10.2478/acsc-2013-0023

Archives of Control Sciences
Volume 23(LIX), 2013
No. 4, *pages 381–393*

Investigation of probability density functions in modeling sample distribution of surface electromyographic (sEMG) signals

IGOR DA GUIA ROSA, MARCO ANTONIO CAVALCANTI GARCIA and MARCIO NOGUEIRA DE SOUZA

The surface electromyography signal (sEMG) has been typically modeled as a Gaussian random process. However, some authors have reported that the probability density functions (*pdfs*) associated with the sample distribution of sEMG signal exhibits a more peaked shape than one could expected for a Gaussian *pdf*. This work aimed to reinvestigate the profile of the sEMG *pdfs* during five different load levels of isometric contractions of *biceps brachii* muscle, and compared the adequacy of four different *pdfs* (Gaussian, Logistic, Cauchy, and Laplacian) in describing the sample distribution of such signal. Experimental *pdfs* were estimated for each subject and load condition. The comparison between experimental *pdfs* obtained from sEMG data of forty volunteers and four theoretical *pdfs* was performed by fitting these functions to its experimental counterpart, and using a mean absolute errors in the assessment of the best fit. On average, the Logistic *pdf* seemed to be the best one to describe the sample distribution of sEMG signal, although the probabilistic results, considering binomial trials, were significant for both Gaussian and Logistic *pdfs*.

Key words: electromyography; EMG; surface EMG; sEMG; EMG onset detection

1. Introduction

Along the last 70 years many studies have been dedicated to obtain, decode, and interpret the information content of different non-invasive biological signals. However, it is still a challenge to understand how a signal collected on the body surface reflects the most significant aspects concerning its genesis, as well as how it is affected by mechanisms related to surrounded tissues or organ.

Mathematical modeling of interesting signals is the ordinary approach to improve our understanding of the intrinsic biological phenomenon [1,2]. For instance, mathemat-

The Authors are with Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil. I.G. Rosa and M.N. Souza are with Laboratório de Instrumentação Biomédica. M.A.C. Garcia, the corresponding author, is with Núcleo de Estudos do Movimento Humano (NEMoH) – Departamento de Biociências da Atividade Física – Escola de Educação Física e Desportos (EEFD), email: garcia@ufrj.br

This study was partially supported by the Brazilian Research Council (CNPq), CAPES, FAPERJ and PRONEX.

Received 9.05.2013.

ical parameters derived from the surface electromyographic (sEMG) signal have been widely used for both clinicians and researchers in evaluating the normal and abnormal behavior of a muscle contraction [3]. Moreover, by using such mathematical parameters the user must consider that the sEMG signal is influenced by many factors, e.g., level of effort, the muscle length and muscle fibers architecture, fatigue, electrodes positioning, and others [1,2,4]. Due to these aspects, some authors have characterized the sEMG signal by means of statistical properties and frequency contents [5,6], which are relevant in signal modeling [1,2]. In this aspect the statistical distribution associated with the sampled sEMG signal has been widely adopted [7-9].

Despite the importance of the time domain data in the design of algorithms to detecting the onset, offset, and maintenance of a muscle contraction from the sEMG signal, there is no consensus in the literature about the best probability density function (*pdf*) used to describe the behavior of this characteristic of the signal. Different results have been reported concerning the agreement between experimental results and some *pdfs* adopted to describe the sampled sEMG signal. Some authors have suggested the Gaussian *pdf* as the best option in representing the sample distribution of any EMG signal [10-12], including invasive ones (iEMG), whereas some other authors support different *pdfs* [13,14]. The investigation concerning the best *pdf* to describe the distribution of the sampled sEMG signals becomes relevant because it can help to improve algorithms of onset detection applied in neuroprosthesis [15,16] and biofeedback [7], and can improve other important applications.

Some authors have investigated the suitability of the Gaussian *pdf* in describing the sample distribution of the iEMG signal and concluded that it might be fitted by some other peaked *pdf*. Milner-Brown and Stein [17] observed that the sample distribution of the iEMG signal of the first dorsal interosseous muscle in a condition of constant force (isometric contraction) and angled contraction presented a pattern that is more sharp and peaked around the zero mean than the one predicted by a Gaussian *pdf*. Moreover, this peak in the sample distribution seemed to be less pronounced at higher muscle force levels. On the other hand, Parker et al. [18] recorded iEMG signals from the *biceps brachii* muscle during two different low levels of muscle contraction and concluded, after a comparison performed graphically, that the sample distribution is reasonably modeled by a Gaussian *pdf*. Hunter et al. [19] also examined sEMG signals from the *biceps brachii* muscle during isometric and non-fatiguing contractions at 30% of the maximum voluntary contraction (MVC). After performing a graphical comparison, the authors reported that the shape of the experimental sample distribution was considerably different from a Gaussian one, being more peaked than a normal *pdf* around zero mean.

Bilodeau et al. [20] evaluated the sEMG data normality by means the Shapiro-Wilk test and found that this signal presents, in general, a non-Gaussian sample distribution, being more peaked around the zero mean. The authors [20] performed their study considering isometric and non-fatiguing contractions at 20%, 40% 60% and 80% of the MVC and non-fatiguing but slowly-force-varying contractions as well. They observed that the peaking of the sample distribution was less pronounced at higher levels of muscle contraction.

Clancy and Hogan [14] also investigated the sEMG samples distribution and concluded that experimental distributions may present a shape between a Gaussian *pdf* and a Laplacian *pdf*, with the former being the one that, on average, presents the best fitting on average. However, they used the area of the differences between histograms obtained from the experimental data and the studied *pdfs* to the assessment of the best fitting. We argue that such metric is not supported in the probability theory because it is supposed that the area below any *pdf* should be unitary.

Therefore, due to the controversy observed in the literature, the present work aimed to reinvestigate the sample distribution of sEMG signals and evaluate the adequacy of four *pdfs* in describing the behavior of this characteristic of the electromyographic signal. Instead of using the differences between areas as some previous works [14], we will assess the mentioned adequacy by means of a conservative metric that is based on the absolute value the error between the histogram obtained from the experimental data and the *pdf*.

2. Methods

2.1. Subjects

Forty healthy volunteers (twenty males aged 25.9 ± 5.53 years and twenty females aged 25.0 ± 7.48 years), all right-handed, and undergraduate students from the Physical Education School of the Federal University of Rio de Janeiro, without neuromuscular or orthopaedic diseases, participated in this study. The study was submitted to the Ethical Committee of the Federal University of Rio de Janeiro and was performed after the volunteers gave their written and informed consent.

2.2. Procedures for the sEMG signal acquisition

The acquisition system was based on a personal computer and an A/D converter DaqPad 1200 (NATIONAL INSTRUMENTS, TX-EUA) of 12 bits. The acquisition of sEMG signal was accomplished by using a custom-made signal conditioner developed at the Laboratory of Biomechanics of the Physical Education School of the Federal University of Rio de Janeiro following the Surface Electromyography for the Non-Invasive Assessment of Muscles (SENIAM) recommendations [1,2]. The most significant characteristics of the system were a common mode rejection ratio (CMRR), gain and frequency bandwidth that were set to 106 dB, 1000 and 10-500 Hz, respectively. The software for acquisition and signal processing was designed in LabVIEW 5.0 (NATIONAL INSTRUMENTS, TX-EUA). The sample frequency was set to 2 kHz and surface electrodes (Ag-AgCl; 1 cm diameter, Medtrace 200; Kendall, Canada) were used for collecting the sEMG signal.

An apparatus specially designed for the experiment and a dynamometer system (KRATOS DINAMMETROS LTDA., Brazil) were used for supporting the right superior arm and collecting the data of muscle force during the tests (Fig. 1). Both devices permitted individual adjustments to keep the right shoulder abducted in 70^o angle and the cable of the dynamometer perpendicular to the right forearm.

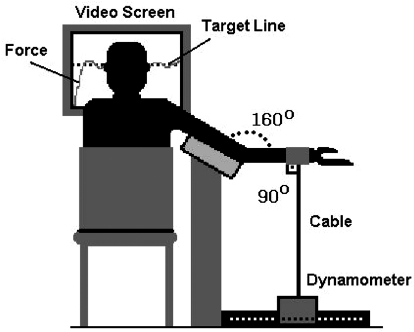

Figure 1. Posterior view of the apparatus, the video screen, and the dynamometer system for collecting the muscle force and the sEMG signals. Individual adjustments were possible in angles and distance from the monitor before collecting the data.

The investigation was performed with sEMG signals collected from the right *biceps brachii* muscle and that provided a database for further evaluation of the adequacy fitting associated with the four studied *pdfs*.

Before collecting of the sEMG signals, the volunteers were sat on a chair, with hips and knees positioned at approximately 90^o angle. The MVC of each volunteer was determined for further estimation of four different and arbitrary loads levels: 20%, 40%, 60%, and 80%, including 100% of MVC. Each test consisted of isometric and non-fatiguing contractions, as illustrated in the apparatus in Fig. 1. A visual feedback was supplied to each subject (by means of a video screen) to help the volunteer to sustain the requested percentile level of the MVC during 6 s. All the relative load levels were randomized previously to data collection and evaluated in the same experimental session. A resting period of about two minutes was observed between each trial.

To collect the sEMG signal, the skin was shaved and cleaned with neutral soap. The surface electrodes were placed on the right *biceps brachii* muscle, following the SENIAM recommendations [2]. The reference electrode was placed on the lateral *epicondilus* of *humerus*.

2.3. Estimation and comparison of the studied *pdfs*

The samples of the experimental sEMG signals were normalized (mean adjusted to zero and standard deviation adjusted to one) according to the equation (1) and then a sample histogram was obtained. The histograms used 500 bins equally spaced over a normalized range.

$$x_a = \frac{x-\mu}{\sigma} \tag{1}$$

where x_a and x are the normalized and the raw sample of the sEMG signal for each subject, respectively; μ and σ are the mean and the standard deviation of the raw sEMG signal, respectively.

For the same experimental condition, a mean histogram for all forty volunteers was calculated as follows

$$\overline{N}(j) = \frac{1}{K}\sum_{i=1}^{K} N_i(j) \tag{2}$$

where $\overline{N}$ is the mean histogram, N is a subject histogram, K is the number of subjects, and j is the bin number ($j = 1...500$).

The so-called experimental *pdfs* were estimated from each mean histogram, and similarly from the histograms of each subject separately. These estimations were obtained through equation (3), which guarantees unitary area below the *pdf* curve [21].

$$p' = \frac{\overline{N}}{LW} \tag{3}$$

where p' is the experimental *pdf*, L is the length of the raw sEMG signal (x), and W is the width of the bin.

From the sEMG database we obtained five compound signals grouping signals acquired at some given relative load levels. These compounded signals were constituted, respectively, by the sEMG signals corresponding to 20% of the MVC of all the subjects; from the signals corresponding from 20% to 40% MVC; from those corresponding from 20% to 60% MVC; from those concerning from 20% to 80% MVC; and from the ones representing MVC from 20% to 100%. For each compound signal an experimental *pdf* was estimated as described previously.

Five mean signals were also derived for each percentile load level (20%, 40%, 60%, 80% and 100%), and thus theirs experimental *pdfs* were also estimated. Besides, the sEMG signal from each subject, at each load level, had its experimental *pdf* individually estimated.

For each experimental *pdf* four theoretical *pdfs* (Gaussian, Logistic, Cauchy and Laplacian) were obtained. The *pdfs* were ordered from the less to the most peaked curve, respectively, what is shown in the associated following equations being ordered in the same way.

$$p_1(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)} \tag{4}$$

where μ is the mean value and σ is the standard deviation.

$$p_2(x) = \frac{e^{\left[\frac{-(x-\alpha)}{\beta}\right]}}{\beta\left[1+e^{\frac{-(x-\alpha)}{\beta}}\right]^2} \tag{5}$$

where β is the scale and α is the location parameter.

$$p_3(x) = \frac{b}{\pi[(x-a)^2 + b^2]} \tag{6}$$

where b and a are the scale and the location parameters, respectively.

$$p_4(x) = \frac{1}{2} e^{\frac{-|x-l|}{s}} \tag{7}$$

where s and l are the shape and the location parameters, respectively.

The assessment of the best fitting among the tested theoretical *pdfs* was done by the mean absolute error (MAE) between an experimental and a candidate *pdf* as shown in equation (8). The best fitting presents the least error. The mean and standard deviation (SD) of the MAE feature for each fitted *pdf* were compared and statistically analyzed.

$$\text{MAE} = \frac{1}{M} \sum_{i=1}^{M} \left| y_{\text{exp}_i} - y_{theo_i} \right| \tag{8}$$

where y_{exp} is the experimental probability density value, y_{theo} is the theoretical probability density value, and M is the number of discrete samples points of each *pdf*.

Finally, we computed the number of times that each theoretical *pdf* best fitted each one of the forty experimental *pdfs*, for each percentile load level.

2.4. Statistical analysis

One-way analysis of variance (One-way ANOVA) followed by Tukey HSD post hoc was used to compare the results of the MAE feature. The experimental *pdf* estimated from the histogram of each sampled sEMG signal was fitted by the four studied theoretical *pdfs* and its corresponding MAE features calculated as previously presented. These forty MAE values were compared and statistically analyzed for each load level condition, where the level of significance (α) was set at 5%.

For counting the best fits of each theoretical *pdf* the probability of having the same or more number of successes was computed. It was considered as the binomial process [14] with a given probability of success of 0.25.

3. Results

The profile of the mean experimental sEMG *pdf*, when different fractional ranges of the percentile load condition are considered, can be observed in Fig. 2, where the five compound signals are shown (from 20% to k times 20% MVC, with $k = 1, 2, 3, 4, 5$).

The estimate sEMG *pdf* for each percentile load level and the best fits of the experimental data for each theoretical *pdf* can be observed in Fig. 3. Contrary to the *pdfs* observed in Fig. 2, where ranges of the percentile load levels were considered, each

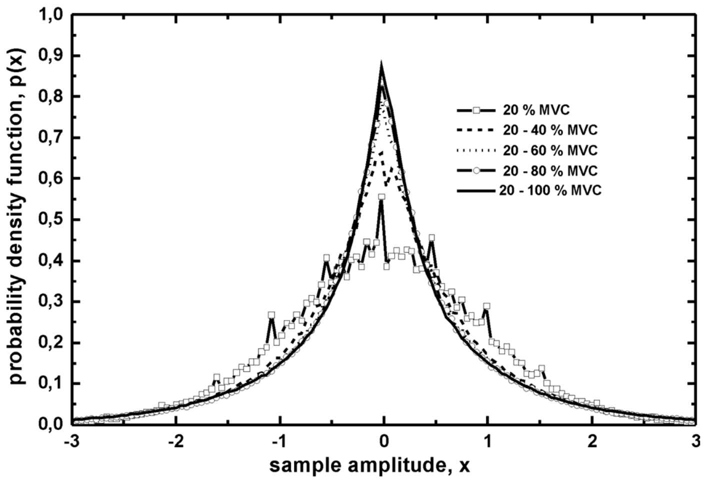

Figure 2. Changes in the shape of the mean sample distribution of sEMG signals associated with different ranges of percentile load levels of the MVC.

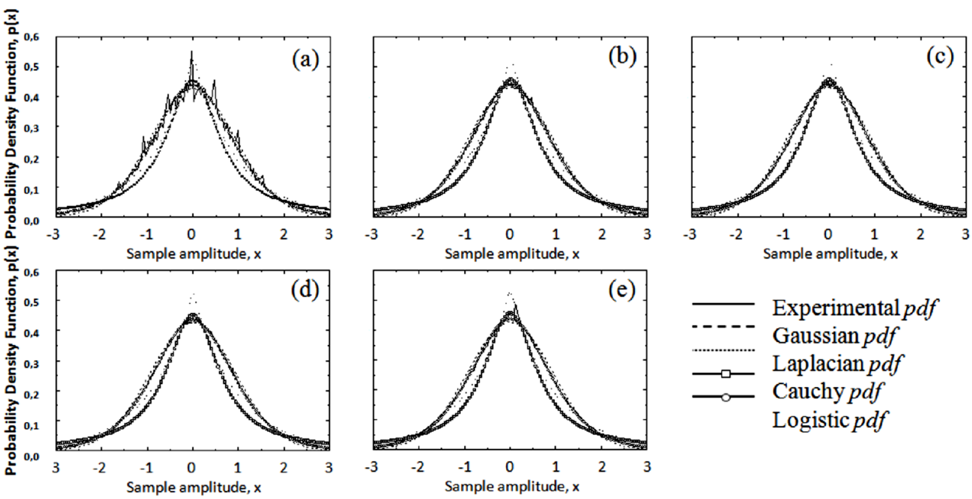

Figure 3. Changes in shape of the mean distribution of sEMG signals amplitudes associated with different percentile load level conditions: (a) 20%, (b) 40%, (c) 60%, (d) 80%, and (e) 100% of the MVC.

graph of Fig. 3 represents the *pdf* obtained from the sEMG signal of a given percentile load, not a range of it.

The agreement between the experimental estimated *pdf* and the four studied theoretical *pdfs* can be observed in Tables 1 to 5, where the mean and SD of the MAE feature of each theoretical *pdf* best fit are presented for each percentile load level condition.

Table 1. MAE for 20% MVC

Theoretical *pdf*	Mean	SD ($\times 10^{-3}$)
Gaussian	0.0036	0.1003
Laplacian	0.0081	0.2752
Cauchy	0.0129	0.4244
Logistic	0.0027	0.0901

Table 2. MAE for 40% MVC

Theoretical *pdf*	Mean	SD ($\times 10^{-3}$)
Gaussian	0.0025	0.0250
Laplacian	0.0075	0.1896
Cauchy	0.0123	0.3348
Logistic	0.0012	0.0075

Table 3. MAE for 60% MVC

Theoretical *pdf*	Mean	SD ($\times 10^{-3}$)
Gaussian	0.0024	0.0222
Laplacian	0.0076	0.1957
Cauchy	0.0123	0.3350
Logistic	0.0009	0.0045

Table 4. MAE for 80% MVC

Theoretical *pdf*	Mean	SD ($\times 10^{-3}$)
Gaussian	0.0022	0.0199
Laplacian	0.0077	0.1893
Cauchy	0.0124	0.3357
Logistic	0.0011	0.0048

Table 5. MAE for 80% MVC

Theoretical *pdf*	Mean	SD ($\times 10^{-3}$)
Gaussian	0.0028	0.0382
Laplacian	0.0071	0.1545
Cauchy	0.0122	0.3061
Logistic	0.0015	0.0141

The profile of the MAE feature of each theoretical *pdf* considering each percentile load condition is presented in Fig. 4, where each load level was considered separately.

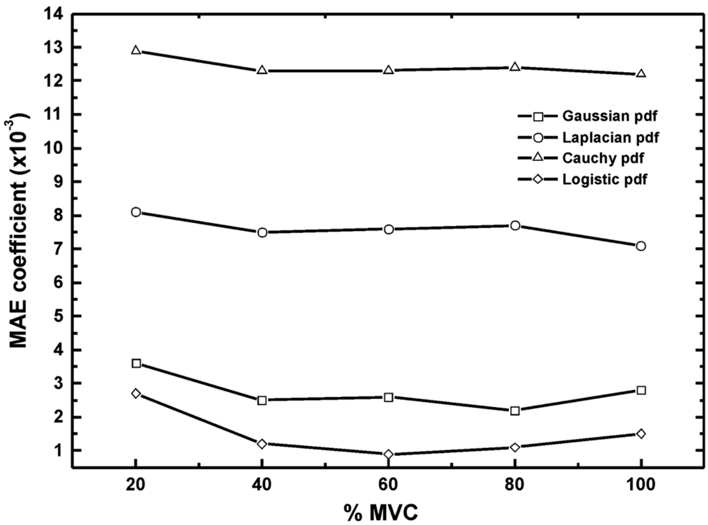

Figure 4. Behavior of the MAE feature of each theoretical *pdf* considering each percentile load level condition.

On the other hand, the behavior of the MAE for each theoretical *pdf* considering the fractional ranges of the muscular contraction is shown in Fig. 5.

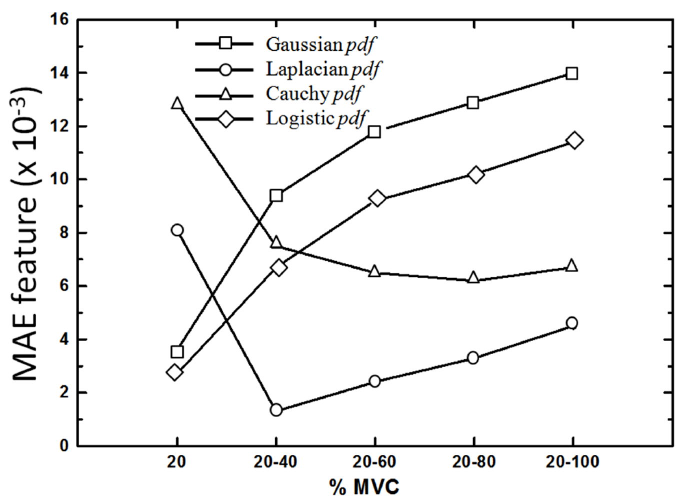

Figure 5. Behavior of the MAE feature of each theoretical *pdf* considering the fractional ranges of the muscle contraction at different percentile load levels.

Tab. 6 resumes the degree of the agreement between the experimental estimated sEMG sample distribution and each one of the four considered theoretical *pdfs*.

Table 6. Number of best fits of each theoretical *pdf* in all different load levels of contractions for each subject (between brackets: probability of having the observed or more number of successes in forty trials for each load) (**G**: Gaussian; **Lo**: Logistic; **C**: Cauchy; **La**: Laplacian)

	% of MVC				
pdf	20	40	60	80	100
G	18 (0.0046)	12 (0.2848)	15 (0.0544)	19 (0.0017)	19 (0.0017)
Lo	21 (0.0001)	27 (1.87×10^{-8})	24 (2.82×10^{-6})	20 (5.72×10^{-6})	18 (0.0046)
C	1 (0.999)	1 (0.999)	0 (1)	0 (1)	1 (0.999)
La	0 (1)	1 (0.999)	1 (0.999)	1 (0.999)	2 (0.999)

4. Discussion

According to the results presented in Fig. 4, there are evidences that the sEMG *pdf* cannot be so peaked as some previous studies have reported [17,19,20]. Most the individual experimentally sample distributions of the sEMG signal could be modeled by the one *pdf* that presents a behavior between the shape of a Gaussian and a Logistic *pdf*.

Both findings obtained graphically and by means of the MAE feature seemed to point to the Logistic *pdf* as the theoretical *pdf* that best fit the experimental-data distribution. However, results from ANOVA suggested equivalence between the Gaussian and Logistic *pdf*, i.e., they are not statistically different, whereas the results were statistically different in what concerns the Laplacian and Cauchy *pdfs*. The probabilities obtained when the numbers of best fitting of each theoretical *pdf* were considered binomial trials also indicated that the result is statistically significant for Gaussian and Logistic *pdf*.

It was also observed the importance of the signal-to-noise ratio (SNR) of the sEMG signal in the fitting of a theoretical *pdf*. Experimental sEMG signal can be considered the composition of two distinct random processes, one due to the interference pattern from the activity of the recruited motor units, and the other the result of all instrumentation noises (normally considered a Gaussian noise). Therefore, when the SNR decreases, the resulting *pdf* seems to be more influenced by the contribution of the noise, which has a Gaussian *pdf*. On the other hand, when the SNR increases, the result is a peaked and thinner *pdf*. This hypothesis must be further investigated using a sEMG signal model.

5. Conclusion

The results of the present study were also similar to those observed in some previous investigations [14,17,19,20]. Peaked *pdfs* around the zero mean are observed in sample distributions of sEMG signal, when different muscular loading conditions were performed. In the present work four theoretical *pdfs* were investigated in the fitting process of sample distribution of the sEMG signal and the corresponding profiles were discussed. The Logistic *pdf* seems to be a better option than the Gaussian one to model the sample distribution of the sEMG signal. However, to perform a better comparison of individual results, the signal must be acquired at a higher sample frequency, in order to reduce the fluctuation of the experimental estimated *pdfs*.

Since a given theoretical sEMG *pdf* may be assumed in sEMG data analyzers and muscle activity detection techniques, these results can contribute to the development of a new approaches with the same purposes. In a similar view, these findings can guide the development of sEMG simulators that could more accurately mimic experimental signals. Such simulators could improve the comparison among several muscular contraction detection techniques, since any specific sEMG signal database would not need to be used. Moreover, if researchers in sEMG signal-processing field would adopt such standard procedure some contradictory results reported could be avoided.

References

[1] R. Merletti and P.A. Parker: Electromyography: Physiology, and oninvasive Applications. John Wiley & Sons, Inc., Hoboken, N.J. 2004.

[2] H.J. Hermens, B. Freriks, R. Merletti, D.A. Stegeman, J. Blok, G. Rau, C. Disselhorst-Klug and G. Hǧg: SENIAM 8: European Recommendations for Surface Electromyography. Roessingh Research and Development b.v., Enchede 1999.

[3] S. Micera, G. Vannozzi, A.M. Sabatini and P. Dario: Improving detection of muscle activation intervals. *IEEE Engineering, Medicine & Biology Magazine*, **20**(6), (2001), 38-46.

[4] J.R. Cram and G. S. Kasman: Instrumentation, in Introduction to Surface Electromyography, Chap. 3. Aspen Publishers, Inc., Gaithersburg, Maryland 1998.

[5] G. Rasool, K. Iqbal and G.A. White: Myoelectric activity detection during a Sit-to-Stand movement using threshold methods. *Computers and Mathematics with Applications*, **64**(5), (2012), 1473-1483.

[6] K. Englehart, B. Hudgins, P. Parker and M. Stevenson: Time-frequency representation for classification of the transient myoelectric signal. *Proc. 20th Annual Int. Conf. IEEE/EMBS*, **5** (1998), 2627-2630.

[7] G. SEVERINI, S. CONFORTO, M. SCHMID and T. D'ALESSIO: Novel formulation of a double threshold algorithm for the estimation of muscle activation intervals designed for variable SNR environments. *J. Electromyography and Kinesiology*, **22**(6), (2012), 878-885.

[8] X. LI, P. ZHOU and A.S. ARUIN: Teager-Kaiser energy operation of surface EMG improves muscle activity onset detection. *Annals of Biomedical Engineering*, **35**(9), (2007), 1532-1538.

[9] P.W. HODGES and B.H. BUI: A comparison of computer-based methods for the determination of onset of muscle contraction using electromyography. *Electroencephalography and Clinical Neurophysiology - electromyography and Motor Control*, **101**(6), (1996), 511-519.

[10] J.H. ABBINK, A. VAN DER BILT and H.W. VAN DER GLAS: Detection of onset and termination of muscle activity in surface electromyograms. *J. Oral Rehabilitation*, **25**(5), (1998), 365-369.

[11] P. BONATO, T. D'ALESSIO and M. KNAFLITZ: A statistical method for the measurement of muscle activation intervals from surface myoelectric signal during gait. *IEEE Trans. in Biomedical Engineering*, **45**(3), (1998), 287-299.

[12] G. STAUDE and W. WOLF: Objective motor response onset detection in surface myoelectric signals. *Medical Engineering & Physics*, **21**(6-7), (1999), 449-467.

[13] E.A. CLANCY: Electromyogram amplitude estimation with adaptive smoothing window length. *IEEE Trans. in Biomedical Engineering*, **46**(6), (1999), 717-729.

[14] E.A. CLANCY and N. HOGAN: Probability density of the surface electromyogram and its relation to amplitude detectors. *IEEE Trans. in Biomedical Engineering*, **46**(6), (1999), 730-739.

[15] S. MICERA, J. CARPANETO and S. RASPOPOVIC: Control of hand prostheses using peripheral information. *IEEE Reviews in Biomedical Engineering*, **3** (2010), 48-68.

[16] M. ZECCA, S. MICERA, M.C. CARROZZA and P. DARIO: Control of multifunctional prosthetic hands by processing the electromyographic signal. *Critical Reviews in Biomedical Engineering*, **30**(4-6), (2002), 459-485.

[17] H.S. MILNER-BROWN and R.B. STEIN: The relation between the surface electromyogram and muscular force. *J. Physiology*, **246**(3), (1975), 549-569.

[18] P.A. PARKER, J.A. STULLER and R.N. SCOTT: Signal processing for the multistate myoelectric channel. *Proc. IEEE*, **65**(5), (1977), 662-674.

[19] I.W. HUNTER, R.E. KEARNEY and L.A. JONES: Estimation of the conduction velocity of muscle action potentials using phase and impulse response function techniques. *Medical & Biological Engineering & Computing*, **25**(2), (1987), 121-126.

[20] M. BILODEAU, M. CINCERA, A.B. ARSENAULT and D. GRAVEL: Normality and stationarity of EMG signals of elbow flexor muscles during ramp and step isometric contractions. *J. Electromyography and Kinesiology*, **7**(2), (1997), 87-96.

[21] J. BENDAT J AND A. PIERSOL: Random Data, Analysis and Measurement Procedures. John Wiley & Sons, Inc., Hoboken, N. J. 2000.