Optimal autoregressive orders for myopathic electromyograms

José G. Vicente–*EMBS Member*, Cinthia Itiki–*EMBS Member*

***Abstract*— This paper aims to describe the optimal autoregressive order of varying-length electromyograms for myopathic subjects. Epochs of electromyography signals are modeled as outputs of autoregressive systems, for orders varying from 1 to 100. The optimal order to represent each epoch is chosen by the minimum description length criterion. Probability density functions are fitted to the histograms of the optimal orders. The lognormal function provides the best fitting, and its mean value varies linearly with the epoch length.**

I. Introduction

Electromyography (EMG) consists of recording and evaluating the electrical activity from skeletal muscle. It is an important tool for the diagnosis of neuromuscular diseases [1]-[2]

A simple mathematical description of EMG signals would be as outputs of linear time-invariant systems. The use of such systems is justified, since some parameters of linear systems—autoregressive (AR) and cepstral coefficients—have provided useful information to help clinical diagnosis [3]-[6].

However, the use of AR systems poses the issue of choosing the appropriate number of parameters to represent the EMG signals. In applications involving movement recognition, prosthesis control and muscle fatigue, AR order is well established in the low range from two to six [7]-[12]. On the other hand, in the diagnostic classification, higher AR orders—from twelve to twenty—are used. Different orders provide different rates of diagnostic classification, sweeping the range from 47.6% to 87.5% [3]-[4]. These results show that the number of coefficients of the AR model is a parameter that might influence the classification results.

A previous paper has shown how AR orders varied with epoch length, in normal subjects—the mean values of the estimated orders increased with epoch length [13]. Several criteria have been used to estimate optimal AR orders, and the Minimum Description Length (MDL) criterion was shown to be the most appropriate, since it provided intermediate estimates of AR order [13], in comparison to Akaike's Information Criterion [14] and Bayesian Information Criterion [15].

This paper focuses in the pathological cases, specifically in the myopathic subjects. The optimal AR orders are estimated by the MDL criterion, for several epoch lengths. Then probability density functions are fitted to the histograms of the estimated orders.

The results presented here may be helpful in two applications. The description of AR orders by probability density functions may help the development of an EMG model for simulation purposes, including normal and pathological cases. It can also be used as an aid to the detection of myopathies. For both applications, further research on the model coefficients should be performed for AR systems whose orders are described by the probability density functions given in this paper.

II. Methods

A. Electromyography Signals

EMG signals were obtained from a database, from which six myopathic subjects were chosen. Signals had been recorded by concentric needle electrodes at several locations of the *biceps brachii* muscle. Signal conditioning included proper amplification up to 1,000 times, and low-pass filtering with a cutoff-frequency of 10 kHz.

Seven signals were selected for each subject, and corresponded to isometric and isotonic contractions at fifty percent of the maximum voluntary contraction force. The selected signals were acquired at the rate of 25,000 samples per second by a twelve-bit analog-to-digital converter.

Epochs of 500-ms length were tested for stationarity in mean and variance by the run test [16]. The selected stationary epochs were further divided into smaller epochs, in the range of 50 ms to 500 ms, in order to incorporate usual lengths mentioned in the literature [3], [13].

B. Autoregressive modeling

Each epoch was considered as the output of an autoregressive (AR) system described by (Kay, 1987)

$$y(t) = -\sum_{k=1}^{n} a_n(k)\, y(t-k) + e(t)\,, \qquad (1)$$

where $y(t)$ is the EMG signal at the system's output, $e(t)$ is the unknown white noise at the system's input, t is the sample number, n is the model's order, and $a_n(k)$ are the model coefficients for a n^{th}-order model.

The AR coefficients and the variance of the input (zero-mean white noise) were estimated by the modified covariance method. This method was chosen because it presents good spectral resolution and does not suffer from spectral interference [17]-[18]. The modified covariance method is a least-squares technique for estimating the

J. G. Vicente is with Laboratory of Mechanical Tests and Implants (LEME), Technological Research Institute (IPT), Sao Paulo, SP 05508-901, Brazil (e-mail: gvicente@ipt.br).

C. Itiki is with the Biomedical Engineering Laboratory (LEB), Department of Telecommunications and Control Engineering, Escola Politécnica, University of São Paulo, Sao Paulo, SP 05508-010, Brazil (phone: +55(11)3091-5150; fax: +55(11)3091-5718; e-mail: cinthia@leb.usp.br).

978-1-4244-7929-0/14/$26.00 ©2014 IEEE

autoregressive coefficients. It minimizes the sum of squares of the forward and backward linear-prediction errors [19]. The forward and backward linear-prediction errors are represented by

$$e_n^f(t) = y(t) + \sum_{k=1}^{n} a_n(k)\, y(t-k)\,, \qquad (2)$$

for $t \le N-n$, and

$$e_n^b(t) = y(t-n) + \sum_{k=1}^{n} a_n^*(k)\, y(t-n+k)\,, \qquad (3)$$

for $n-1 \le t \le N-n$, where the superscript asterisk * denotes complex conjugation, and N is the total number of samples.

The optimal order of the AR system that best fitted the EMG signal was determined by the MDL function [20]

$$MDL\ (n) = N\ \ln(\hat{\sigma}_n^2) + \ln(N)\, n\,, \qquad (4)$$

where n is the AR order, N is the epoch length in number of samples, and σ_n^2 is the variance estimate of the input noise.

C. Histograms and fitting of probability-density functions

For each epoch length, seven histograms were built—six individual histograms for the AR orders from each individual subject as well as one histogram for the ensemble of AR orders from all the subjects. Since the AR orders were integer numbers varying from 1 to 100, the histogram bins had unitary width and were computed in the range of 1 to 100. The obtained histograms were then normalized by the corresponding number of epochs, resulting in the normalized histograms $H(n)$.

Several probability density functions were fitted to the normalized histograms. Three of them—Gaussian, lognormal and Weibull—provided the most representative results and are therefore reported in this paper. The Gaussian probability density function is described by

$$f_G(x \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\,, \qquad (5)$$

for $x \in \mathbf{R}$, where μ is the mean and σ is the standard deviation ; while the lognormal probability density function is given by

$$f_{LN}(x \mid \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln x-\mu)^2}{2\sigma^2}}\,, \qquad (6)$$

for $x>0$. The Weibull probability density function is

$$f_W(x \mid a, b) = \frac{b}{a}\left(\frac{x}{a}\right)^{b-1} e^{-(x/a)^b}\,, \qquad (7)$$

for $x>0$, where a is the scale parameter and b is the shape parameter.

D. Error computation

Each probability density function $f(x)$ was integrated in the intervals $(n-0.5,\ n+0.5]$, resulting in the discrete distribution

$$P(X = n) = \int_{n-0.5}^{n+0.5} f(x)\, dx\,, \qquad (8)$$

for integer values of n.

The total quadratic errors (E_T) between the discrete distribution and the corresponding normalized histogram were given by

$$E_T = \lim_{L\to\infty} \sqrt{\sum_{-L}^{+L} \left(P(X=n) - H(n)\right)^2}\,, \qquad (9)$$

and the summation was truncated for L=1,000.

RESULTS

Fig. 1 shows the fitting of probability density functions to an individual histogram, as well as to the ensemble histogram. The optimal orders represented in these histograms were computed for 100-ms epochs. The fitted Gaussian function showed the widest peak, while the lognormal function provided the narrowest peak.

Fig. 2 and Fig. 3 show the estimated parameters for the Gaussian, Weibull and lognormal distributions. Fig. 2 shows that, in average, the parameters μ (for the Gaussian and lognormal distribution) and a (for Weibull distribution) increase with the epoch length. Friedman's test shows that these increases are statistically significant ($p < 1$ %).

Fig. 3 shows an increase on parameter σ with epoch length, on average, only for the Gaussian distribution ($p < 1$ %). The remaining parameters do not show statistically significant variation with epoch length.

Fig. 4 shows the total quadratic errors in fitting the functions to the normalized histograms computed for the optimal AR orders at varying epoch lengths (in milliseconds). Considering the ensemble histograms, the smaller errors are provided by the lognormal probability density function, and the largest errors are given by the Gaussian distribution. Intermediate results are obtained with Weibull distribution. If one considers the mean errors, similar results are obtained for the individual fittings.

The best fitting to the ensemble histogram can be described by a lognormal distribution with standard deviation σ equal to 0.6889 and a mean value given by the linear relation μ=0.0027D+1.5415, for epoch lengths D in milliseconds.

Table I shows the mean values and standard deviations of the optimal AR-order histograms, for several epochs of normal [13] and myopathic subjects. Apparently mean values increase with epoch length, for both normal and myopathic cases. A similar trend is observed for standard deviations. Besides it, for 100 ms, 250 ms and 500 ms, both mean values and standard deviations are higher (or equal) for myopathic orders than for normal ones.

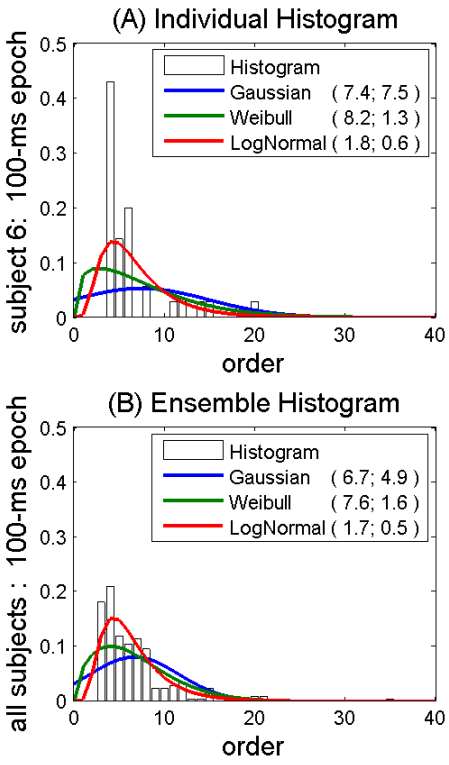

Figure 1. Fitting of probability density functions to the (A) individual histogram from subject 6, and to the (B) ensemble histogram.

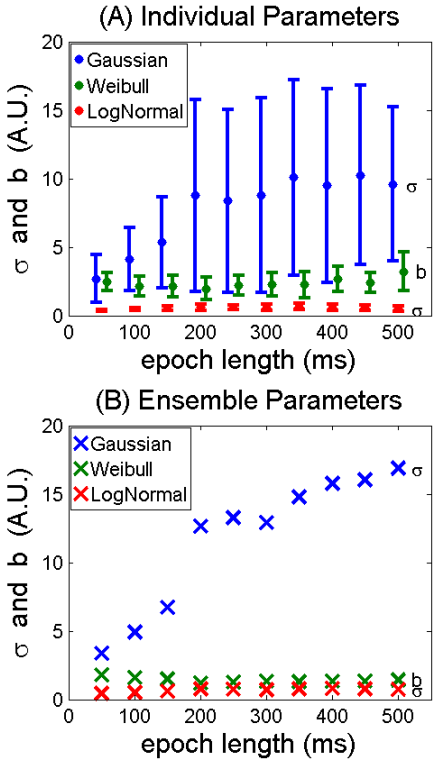

Figure 3. Parameters σ and b for probability density functions. (A) circles and bars represent mean value ± standard deviation for parameters from individual fittings, while (B) crosses represent the result for the ensemble.

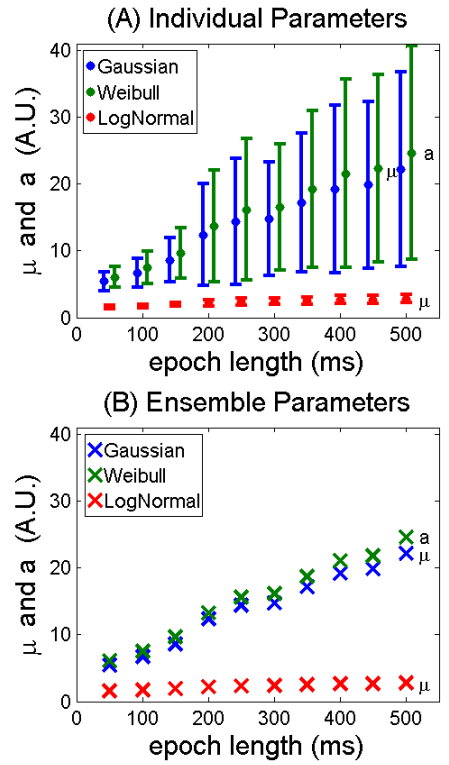

Figure 2. Parameters μ and a for probability density functions. (A) circles and bars represent mean value ± standard deviation for the parameters from individual fittings, while (B) crosses represent the result for the ensemble.

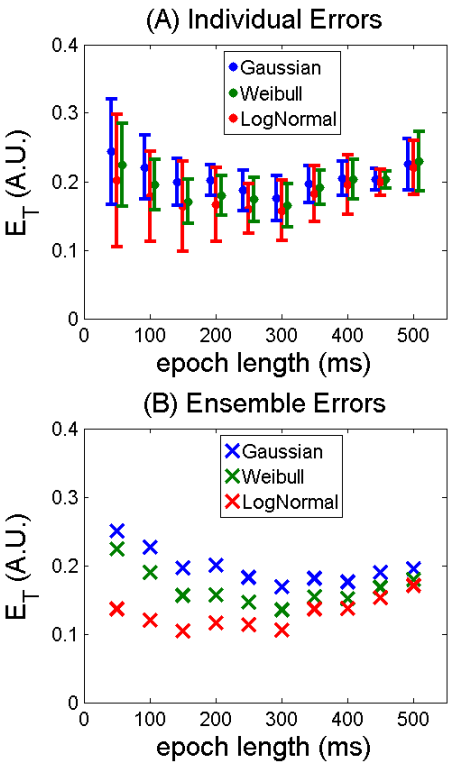

Figure 4. Total quadratic error for fitting probability density functions to the AR order histograms. (A) circles and bars represent mean value ± standard deviation of errors for individual fittings, while (B) crosses represent the errors for the ensemble of subjects.

TABLE I. COMPARISON OF NORMAL [13] AND MYOPATHIC AR-ORDER ENSEMBLE HISTOGRAMS (MEAN VALUE ± STANDARD DEVIATION)

epoch (ms)	Normal	Myopathic
25	5 ± 3	
50	6 ± 2	5 ± 3
100	7 ± 3	7 ± 5
150		9 ± 6
200		12 ± 13
250	10 ± 5	14 ± 13
300		15 ± 13
350		17 ± 15
400		19 ± 16
450		20 ± 16
500	12 ± 8	22 ± 17

III. DISCUSSION

The increase on the parameters μ (Gaussian and lognormal mean values) and a (Weibull scale parameter) with epoch length reflects a known behavior—the AR order estimates increase with epoch length. As a consequence, the mean value of the probability density function accompanies the increase of order estimates.

On the other hand, the increase of the parameter σ for the Gaussian distribution shows that the fitted functions spread over a wider range of values, as the epoch length increases.

Errors indicate that the lognormal distribution provides the best fitting to the histograms and that the Gaussian distribution is the worst curve adjustment to the histograms. One possible reason is given by the fact that, unlike the Gaussian distribution, the lognormal probability density function is asymmetric and able to represent peaks for orders concentrated in the low end of the histograms. Besides it, Gaussian distribution attributes non-zero probability to negative orders, which increases the error.

Comparing myopathic to normal subjects, higher AR orders are observed for myopathic subjects, on average. This result suggests that normal EMG signals are better described by smaller AR models, and provide more consistent order estimates, as shown by smaller variances.

IV. CONCLUSION

Results suggest that the optimal AR order varies with the epoch length and that, for myopathic subjects, it may be described by a lognormal probability density function of linearly varying mean and constant variance.

ACKNOWLEDGMENT

C. I. thanks Prof. D. Sanders from Duke University for the EMG database and G. Vicinansa for selecting the signals.

REFERENCES

[1] A. Fuglsang-Frederiksen, "Interference EMG analysis," in *Computer-Aided Electromyography and Expert Systems*, J. E. Desmedt, Ed. Amsterdam: Elsevier Science, 1989, pp. 161–179.

[2] C. Berzuini, M. Maranzana-Figini, and L. Bernardinelli, "Effective use of EMG parameters in the assessment of neuromuscular diseases," *International Journal of Bio-Medical Computing*, vol. 13, pp. 481–499, 1985.

[3] G. F. Inbar, A. E. Noujaim, "On surface EMG spectral characterization and its application to diagnostic classification," *IEEE Transactions on Biomedical Engineering*, vol. 31, no. 9, pp. 597–694, 1984.

[4] C. S. Pattichis and A. G. Elia, "Autoregressive and cepstral analyses of motor unit action potentials," *Medical Engineering and Physics*, vol. 21, no. 6–7, pp. 405–419, 1999.

[5] C. Itiki, "Dynamic programming and diagnostic classification," Journal of Optimization Theory and Applications, vol. 127, ed. 3, Dec. 2005, pp.579-586

[6] C. Shirota, M.Y. Barretto, C. Itiki. "Associative memories and diagnostic classification of EMG signals," In: *Adaptive and Natural Computing Algorithms*. B. Ribeiro, R. F. Albrecht, A. Dobnikar, Ed. Coimbra: Springer, 2005, pp.482-485.

[7] J. Chiang, Z. Wang, M.J. McKeown. "Hidden Markov multivariate autoregressive (HMM-mAR) modeling framework for surface electromyography (sEMG) data," *Conf. Proc. IEEE Eng. Med. Biol. Soc*., 2007, pp. 4826–4829, 2007.

[8] A. Latwesen, P.E. Patterson. "Identification of lower arm motions using the EMG signals of shoulder muscles," *Med. Eng. Phys*., vol. 16, no. 2, pp. 113–121, 1994.

[9] Y. H. Chiou, J. J. Luh, S. C. Chen, J. S. Lai, T. S. Kuo. "The comparison of electromyographic pattern classifications with active and passive electrodes," *Med. Eng. Phys.*, vol. 26, no. 7, pp. 605–610, 2004.

[10] X. Chen, X. Zhu, D. Zhang. "Use of the discriminant Fourier-derived cepstrum with feature-level post-processing for surface electromyographic signal classification," *Physiol. Meas*., vol. 30, no. 12, pp. 1399–1413, 2009.

[11] E. A. Clancy, D. Farina, R. Merletti. "Cross-comparison of time- and frequency-domain methods for monitoring the myoelectric signal during a cyclic, force-varying, fatiguing hand-grip task," *J. Electromyogr. Kinesiol.*, vol. 15, no. 3, pp. 256–265, 2005.

[12] S. Karlsson, J. Yu. "Estimation of surface electromyogram spectral alteration using reduced-order autoregressive model," *Med. Biol. Eng. Comput.*, vol. 38, no. 5, pp. 520–527, 2000.

[13] C. Itiki. Epoch length and autoregressive-order selection for electromyography signals. *Proceedings of the 34th Annual International Conference of the IEEE-EMBS*, San Diego, USA, 28 Aug.-1 Sept., pp. 3476–3479, 2012

[14] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. AC-19, no. 6, pp. 716–723, 1974.

[15] H. Akaike, "A Bayesian analysis of the minimum AIC procedure," *Annals of the Institute of Statistical Mathematics*, vol. 30, part A, pp. 9–14, 1978.

[16] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Annals of Mathematical Statistics*, vol. 19, no.3, pp. 326–339, 1948.

[17] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Englewood Cliffs: Prentice- Hall, 1988, pp. 234–237.

[18] S. L. Marple Jr. *Digital Spectral Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[19] S. L. Marple Jr. "A fast computational algorithm for the modified covariance method of linear prediction," *Digital Signal Processing*, vol. 1, pp. 124–133, 1991.

[20] J. Rissanen, "A universal prior for the integers and estimation by minimum description length," *The Annals of Statistics*, vol. 11, pp. 417–431, 1983.