Spectral Moments for Feature Extraction from Temporal Signals

Marko Vuskovic[1] and Sijiang Du[2]

[1] Department of Computer Science, SDSU,
San Diego, CA 92182-7720, USA
marko@cs.sdsu.edu

[2] Department of Computer Science, SDSU,
San Diego, CA 92182-7720, USA
sijiangdu@hotmail.com

Abstract

A new approach to computation of spectral moments of temporal signals is proposed. The approach is based on the auto correlation sequence of the original temporal signal, and makes use of the fact that the power spectral density is a discrete-time continuous-frequency function. The new approach offers more efficient generation of moments than the approaches based on numerical integration of the power spectral density function. The impact of noise is also analyzed, which was found to be very high at higher-order moments. Based on the analysis, a simple linear transformation of moments is suggested. It is shown that the new features are very little affected by additive white Gaussian noise.

Keyword: Feature extraction, spectral moments, temporal signals, EMG, pattern recognition.

1 Introduction

Usage of power spectral density functions (PSD) for extraction of features from temporal and spatial signals is a standard approach in various areas of pattern recognition, including acoustics, seismology, oceanography, imaging, and biomedical engineering. PSD reduces the redundancy in signals by concentrating its energy (information) into smaller areas of the frequency domain. However, the PSD still represents a significant amount of data, which is inconvenient to be directly presented to a pattern classifier. Therefore a further data reduction is needed. The goal is to come up with fewer quantities (features) that best characterize the PSD in some efficient manner. A straightforward approach is to use magnitude averages of PSD over a few frequency intervals. Another approach is to use the moments of PSD function.

This work has been motivated by the practical needs which the authors have met in the area of classification of prehensile electromyographic signals (EMG). The magnitude averages of PSD obtained from EMG patterns are used in [1], [2] and [3], which we consider as notable contributions to biomedical applications. The spectral moments are used by the authors of this

work [4] who have shown empirically the advantage of moments over the magnitude averages in prehensile EMG pattern classification.

This work focuses on a new method for efficient computation of spectral moments, and on the analysis of their performance in presence of noise. The concrete real-life temporal signals used in examples in subsequent sections deal with the classification of different types of grasps as defined by Schlesinger [5]: spherical, cylindrical, lateral and precision grasp. The EMGs were recorded from real subjects who have repeatedly performed grasping of spherical and cylindrical objects, keys and pens. The purpose of the experiments was to build multifunctional hand prosthesis. The experiments were described in detail in [6] and [7]. Although the examples used here are based on the prehensile EMG patterns, we believe that the ideas discussed here can be equally successfully applied to the classification of other types of temporal signals.

2 Computation of Spectral Moments

The PSD for a time sequence $s[i], i = 0,1,2,..., N-1$ can be estimated via periodogram

$$P(f) = \frac{1}{N}|S(f)|^2 = \frac{1}{N} S(f) S(f)^*, \qquad (1)$$

where $S(f)$ and $S(f)^*$ are discrete-time Fourier transform of $s[i]$ and its conjugate

$$S(f) = \sum_{i=0}^{N-1} s[i] e^{-j2\pi fi}. \qquad (2)$$

Note that $P(f)$ is a discrete-time, continuous-frequency version of PSD, with discrete time $t = i\Delta t$ and continuous frequency $f \in [0, \frac{1}{2}]$. The raw spectral moments are defined as

$$M_m = \int_0^{1/2} P(f) f^m df, m = 1,2,..., L. \qquad (3)$$

In order to compute Mm, the frequency can be discretized, $f = j\Delta f = j/N$, which would give

$$M_m = \frac{1}{N^{m+1}} \sum_{j=0}^{\lfloor N/2 \rfloor} P_j j^m, \qquad (4)$$

where $P_j = P(j/N)$ can be obtained via discrete Fourier transform of $s[i]$ followed by subsequent absolute operation.

In our approach we compute the moments by using a different representation of the PSD, based on Wiener-Khintchine theorem

$$P(f) = \sum_{k=-N+1}^{N-1} C_{ss}[k] e^{-j2\pi fk}, \quad (5)$$

where $C_{ss}[k]$ is the autocorrelation function of the sequence $s[i]$

$$C_{ss}[k] = \frac{1}{N} \sum_{i=0}^{N-|k|-1} s[i]s[i+k], k = \pm 1, \pm 2, ..., \pm(N-1). \quad (6)$$

Since $C_{ss}[k]$ is symmetric, i.e. $C_{ss}[-k] = C_{ss}[k]$, (5) can also be written as

$$P(f) = C_{ss}[0] + 2\sum_{k=1}^{K} C_{ss}[k]\cos(2\pi fk), \quad (7)$$

where $K \le N-1$. Now, by substituting (7) into (3) we finally get

$$M_m = C_{ss}[0] I_m(0) + 2\sum_{k=1}^{K} C_{ss}[k] I_m(k). \quad (8)$$

The equation (8) represents the new approach in computation of spectral moments of any degree. The coefficients $I_m(k)$ are defined as

$$I_m(0) = \int_0^{1/2} f^m df, \; I_m(k) = \int_0^{1/2} \cos(2\pi fk) f^m df. \quad (9)$$

After integration, the first few I-coefficients yield

$$I_m(0) = \frac{1}{2^{m+1}(m+1)},$$

$$I_0(k) = 0, I_1(k) = \frac{(-1)^k - 1}{(2\pi k)^2}, \; I_2(k) = \frac{(-1)^k}{(2\pi k)^2}, \; k = 1, 2, ...$$

$$I_3(k) = 3\frac{(-1)^k((\pi k)^k - 2) + 2}{(2\pi k)^4}, \; I_4(k) = 2\frac{(-1)^k((\pi k)^k - 6)}{(2\pi k)^4}. \quad (10)$$

The rest of the I-coefficients can be obtained from the following recurrent relation (see appendix)

$$I_m(k) = \frac{m}{2^{m-1}(2\pi k)^2}((-1)^k - 2^{m-1}(m-1)I_{m-2}(k)). \quad (11)$$

It is important to notice that the coefficients $I_m(k)$ do not depend on input data, and therefore can be computed in advance and kept in memory as reusable constants.

We also emphasize that this approach does not use the discretization of frequency. Instead, it takes the advantage of the fact that the PSD as defined by (1) and (2) is continuous in frequency. The continuous integration over the frequency is encapsulated in coefficients $I_m(k)$. It

is expected that this may reduce the integration error due to frequency discretization, which is inherent in straightforward approach.

3 Efficiency

From (10) and (11) it is apparent that I-coefficients decrease with the square of time lags k. Similarly, the autocorrelation function also tends to generally decrease with the time lags. Therefore we can expect that the maximal number of time lags, K, which is necessary to compute the moments through equation (8), can be much smaller than the length of the time sequences $s[i]$. The figures below illustrate this. (Note the coefficients $I_m(k)$ in figure 1 are multiplied by k for a better graphical presentation.)

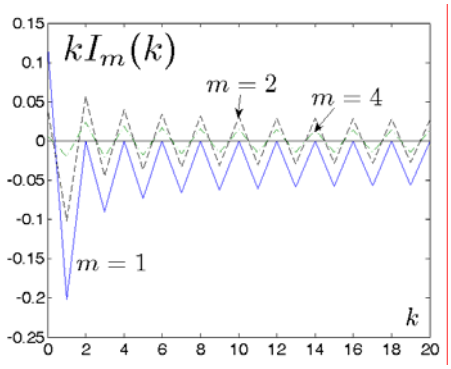


Fig. 1. Coefficients $I_m(k)$ for $m = 1,2,3$

The error diagram in figure 2 shows the relative differences between moments computed for N and K time lags. The diagram is obtained by averaging 30 time sequences of length 400 ms recorded from a surface EMG electrode while the subject was grasping a cylindrical object. Evidently, the error is smaller than 1% if $K > 10$. Similar error rates are obtained for other grasp types.

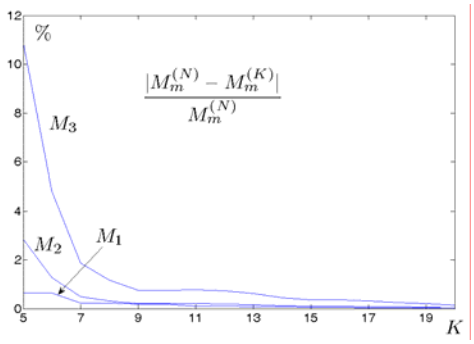


Fig. 2. Error in computing moments at $K < N$

Table 1. No. of floating-point operations for two approaches.

(Example: $m=5,\ N=400$ and $K=20.$)

Approach	$S(f)$	$C_{ss}[k]$	$\lvert S(f)\rvert^2$	M_m	Total	Example	
Standard		*	-	$3/2\,N$	$(m+1)N/2$	$(m+4)N/2$	1,800
Here	-	*	-	$2mK$	$2mK$	200	

(*) The sequence computed recurrently during signal acquisition.

The proposed approach for computation of spectral moments eliminates the need for FFT and for the subsequent absolute operation. At this point it can be argued that the FFT has to be used anyway for fast computation of autocorrelation functions [8]. However, this is not the case here because $C_{ss}[k]$ can be computed recurrently during the signal acquisition. Consequently, by the time the last sample $s[N-1]$ is acquired the sequence $C_{ss}[k]$ is already available for use in (8).

The table 1 summarizes the number of floating-point operations needed to compute m raw moments by using (4) and (8) respectively.

4 Impact of Noise

In real life the time sequences are contaminated with noise. This is especially true in case of surface EMG electrodes. The most common method to model the instrumental noise is the additive white Gaussian noise (AWGN). Based on (6) the autocorrelation function of the noisy sequence becomes

$$\tilde{C}_{ss}[k] = C_{s+n,s+n}[k] = C_{ss}[k] + C_{nn}[k] + C_{sn}[k] + C_{ns}[k]\,. \qquad (12)$$

AWGN is uncorrelated with itself and with other signals, therefore $C_{sn}[k] = C_{ns}[k] = 0$ for all $k,$ while $C_{nn}[k] = N_0^2\delta(k)$, where N_0^2 is the noise energy, while $\delta(0)=1, \delta(k>0)=0.$ The noise-to-signal ratio can be defined as $\lambda_n = C_{nn}[0]/C_{ss}[0]$, hence

$$C_{nn}[0] = \lambda_n C_{ss}[0]. \qquad (13)$$

In the case of realistic noise and shorter time sequences, the correlation sequences $C_{sn}[k]$ and $C_{ns}[k]$ for all k, and $C_{nn}[k]$ for $k>0$ can not be ignored. Substitution of (12) and (13) into (8) gives the noise-contaminated raw moments

$$\tilde{M}_m = M_m + 2\lambda_n M_0 I_m(0) + \varepsilon_m\,, \qquad (14)$$

where:

$$\varepsilon_m = (C_{sn}[0] + C_{ns}[0]) I_m(0) + 2\sum_{k=1}^{K} (C_{nn}[k] + C_{sn}[k] + C_{ns}[k]) I_m(k) \qquad (15)$$

Note that the index k in the sum above starts from 1.The quantity ε_m vanishes if the noise is AWGN, or $|\varepsilon_m| \approx 2\lambda_n M_0 I_m(0)$ if the noise is close enough to AWGN. The relative variation of moments due to noise can be finally expressed as

$$\frac{\Delta M_m}{M_m} = \frac{|\tilde{M} - M_m|}{M_m} \le 2\lambda_n I_m(0) \frac{M_0}{M_m} + \frac{|\varepsilon_m|}{M_m} \qquad (16)$$

It is apparent that the variations are very big even for ideal AWGN.

5 Reduced Moments

The high impact of noise on the accuracy of spectral moments suggests a modification of features based on the following linear transformation

$$R_m = \frac{M_m - 2I_m(0)M_0}{2} = \sum_{k=1}^{K} C_{ss}[k] I_m(k)\,, \qquad (17)$$

We will call the new moments R_m "reduced moments". It can easily be shown that the relative variation of reduced moments due to noise is

$$\frac{\Delta R_m}{R_m} = \frac{\varepsilon'_m}{R_m}\,, \qquad (18)$$

where

$$\varepsilon'_m = \frac{\varepsilon_m - (C_{sn}[0] + C_{ns}[0]) I_m(0)}{2} \qquad (19)$$

Since $\varepsilon'_m \approx R_m$ it appears that the reduced moments are much less impacted by noise. This is best illustrated in figure 5, which shows that the relative variations of reduced moments R_1, R_2 and R_3 are below 5% even at the noise-to-signal ratio $\lambda_n = 0.5$, while the variations of straight moments can be hundred times higher.

An example of ability of spectral moments (straight and reduced) to cluster and classify is shown in figure 4. Two sets of five-moment features, one for spherical grasps (dots) and the other for cylindrical grasps (circles) are considered. The five-moment feature vectors are projected onto 2D plane for the purpose of visual presentation. The projection uses two not necessarily orthogonal vectors. The first vector is a normal to the hyper-plane obtained by the support vector machine (SVM), which provides a separation of clusters with a maximal margin. The SVM are a family of learning algorithms [10] based on the work of Vapnik [9], which have recently gained a considerable interest in pattern recognition community. The success of SVM comes from their good generalization ability, robustness in high dimensional feature spaces and

good computational efficiency. The other vector is along the direction of minimal pooled variance of both cluster

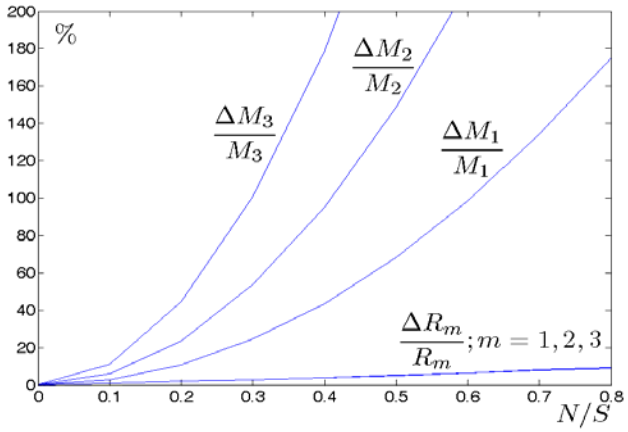


Fig.3. Relative variations of straight and reduced moments due to noise

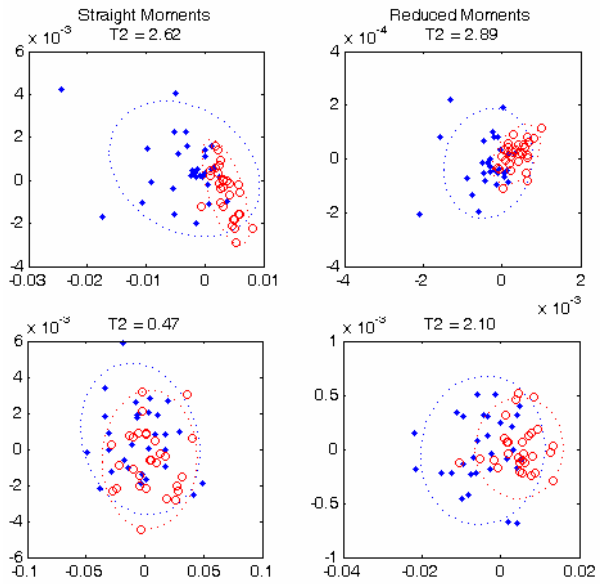


Fig.4. Comparison of straight $M_1,\ldots,M_5$ and reduced moments $R_1,\ldots,R_5$ without noise (above), and with AWGN at $N/S = 1$ (below).

The figure compares the straight and the reduced moments, without and with AWGN at $\lambda_n = 1$. For each case the Hotelling T^2 statistic [11] is computed, in order to compare the distance between clusters (the higher is the T^2 value, the better separation of clusters). The figure shows that the T^2 value has dropped due to noise more than five times in case of straight moments, while in case of reduced moments T^2 has dropped only by 27%. The figure also shows that the reduced moments performed slightly better than the straight moments in absence of noise. This could be explained by the fact that the time sequences recorded from real subjects might already be contaminated with noise.

In some applications the zero moment M_0 may be a very important feature. In such cases the zero moment might be retained in the set of reduced moments, i.e. $R_0 = M_0$. This can be afforded since the zero moment is impacted by noise only linearly with the N/S ratio, while the higher moments are more critical.

If we represent the straight and reduced moments by $(m+1)\times 1$ vectors $M = (M_0, M_1, ..., M_m)^T$ and $R = (R_0, R_1, ..., R_m)^T$ respectively, the moment transformation (17) can be expressed in matrix form:

$$R = TM \tag{20}$$

where

$$T = \begin{bmatrix} 1 & 0 & 0 & 0 & ... & 0 \\ -I_1(0) & 1/2 & 0 & 0 & ... & 0 \\ -I_2(0) & 0 & 1/2 & 0 & ... & 0 \\ -I_3(0) & 0 & 0 & 1/2 & ... & 0 \\ ... & ... & ... & ... & ... & 0 \\ -I_m(0) & 0 & 0 & 0 & ... & 1/2 \end{bmatrix} \tag{21}$$

The matrix T is invertible since $\det(T) = 1/2^m$, therefore the transformation (21) has an inverse, i.e. $M = T^{-1}R$.

6 Conclusion

A new algorithm for computation of spectral moments of any degree is proposed in this work. The algorithm utilizes autocorrelation sequence of the original temporal signal, rather than the signal's power spectral density, and offers much faster computation of moments than the standard method based on direct numerical integration of the power spectral density. The algorithm does not discretize the frequency and is therefore more accurate than the traditional approach. The algorithm enables to compute the raw spectral moments of any order by simply multiplying and summing the autocorrelation function of the input data signal with a series of coefficients that do not depend on data, and can therefore be pre-computed and kept in memory for use with any signal. A recurrent equation has been derived to generate these coefficients for a moment of any order.

The impact of noise has also been analyzed. It has been shown that the higher-order moments are very vulnerable to noise. Based on this analysis, the new set of modified moments is suggested. It is shown that the new moments are less impacted by the additive white Gaussian noise.

Acknowledgments

The authors would like to acknowledge the help of Ms. Hongyu Xu from C.A.C.I. in usage of support vector machine in experiments, and a numerous valuable discussions regarding the algorithms used in this work.

References

[1] Hannaford, B., Lehman, S.: Short Time Fourier Analysis of the Electromyogram: Fast Movements and Constant Contraction. IEEE Transactions on Biomedical Engineering, Vol. BME-33 (1986), pp. 1173-1181.
[2] Farry, K.A., Walker, I.D., Baraniuk,R.G.: Myoelectric Teleoperation of a Complex Robotic Hand. IEEE Trans On Robotic and Automation, Vol. 12 (5) (1996), pp. 777-787.
[3] Englehart, K., Hudgins, B., Parker, P, Stevenson, M.: Time-frequency representation for classification of the transient myoelectric signal. Proc. of the 20th Annual Int. Conference on Eng. in Med. and Biol. Vol. 5 (1998), pp. 2627-2630.
[4] Du, S. and Vuskovic, M.: Temporal vs. Spectral approach to Feature Extraction from Prehensile EMG Signals. The IEEE Int. Conf. on Information Reuse and Integration (IEEE IRI-2004), Las Vegas, Nevada (2004).
[5] Schlesinger, G.: Der Mechanische Aufbau der Kunstlishen Glieder. In Borchardt, M. etal (Edt.), Ersatzglieder und Arbeitshilfen für Kriegsbeschadigte und Unfallverzletzte. Berlin, Springer (1919).
[6] Vuskovic, M. I., Pozos, A. L., Pozos, R.: Classi.cation of Grasp Modes Based on Electromyographic Patterns of Preshaping Motions. Proc. of the IEEE Int. Conference on Systems, Man and Cybedrnetics, Vancouver, B.C., Canada, October 22-25 (1995), pp. 89-95.
[7] Vuskovic, M., Du, S: Classification of EMG Patterns with Simplified Fyzzy ART Networks. Proc. of the Int. Joint Conf. on Neural Networks. Honolulu, HI, May 12-17 (2002).
[8] Blahut, R.E.: Fast Algorithms for Digital Signal Processing. Addison-Wesley Pub. Co., Reading (1985), pp. 65-76.
[9] Vapnik V. N.: Statistical Learning Theory. John Wiley & Sons (1998).
[10] Cristianini, N., Shawe-Taylor, J.: Introduction to Support Vector Machines and other Kernel-Based Learning Methods. Cambridge Univ. Press, Cambridge (2000).
[11] Seber G.A.F.: Multivariate Observations, John Wiley & Sons (1984), pp. 102-117.

Appendix

Equation (11) can be derived if we apply partial integration to $x^m \cos(x)$ and $x^m \sin(x)$,

$$\int_0^{\pi k} x^m \cos(x)dx = [x^m \sin(x)]_0^{\pi k} - m\int_0^{\pi k} x^{m-1} \sin(x)dx , \qquad (22)$$

$$\int_0^{\pi k} x^m \sin(x)dx = -[x^m \cos(x)]_0^{\pi k} + m\int_0^{\pi k} x^{m-1} \cos(x)dx . \qquad (23)$$

where $x = 2\pi fk$. If we substitute definite integrals with C_m and S_m respectively, the equations above become

$$C_m = -mS_{m-1},$$

$$S_m = -(\pi k)^m \cos(\pi k) + mC_{m-1}. \quad (24)$$

After applying induction to the second equation, we get

$$S_{m-1} = -(\pi k)^{m-1} \cos(\pi k) + (m-1)C_{m-2}, \quad (25)$$

which substituted into (24) gives

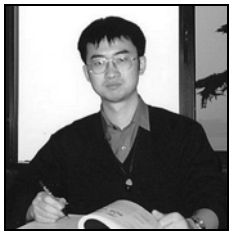
$$C_m = -m[-(\pi k)^{m-1} \cos(\pi k) + (m-1)C_{m-2}]. \quad (26)$$

The equation (11) follows directly from (26) since $\cos(\pi k) = (-1)^k$ and

$$C_m = \int_0^{\pi k} x^m \cos(x)dx = (2\pi k)^{m+1} \int_0^{\pi k} f^m \cos(2\pi kf)df = (2\pi k)^{m+1} I_m(k), \quad (27)$$

$$C_{m-2} = (2\pi k)^{m-1} I_{m-2}(k). \quad (28)$$

Marko Vuskovic received his B.S. and M.S. degrees in electrical and computer engineering from University of Belgrade, Yugoslavia in 1964 and 1969 respectively, and the Dr. Sci. degree in technical sciences from University of Zagreb, Croatia in 1975. From 1964 to 1986 he was a research associate, then senior research associate with the Automation Department of Institute Mihailo Pupin, Belgrade. In 1986 he joined the Computer Science Department of San Diego State University as a professor of computer science. In 1987 he has founded Robotics and Neural Network Laboratory at SDSU. His research interests include robot kinematics and dynamics, prehensile robotics, pattern recognition, neural networks, feature extraction, ART-based neural networks. His current focus is on feature extraction and classification of glycan microarrays applied to diagnosis and prognosis of cancer.



Sijiang Du received his B.S. degree in computer science from Nanjing University of Science and Technology, Nanjing, China in 1996, and M.S. degree in computer science from San Diego State University, San Diego, California in 2003. He was research member and TA in Robotics and Neural Networks Laboratory at SDSU from 2001 to 2003. His research at SDSU includes pattern recognition and feature extraction of prehensile EMG signals. He is currently a software design engineer at Nokia Mobile Phones, Inc. where he works on embedded UI software for CDMA handset.