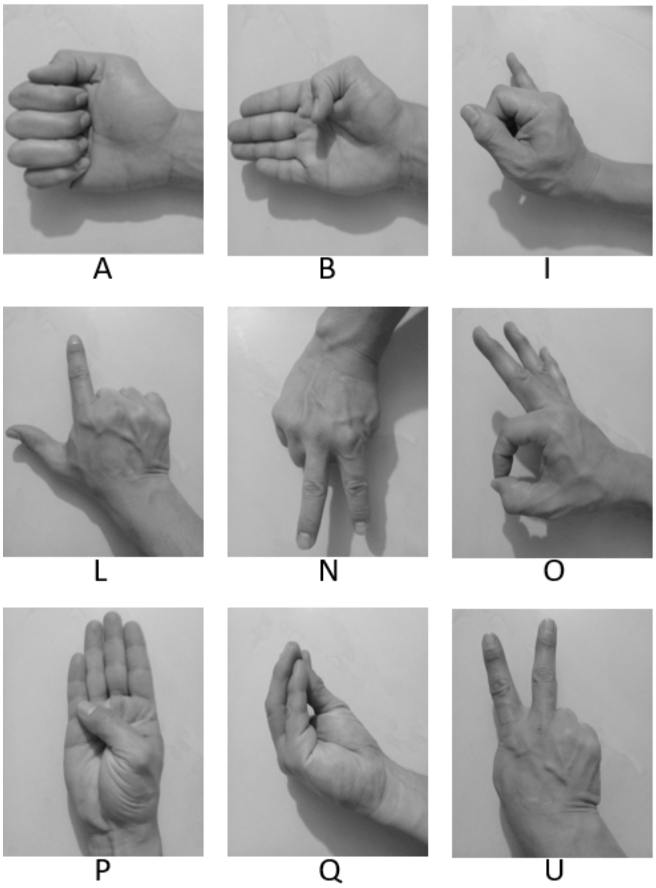

Fig. 1. Hand gesture reproduction to the visual stimuli used to acquire the EMG signals. Procedure also described in [4]

The thirty surface electrodes were uniformly distributed in the subject's forearm as shown in Fig. 2. This uniform distribution in the right forearm allowed to record the muscle activity completely in each experiment.

Each subject performed nine repetitions presented pseudo-randomly each of the nine gestures (A, B, I, L, N, O, P, Q, U). The EMG signals (32 electrodes: two as land reference) were recorded in parallel (All sensors at a time) during the experiment. The sampling frequency was 1000Hz, and each event lasted 6.6 seconds [4].

To ensure stability recording the EMG signals, we used only the last 3600 samples of the 6600 acquired for each tasks or letter performed, considering that the subject by that time had already maintained the forearm and the gesture [4].

Fig. 3 shows a plot of electrical signals in the time domain, which represents the muscular activity of the right forearm of subject 1. The performed task was mimicking the A letter in the first repetition. The electrical signal contains 3600 samples from thirty electrodes registered simultaneously; the signals contain DC components or low frequency artifacts (close to 0Hz), mainly due to sweating on the skin or relative movement between the skin and the electrodes. Other sources of noise have frequency components, such as the heart rate (1Hz to 2Hz) and the frequency of the electrical network (50Hz) [4].

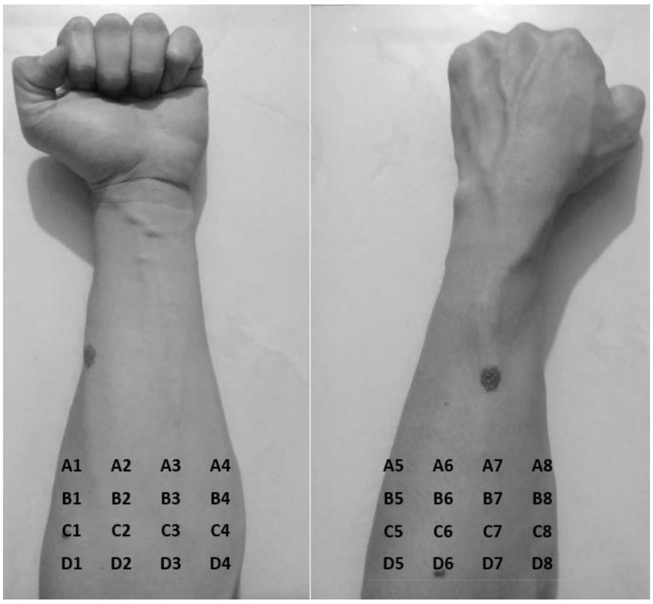

Fig. 2. Distribution of thirty-two surface electrodes around extensor digitorum communis muscle. Electrodes named A6 and D6 were used as reference to ground [4]

The preprocessing data analysis was performed using Matlab 2017b (The MathWorks, Natick, MA) and the *Signal Processing Toolbox*. The clustering algorithms were performed in RStudio V1.0.143. The EMG signals registers were grouped and stored in one Matlab file (D_{EMG}), containing 3600 samples of all 30 electrodes for each of the 9 gestures, for 9 repetitions performed by the 6 subjects (3600 x 30 x 9 x 9 x 6).

Due to the presence of noise and DC artifacts, the data does not present a normal distribution. Once identified the presence of DC artifacts the next step, in this methodology, was to perform the normality hypothesis test of the EMG signals. The test performed with a significance level of 5% showed that, for the null hypothesis H_o of the captured EMG signals, it did not present a normal distribution with zero mean and variance with value 1.

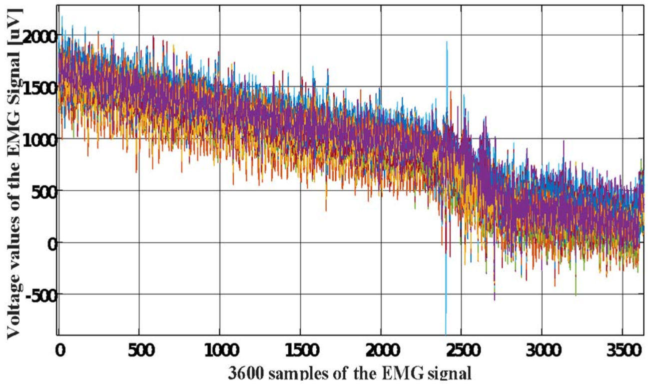

Fig. 3. DC artefacts present on the thirty electrodes of the EMG signal for the letter A

Whereas for hypothesis H_1, the EMG signals showed a normal distribution with zero mean and variance with value 1. Therefore, there is insufficient evidence to reject the null H_o hypothesis, given that the captured signals were very susceptible to noise-induced interference from DC artifacts, as well as from the electricity network (50Hz). The presence of DC artifacts modifies the characteristics of descriptive statistics in the time domain, such as mean values and variance. The Fig. 4 shows the histogram of all EMG signals from the thirty superficial electrodes, with low frequency noise close to 0Hz and noise with frequencies higher than 20Hz.

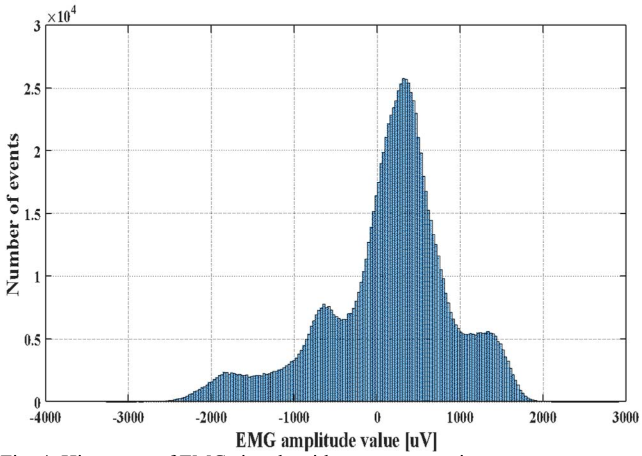

Fig. 4. Histogram of EMG signals without preprocessing

The proposed methodology also considers to perform an analysis in the frequency domain, to determine the presence of noise generated by the electrical distribution network. This analysis was done using the Fast Fourier Transform (FFT) to the D_{EMG} data Matlab file.

The Fig. 5 shows that all electrodes have a strong activity due to the presence of DC artifacts values close to 0 Hz frequency. Also, it is possible to observe a pronounced activity in the frequency of 50Hz; which is known as fundamental harmonic generated by the electric network.

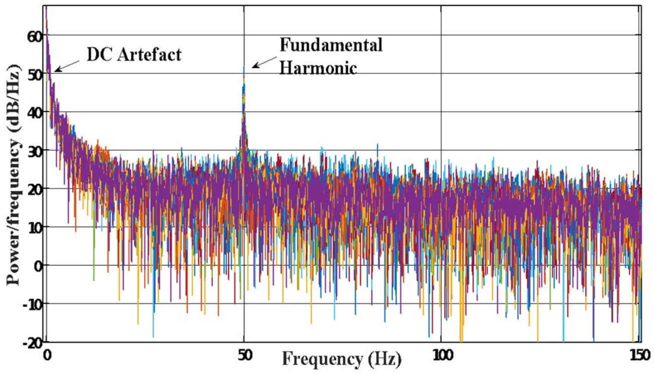

Fig. 5. Frequency analysis using the FFT of EMG signals with noise components

The muscular activity measured by the EMG surface electrodes, recorded the muscular activity in the range frequency of 7Hz to 20Hz. Fig. 6 shows the *Bandpass filter Butterworth-IIR*, the constant coefficient was designed for the band frequency of 7Hz to 20Hz. To produce a frequency response as stable as possible, keeping the signal of interest by removing the noise from the acquired EMG data [7].

Once the acquired data of the thirty EMG electrodes were preprocessed, some temporal features were extracted, for instance: the maximum value, the minimum value, the arithmetic mean and median values during 3600 samples (4 features for each electrode). Additionally, the variance, covariance, and the correlation between all thirty electrodes (90 features for each electrode) were extracted. Furthermore, features from frequency domain signals such as: the maximum value, the minimum value, the arithmetic mean, the median and the maximum-value, index (which-Max) in the periodogram (5 features for each electrode) were also extracted. We have also considered the variance value, the covariance and correlation values between all thirty electrodes (90 features for each electrode).

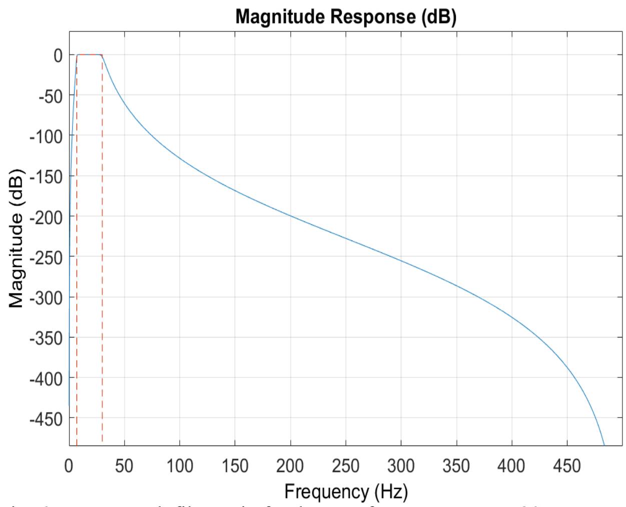

Fig. 6. Butterworth filter gain for the rage frequency 7Hz to 20Hz

These 189 features extracted from each of the 30 EMG electrodes (5670 features in total), were created and stored in a new two-dimensional Matlab file (F_{EMG}), where its rows represent the nine gestures or tasks performed by the six subjects during nine repetitions (486 rows), and the columns contain the time and frequency features (5670 columns). Additionally, a first column was added with the label of the letter reproduced by the subjects, this label allowed us to evaluate the success rate of the clustering algorithms used.

The first clustering algorithm used with F_{EMG} (486x5671) file was *K-means*. Following Hartigan and Wong [7, 8] we determine the appropriate number of clusters (k = 9). Using a convex type configuration, this algorithm obtained nine clusters, each corresponding to the nine gestures. The results correspond to: Cluster 1 – the task performed for gesture letter A, Cluster 2 – letter B, Cluster 3 – letter I, Cluster 4 – letter L, Cluster 5 – letter N, Cluster 6 – letter O, Cluster 7 – letter P, Cluster 8 – letter Q, and Cluster 2 – letter U.

The *Density-Based Clustering algorithm (DBSCAN)* was also tested with the F_{EMG} (486x5671) file. The minpoints are how many neighbors a point must have to be considered into a cluster; but before using this algorithm, we calculated the Epsilon (EPS) distance value between the points to be considered part of the same cluster apply with a minpoints = 7 and a distance value EPS = 65 [9, 10]. The DBSCAN algorithm obtained: Cluster 1 corresponding to gestures A, B, I, L and N. Cluster 2 correspond to letter O, P, Q and letter W.

The third algorithm used was the *Hierarchical Algorithm*, which this cluster analysis agglomerated the F_{EMG} (486x5671) file into two clusters tree: Cluster 1 conformed by letters A, I and L. Cluster 2 corresponding to the tasks performed to letter B, O, P, Q, W and letter N.

III. RESULTS

The designed filter to preprocess EMG signals was able to reduce the noise identified. Fig. 7 shows the reduction of noise in the time domain, corresponding to letter A performed by the subject one, during the first repetition. The EMG signals showed a peak response during the initial 500 samples, then the signal is stabilized close to zero volts, which is a typical response of *Butterworth-IIR Bandpass* filter.

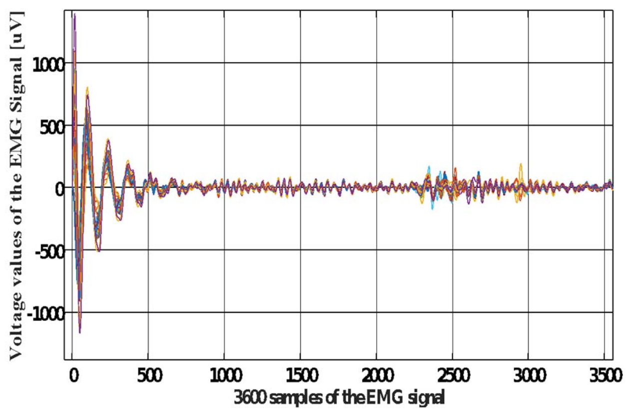

Fig. 7. EMG signal without DC artifacts present on the thirty electrodes of a letter A

The *Bandpass* filter designed reduced the 50Hz frequency of fundamental harmonics, generated by the electric network and the DC artifacts of low frequency close to 0Hz, the preprocessing of D_{EMG} (3600 x 30 x 9 x 9 x 6) Matlab file, allows us to evaluate the muscular activity during the tasks' execution by the subjects. Fig. 8 shows the frequency activity of EMG signals in the range frequency of 7Hz to 20Hz, during the performed gesture correspond to letter A, by the subject one during first repetition.

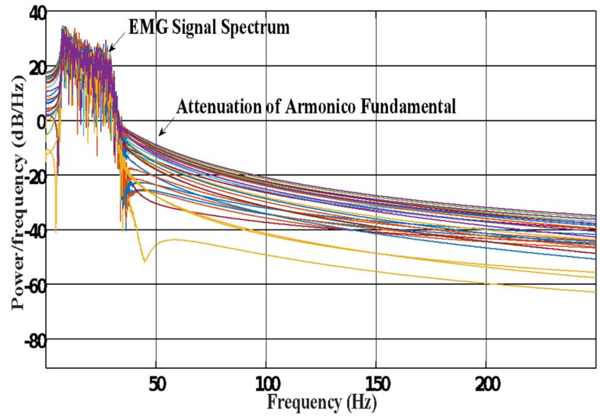

Fig. 8. Frequency analysis of the filtered EMG signals using FFT

After the preprocessing of the EMG signals, the normality hypothesis test was performed, again for a significance level of 5%. The null hypothesis H_o: The EMG data does not show a normal distribution (zero mean and variance 1) vs. H_1: The EMG data shows a normal distribution (zero mean and variance 1). The test shows that there is enough evidence to reject the null hypothesis H_o. Therefore, the hypothesis H_1 is accepted, which indicates that EMG signals have a normal distribution.

Fig. 9 shows the histogram of all filtered EMG signals of the thirty electrodes, where the normal distribution is observed with zero mean and a variance 1, which is consistent since all the electrodes have activity in the band of 7Hz to 20Hz, without DC artifacts. Unlike the histogram shown in Fig. 4, with deviation of the mean product of low frequency noise near 0Hz, and very wide variance product of the noise of the frequency noises greater than 20Hz.

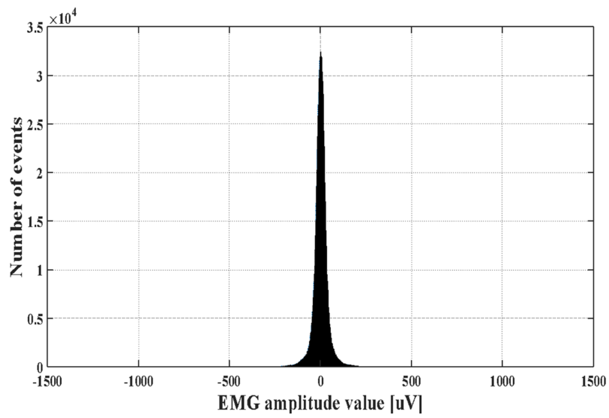

Fig. 9. Histogram of Filtered EMG signals with normal distribution

Fig. 10 shows the sum of square error (SSE) vs. the number of clusters; from where we infer that the appropriate number of clusters is k = 8. This gives a percentage between_SSE and total_SSE of 74.3%. However, because the number of clusters must be nine, as the number of gestures, the value between_SSE and total_SSE was the same percentage. In the detection of the nine gestures, the K-means algorithm was able to cluster each one of the nine letters performed by the subjects, and obtained a mean success rate close to 50% with all nine clusters.

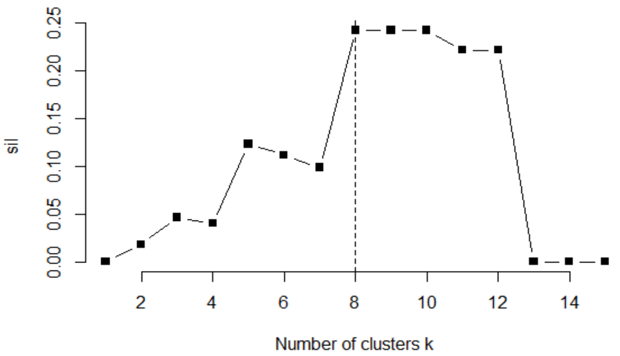

Fig. 10. Evaluation of the K-means algorithm with nine centroids

The evaluation of the *DBSCAN algorithm* is shown in Fig. 11, where the best number of clusters were two groups, and has a very low ability to detect the letters performed by the subjects. In this case the *DBSCAN algorithm* has a success rate close to zero because it was no able to cluster the nine letters and was no able to differentiate the letters A, B, I, L or letter N. Likewise, this algorithm was no able to differentiate the letters O, P, Q nor letter U.

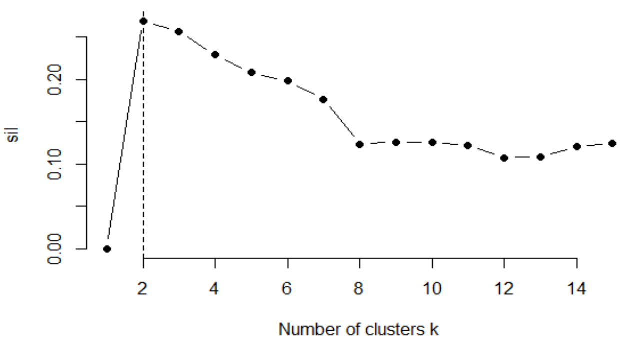

Fig. 11. Evaluating the DBSCAN algorithm

Fig. 12 shows the result of evaluating the cluster quantity of the *Hierarchical Agglomeration algorithm*, where the best number of clusters are two groups, this algorithm was not able to detect the nine letters performed by the subjects. The *Hierarchical Clustering algorithm* obtained a success rate close to zero and it was no able to differentiate the letters A, I or letter L. Also, this algorithm was no able to differentiate the letters B, O, P, Q, U or letter N.

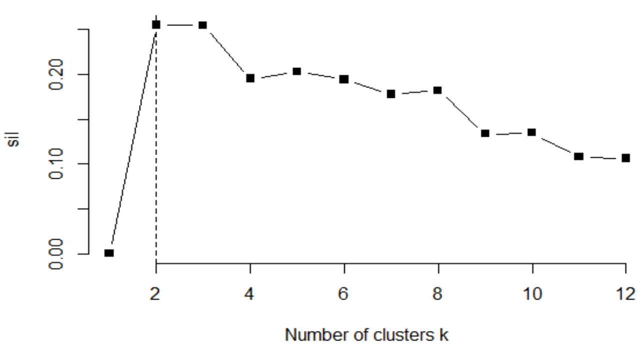

Fig. 12. Evaluation of Hierarchical Clustering algorithm

IV. Discussion and Conclusions

In this work, we used a normal distribution to detect the presence of noise on EMG signals; based on descriptive statistics parameters such as mean and variance. This detection allows us to determine the need of preprocessing, because the noise of low frequency varied the mean distribution of the data set and the noise with frequency greater than 20Hz, increased the variance greater than one. In addition, the use of the same test of normality after the preprocessing with the *Bandpass* filter, allowed us to evaluate the efficiency of the filter in the range frequency of 7Hz to 20Hz.

The used of FFT in the EMG signal without preprocessing let us know that there is a strong activity at 50Hz and near 0Hz, which is an effect of the fundamental harmonic of the electric network and of artifacts DC respectively. Furthermore, the EMG signals were processed without time domain processing to detect the trends produced by the DC artifacts. However, with the use of the normality test, the presence of noise can be detected without the need to perform the representation in the time domain and FFT.

The hypotheses tested helped us to detect the presence of noise. Low frequency DC artifacts affect the temporal features, such as: the average and variance. The signals without preprocessing were analyzed in the frequency domain; these signals showed a greater activity close to 50Hz, which correspond to the artifact of the electric network. This noise affected the characteristics in the frequency domain of the signal.

From the evaluated algorithms, we can conclude that the *K-means Clustering algorithm* was able to detect each one of nine tasks performed with acceptable success rate. This algorithm is based on Euclidean distance to the centroid of the cluster; hence, it has a better response in clustering the features autocorrelated in the time domain [11]. The algorithms based on density *DBSCAN* and *Hierarchical Clustering* presented a low success rate. They group all letters into two clusters because the features based on the frequency do not have enough information about the sharper time resolutions [12].

For future work, we will evaluate and compare other clustering algorithms with additional features based on Wavelet Transform with sharper time resolution [12, 13]. Additionally, the dataset obtained has information about the cerebral cortical activity during the motor task execution of the nine gestures, this information corresponds to 32 superficial electrodes of Electroencephalography (EEG), under the international system 10/20 standard, with a sample frequency of 1000Hz, which will allow us, in the future, to combine EMG and EEG features for detecting the motor tasks based on Brain Computer Interfaces (BCI) [4, 14].

References

[1] Reaz, M., Hussain, M., and Mohd-Yasin, F. Techniques of emg signal analysis: detection, processing, classication and applications. Biological

[1] procedures online 8, 1 (2006), 11-35.

[2] Pantoja, W. C. Neuroprotesis: Hacia un futuro de cuerpos cibernéticos. Ingeniería Industrial, 26 (2008),251-266.

[3] Forero, A., Gonzalez, B., Reyes, C., Arango, D., and Avila, E. Diseño y construcción de un electro-miógrafo con su interfaz de visualizacion en matlab.

[4] Mesa, I., Rubio, A., Tubia, I., De No, J., and Diaz, J. Channel and feature selection for a surface electromyographic pattern recognition task. Expert Systems with Applications 41, 11 (2014), 5190-5200.

[5] Pinzon, J. V., Mayorga, R. P., and Hurtado, G. C. Brazo robotico controlado por electromiografía. Scientia Et Technica 1, 52 (2012), 165-173.

[6] Jerbi, K., Vidal, J., Mattout, J., Maby, E., Lecaignard, F., Ossandon, T., Hamamé, C., Dalal, S., Bouet, R., Lachaux, J.-P., et al. Inferring hand movement kinematics from meg, eeg and intracranial eeg: From brain-machine interfaces to motor rehabilitation. IRBM 32, 1 (2011), 8-18.

[7] Kabacoff, R. R in action: data analysis and graphics with R. Manning Publications Co., 2015.

[8] Hartigan, J. A., and Wong, M. A. Algorithm as 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28, 1 (1979), 100-108.

[9] Bhargav, S., and Pawar, M. A review of clustering methods forming non-convex clusters with, missing and noisy data.

[10] Mulroy, S., Gronley, J., Weiss, W., Newsam, C., and Perry, J. Use of cluster analysis for gait pattern classication of patients in the early and late recovery phases following stroke. Gait & posture 18, 1(2003), 114-125.

[11] Hartigan, J. A., and Wong, M. A. Algorithm as 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28, 1 (1979), 100-108.

[12] Liu, Y. H., Lin, L. F., Chou, C. W., Chang, Y., Hsiao, Y. T., & Hsu, W. C. (2018). Analysis of Electroencephalography Event-Related Desynchronisation and Synchronisation Induced by Lower-Limb Stepping Motor Imagery. Journal of Medical and Biological Engineering, 1-16.

[13] Zhang, Y., Liu, B., Ji, X., & Huang, D. (2017). Classification of EEG signals based on autoregressive model and wavelet packet decomposition. Neural Processing Letters, 45(2), 365-378

[14] Asanza, V., Pelaez, E., & Loayza, F. (2017, October). EEG signal clustering for motor and imaginary motor tasks on hands and feet. In Ecuador Technical Chapters Meeting (ETCM), 2017 IEEE (pp. 1-5). IEEE.