Robust Functional Statistics applied to Probability Density Function Shape screening of sEMG data

S. BOUDAOUD, H. RIX, IEEE *Member*, M. AL HARRACH, and F. MARIN

***Abstract*— Recent studies pointed out possible shape modifications of the Probability Density Function (PDF) of surface electromyographical (sEMG) data according to several contexts like fatigue and muscle force increase. Following this idea, criteria have been proposed to monitor these shape modifications mainly using High Order Statistics (HOS) parameters like skewness and kurtosis. In experimental conditions, these parameters are confronted with small sample size in the estimation process. This small sample size induces errors in the estimated HOS parameters restraining real-time and precise sEMG PDF shape monitoring. Recently, a functional formalism, the Core Shape Model (CSM), has been used to analyse shape modifications of PDF curves. In this work, taking inspiration from CSM method, robust functional statistics are proposed to emulate both skewness and kurtosis behaviors. These functional statistics combine both kernel density estimation and PDF shape distances to evaluate shape modifications even in presence of small sample size. Then, the proposed statistics are tested, using Monte Carlo simulations, on both normal and Log-normal PDFs that mimic observed sEMG PDF shape behavior during muscle contraction. According to the obtained results, the functional statistics seem to be more robust than HOS parameters to small sample size effect and more accurate in sEMG PDF shape screening applications.**

I. INTRODUCTION

Classically, muscle activation and fatigue are often evaluated using time parameters as Averaged Rectified Value (ARV) and Root Mean Square (RMS) value and frequential parameters as Mean Frequency (MNF) and Median Frequency (MDF) that are mainly linked to energetic dynamics of the sEMG signals. Recently, another signal representation, namely, the amplitude PDF, has been explored and obtained promising results concerning the evaluation of both fatigue [1, 2] and muscle activation [3, 4] with force increase. These amplitude statistics are derived from HOS parameters (skewness and kurtosis) and also from a recent functional formalism, from the Functional Data Analysis (FDA) community. It allows a shape screening of the sEMG PDF [5], [6]. Unfortunately, in experimental conditions, these statistical parameters need large sample size for accurate PDF estimation that strongly restrains real-time or fine time resolution PDF shape screening [3]. In this work and by taking inspiration from the CSM formalism, robust functional statistics are proposed. They measure localized shape distances between an observed PDF curve, estimated by kernel density estimation [7] from sEMG data and a reference one, the Normal PDF shape, to evaluate a possible departure from Gaussianity. After describing their formalism, Monte-Carlo simulations, using small sample size (500 samples) are launched to compare these statistics to HOS ones, according to several PDF shapes simulated using centered Log-Normal random processes. These PDF shapes mimic the ones obtained using Laplacian electrode arrangement and recently observed in both simulation [3] and experimentation [8]. Finally, the obtained results are discussed and some perspectives exposed.

This work was carried out and funded in the framework of the Labex MS2T. It was supported by the French Government, through the program "Investments for the future" managed by the National Agency for Research (Reference ANR-11-IDEX-0004-02).

S. Boudaoud, M. AL Harrach, and F. Marin are with Université de Technologie de Compiègne, CNRS UMR 7338 BMBI, 60200 Compiègne, France, (e-mail : sofiane.boudaoud@utc.fr, mariam.al-harrach@utc.fr, frederic.marin@utc.fr).

H. Rix is with Université de Nice Sophia-Antipolis, CNRS UMR6070 I3S, 06903 Sophia Antipolis Cedex – France. (e-mail: rix@i3s.unice.fr)

II. METHODS

A. Robust functional statistics

The CSM method is a recent formalism that allows, using function estimation procedure, the evaluation of subtle shape modifications in a set of curves [5]. It has been used with success to evaluate deterministic signals [5,6] or random ones [3]. By taking inspiration from this formalism (with polynomial order equals 1), we propose metrics that measure the distance of a PDF from Gaussianity including asymmetry and flatness measures as for the HOS parameters. Assume that there are two PDFs $p(x)$ and $g(x)$ that represent a random sEMG signal PDF and a normal PDF respectively defined on $[-l,l]$. The Normal PDF $g(x)$ is obtained using its analytic definition. The sEMG PDF $p(x)$ is obtained using a kernel estimation procedure [7] to obtain a smooth continuous (better abscissa resolution) curve less noisy than the classical histogram. Their normalized integrals (distribution function in F) can be defined as [5]:

$$P(x) = \int_{a_p}^{x} p(u)du \Big/ \int_{a_p}^{b_p} p(u)du, \qquad (1)$$

$$G(x) = \int_{a_g}^{x} g(u)du \Big/ \int_{a_g}^{b_g} g(u)du$$

Where $[a_p, b_p] \subset [-l, l]$, $[a_g, b_g] \subset [-l, l]$ are the nonzero supports of $p(x)$ and $g(x)$ respectively. The obtained distribution functions $P(x)$ and $G(x)$ can typically be linked through a warping function expressed as φ or ψ. Specifically, we can write:

$$P = G \circ \varphi, \qquad G = P \circ \psi, \tag{2}$$

where $P = G \circ \varphi_i$ is a shorthand notation for the composition function $P(x) = G(\varphi(x))$. The time warping function $\psi = \varphi^{-1}$ links $P(x)$ to $G(x)$ and represents the fluctuations in both shape and abscissa support [5]. The PDF shape analysis, aims at separating intrinsic shape variation from those caused by first and second moment variability. To reach this objective, we propose a representation of φ as [6]:

$$\varphi = v \circ A, \qquad \psi = A^{-1} \circ w, \quad w = v^{-1} \tag{3}$$

where $A(x) = \alpha x + \beta$, $\alpha \in \mathbb{R}^+, \beta \in \mathbb{R}$ is an affine function (polynomial function of order 1) that accounts for mean and variance variability of $p(x)$. The second element w is a monotonically increasing nonlinear function that represents shape fluctuations on a constant support. Therefore, we can rewrite (3) as follows in F and F^{-1} respectively [6]:

$$\begin{aligned} &P = G \circ v \circ A, \quad G = P \circ A^{-1} \circ w \\ &P^{-1} = A^{-1} \circ w \circ G^{-1}, \quad A \circ P^{-1} = w \circ G^{-1} \end{aligned} \tag{4}$$

Where $w(y) = y + n(y)$, $y \in [0,1]$ and $n(y)$ is a function that accounts for small intrinsic shape modification and modeling the departure from linearity of w. By replacing this last equation in (4), we obtain:

$$\alpha P^{-1}(y) + \beta = G^{-1}(y) + n(G^{-1}(y)), \; y \in [0,1] \tag{5}$$

where the term $n(G^{-1}(y))$ represents the shape difference between the realigned function $\hat{A}(P^{-1}(y))$ and $G^{-1}(y)$. The parameters α and β are estimated by constrained linear regression between $P^{-1}(y)$ and $G^{-1}(y)$ like in [6]. A simulation for illustration is depicted on Figure.1. One can observe the realignment procedure and the remaining shape difference between $\hat{A}(P^{-1}(y))$ and $G^{-1}(y)$ since $P^{-1}(y)$ is estimated from a Log-Normal random sequence.

Then, we propose three distances, namely, the Center Shape Distance (CSD), the Left Shape Distance (LSD) and the Right Shape Distance (RSD) that measure shape differences between realigned $\hat{A}(P^{-1}(y))$ and $G^{-1}(y)$ in the center, the

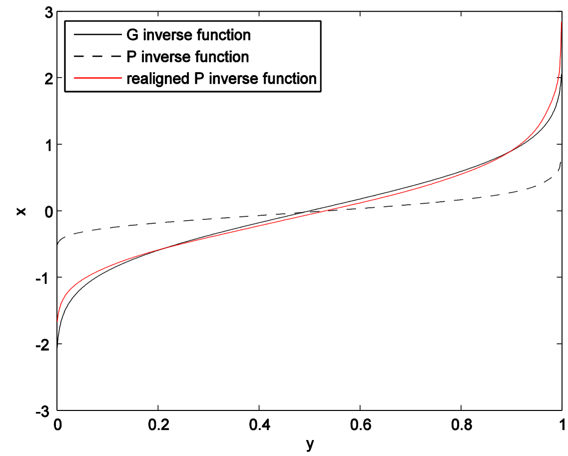

Figure 1. The functions $G^{-1}(y)$ (-), $P^{-1}(y)$ (Log-Normal 1 shape, - -), and realigned $\hat{A}(P^{-1}(y))$ (red line).

left and the right regions respectively. These distances are defined as follows:

$$\begin{aligned} CSD(p,g) &= \sqrt{\int_{0,4}^{0,6} \left(\hat{\alpha} P^{-1}(y) + \hat{\beta} - G^{-1}(y)\right)^2 dy}, \\ LSD(p,g) &= \sqrt{\int_{0}^{0,25} \left(\hat{\alpha} P^{-1}(y) + \hat{\beta} - G^{-1}(y)\right)^2 dy}, \\ RSD(p,g) &= \sqrt{\int_{0,75}^{1} \left(\hat{\alpha} P^{-1}(y) + \hat{\beta} - G^{-1}(y)\right)^2 dy} \end{aligned} \tag{6}$$

From these definitions, on can observe that the CSD distance is a metric that is sensitive to the flatness or peakedness (as the kurtosis) of $p(x)$ and observed in the middle region of its inverse distribution function $P^{-1}(y)$. The LSD and RSD distances scrutinize the departure from gaussianity in the left tail and right tail region of $p(x)$ (see Figure 1). These two distances, if they are not equal, can inform us about possible asymmetry of $p(x)$. One can note that the three proposed distances are positive if $p(x)$ is not a Gaussian PDF and equal to zero if $p(x)$ is a Gaussian PDF.

B. HOS parameters

Classically, PDF shape modifications are monitored with HOS parameters (skewness and kurtosis). We recall briefly the definitions of both normalized HOS parameters in the following equation for a random variable U (sEMG sample values):

$$Sk_U = \frac{E\left(U - \mu_U\right)^3}{\sigma_U^{\,3}}, \quad Kr_U = \frac{E\left(U - \mu_U\right)^4}{\sigma_U^{\,4}} - 3 \tag{7}$$

where $E(.)$ is the expectation operator and μ_U, σ_U are respectively the expected value and standard deviation of U.

These two HOS parameters are both invariant to variance and mean value variability, like the proposed functional distances. Both are also centered on the zero value and are signed parameters in the opposite of the proposed functional metrics.

III. SIMULATION & RESULTS

The proposed functional statistics are tested using Monte-Carlo simulations and their robustness evaluated against small sample size effect and compared to HOS parameters. For this purpose, four PDF shapes are simulated from random sequences of 500 samples. The first studied shape is the Normal (Gaussian) shape to evaluate the ability of both functional statistics and HOS to express the Gaussianity in drastic conditions. The three other PDF shapes are derived from centered Log-Normal distributed sequences using a specific random number generator. In fact, these PDF shapes have been already observed by both simulation [3] and experimentation [8] in sEMG signals recorded using Laplacian electrode arrangements. The four shapes and their respective analytical representations are exposed in Figure 2. Using this methodology, we generated, for each PDF shape, 1000 sequences of 500 samples with each sequence distributed according to the corresponding PDF law. Also, The HOS parameters are computed from these random sequences according to each PDF shapes. Then, smooth and continuous PDFs are estimated using kernel density estimation (*ksdensity* function in Matlab). Each estimated PDF is based on a normal kernel function, and is evaluated at 100 equally spaced points. For each sequence, the estimated PDF is then used for computing the CSD, LSD, and RSD distances following the formalism described earlier. After, mean and standard deviation, of each parameter and each studied PDF shape, are calculated and exposed in the Table.1 and Table. 2 for (skewness, LSD, RSD) and (kurtosis and CSD) respectively. A graphical illustration of the parameter trend and variability according to PDF shape is also provided using boxplot representation in Figure.3 and Figure. 4.

Concerning the parameters sensitive to PDF asymmetry, we can observe that functional statistics, especially the LSD parameter, outperform the skewness parameter in both mean value separation and standard deviation smallness according to PDF shape class as presented in Table.1. In fact, this means that, despite the small sample size (500 samples) used to describe the PDFs, the LSD parameter was able to better discriminate the four classes than the skewness parameter (see Figure 3). As for RSD and the skewness parameter, one can observe an increase with the class number indicating a departure from Gaussianity. In addition, the Normal PDF shape, LSD and RSD are, in average, equal but not null. This should be explained by the unsigned nature of these parameters, that induces a little bias, and by the symmetry of the Normal PDF. This bias is not present for the signed skewness which is well centered on zero. We can also observe, from the data in Table.1, an increase in the (μ_{LSD}-μ_{RSD}) mean difference, from the Normal shape toward the Log-Normal3 shape, indicating probably an asymmetry

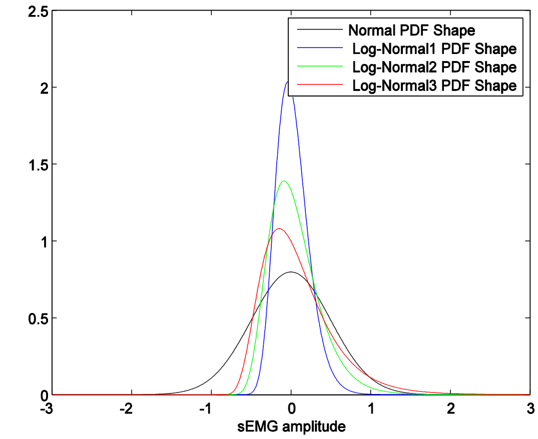

Figure 2. Analytical representations of the four studied PDF shapes.

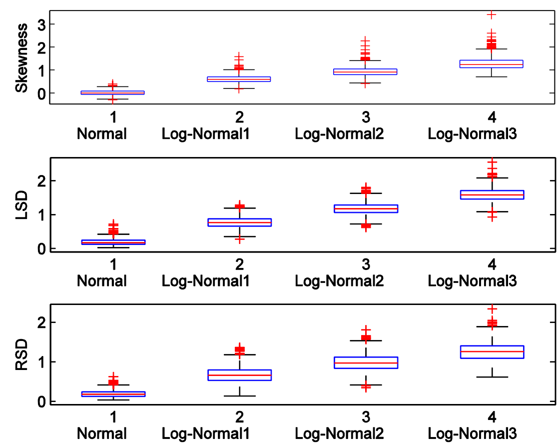

Figure 3. Robustness evaluation of the three parameters (skewness, LSD, and RSD) according to small sample size effect and PDF shape.

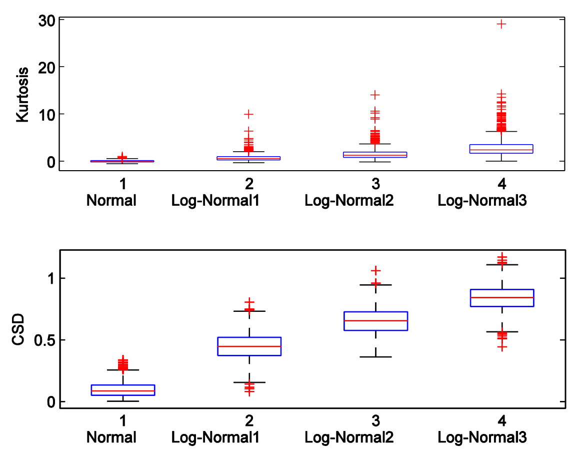

Figure 4. Robustness evaluation of the two parameters (kurtosis and CSD) according to small sample size effect and PDF shape.

TABLE I. ROBUSTNESS EVALUATION (MEAN AND STD OVER 1000 TRIALS)

PDF shapes	***RSD*** ***μ+σ (σ %)***	***LSD*** ***μ+σ (σ %)***	***Skewness*** ***μ+σ (σ %)***
Normal	0.19±0.09(*)	0.19±0.09(*)	0.01±0.11 (*)
Log-Normal 1	0.67±0.2(29)	0.77±0.16(21)	0.61±0.16(26)
Log-Normal 2	0.97±0.21(21)	1.17±0.18(15)	0.93±0.20(22)
Log-Normal 3	1.25±0.25(20)	1.59±0.21(13)	1.28±0.28(22)

(*) Not relevant

TABLE II. ROBUSTNESS EVALUATION (MEAN AND STD OVER 1000 TRIALS)

PDF shapes	***CSD*** ***μ+σ (σ %)***	***Kurtosis*** ***μ+σ (σ %)***
Normal	0.1±0.06(*)	-0.01±0.21(*)
Log-Normal 1	0.44±0.11(25)	0.66±0.72(110)
Log-Normal 2	0.65±0.11(17)	1.52±1.23(81)
Log-Normal 3	0.84±0.10(12)	2.93±2.10(72)

(*) Not relevant

increase validated with the increase of the mean skewness parameter.

For the parameters sensitive to PDF peakedness, the performances differences are more pronounced since the kurtosis is not able to discriminate the four class shapes due to an important standard deviation increase indicating a strong variability. In contrast, the CSD parameter remains stable with increasing the PDF shape deformation as depicted in both Table.2 and Figure.4. We can also observe a little bias as for the RSD and LSD parameters in evaluating the Normal shape for the same explained reason.

In addition, the kurtosis seems to be, by far, less robust to the sample size effect than the skewness according to the obtained results.

IV. DISCUSSIONS AND CONCLUSION

In this work, we proposed robust functional statistics to small sample size effect for evaluating four sEMG PDF shapes. For this purpose, these PDF shapes have been simulated from Normal law and Log-Normal law observed on experimental and simulated data recorded using Laplacian electrode arrangement. These metrics have been designed by taking inspiration from a recent functional formalism used in shape analysis, namely, the CSM formalism. The proposed algorithm is composed of two steps. A first step concerns the estimation of smooth and continuous (better abscissa resolution) PDF from a finite random sequence using kernel density estimation. The second step is the calculus of localized shape distances between a realigned curve (related to the PDF to evaluate) and a reference Normal curve both computed in the inverse distribution function domain. In other words, these shape distances allow the Gaussianity evaluation of a random sequence in both asymmetry and peakedness behaviors as for the HOS parameters.

After, the ability of these metrics to separate PDF shapes in drastic conditions (500 samples by random sequence) was assessed and compared to classical HOS parameters using Monte-Carlo simulations (1000 trials). According to the obtained results, these functional amplitude statistics seem to be more robust and discriminative according to four realistic sEMG PDF shape classes for both asymmetry and peakedness behaviors. In experimental conditions, this robustness should allow a finer time screening of sEMG PDF shape modifications since if the sEMG sampling frequency is classically around 2 kHz, one can obtain a maximum of 4 observation points per second using a windowing of 500 samples. This interesting property has to be verified, using rigorous statistical testing, in real conditions in future studies.

REFERENCES

[1] A. Holtermann, C. Grolund, J.S. Karlsson, and K. Roeleveld. "Motor unit synchronization during fatigue: Described with a novel SEMG method based on large motor unit samples." J Electromyo and Kinesiol, vol.19, pp. 232– 241, 2009.

[2] S. Boudaoud, F. Ayachi, and C. Marque. « Shape Analysis and Clustering of Surface EMG Data." 32nd Conf. of the IEEE EMBS, Buenos Aires, Argentina, Aug. 31 – Sept. 4, 2010.

[3] F Ayachi, S. Boudaoud, and C Marque "Evaluation of Muscle Force Classification Using Shape Analysis of the sEMG Probability Density Function: A simulation study," *Med. & Biol. Eng. & Comput.*, 2014 (accepted).

[4] K. Nazarpour, A. H. Al-Timemy, G. Bugmann, and A. Jackson, "A note on the probability distribution function of the surface electromyogram signal," *Brain Res. Bull.*, vol. 90, pp. 88–91, Jan. 2013.

[5] S. Boudaoud, H. Rix, and O. Meste. "Core shape modelling of a set of curves. "Comput. Stat. and Data Anal., vol. 54(2), pp. 308–325, 2010.

[6] S. Boudaoud, H. Rix, O. Meste, C. Heneghan, and C. O'Brien. "Corrected integral shape averaging applied to the detection of sleep apnea from the electrocardiogram." EURASIP J of Adv in Sig Proc, (Article ID 32570), vol. 2007, 2007. 12 pages.

[7] Bowman, A. W., and A. Azzalini. "Applied Smoothing Techniques for Data Analysis." New York: Oxford University Press Inc., 1997.

[8] M. Al Harrach, S. Boudaoud, D. Gamet, J.F. Grosset, and F. Marin. "Evaluation of High Order Statistic Trends from HD-sEMG recordings during Ramp Exercise" submitted to 36th IEEE EMBS Conf., Chicago, U.S.A, 2014.