34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Epoch Length and Autoregressive-Order Selection for Electromyography Signals

Cinthia Itiki*

Abstract—**This study shows how different EMG-epoch lengths affect the selection of the autoregressive-model orders. Electromyography signals were divided in 25ms, 50ms, 100ms, 250ms and 500ms epochs. Order-selection criteria were applied to the least-square errors of autoregressive models. The Bayesian Information Criterion and the Minimum Description Length indicated that needle-EMG signals recorded from normal subjects at 25kHz could be represented by autoregressive models using orders below 25 for 500ms epochs, and that smaller orders could be used to represent shorter epochs.**

I. Introduction

Electromyography (EMG) signals have been used in several studies, including hand motion classification, electrode shift during movements, muscle fatigue, and diagnosis of neuromuscular diseases. [1–7]

As an aid to clinical diagnosis, several EMG parameters have provided useful information—temporal parameters (amplitude, duration, number of phases and zero crossings), spectral parameters (median and mean frequencies), and linear-model parameters (autoregressive and cepstral coefficients) among others. [4–7]

On the autoregressive (AR) modeling of EMG signals for diagnosis purposes, the epoch duration has been chosen in a wide range, from 25.6ms to 500.0ms The AR-model order has also been selected in the range of 12 to 20. These ranges, associated with several other methodological differences, resulted in global classification rates varying from 47.5% to 87.5%. [6–7]

These results indicate that the choice of the AR-model order is an important issue, and that it may be influenced by the EMG epoch duration. No previous work has studied the variation of the AR-model order with the length of EMG epochs recorded under the same experimental settings. So, this work aims to study how different EMG-epoch lengths may affect the selection of the autoregressive-model orders.

II. Methods

An EMG database was used. Twenty-one signals were selected from three subjects that presented no neuromuscular disease. They lasted 500ms and had been classified as stationary by the Wald-Wolfowitz test. [8] The database signals had been recorded from the *biceps brachii* at 50% of maximum voluntary contraction. They had been amplified 100 to 1,000 times, had been filtered by a low-pass with a cut-off frequency of 10kHz, and had been acquired by a 12-bit A/D converter at the rate of 25 thousand samples per second.

In this work, each 500ms signal was divided into two 250 ms epochs, five 100ms epochs, ten 50ms epochs or twenty 25ms epochs. The range of 25ms to 500ms was chosen to encompass the usual epoch durations found in the literature. [6–7]

Each epoch was modeled as the output of an autoregressive system, described by [9]

$$y(t) + a_1 y(t-1) + \ldots + a_n y(t-n) = e(t), \qquad (1)$$

where $y(t)$ was the output signal, a_k were the model coefficients, $e(t)$ was the unknown input white noise, t=1, …, N was time, N was the total number of samples, and n was the model order. The model coefficients and the input-signal variance were computed by Matlab's *arx.m* function, which minimized the least square error, through QR factorization (see appendix).

For a given epoch, the input variance—that varied with the number (n+1) of coefficients—was used in three order selection criteria [10–12]. These criteria selected the optimal order as the one that minimized a function pondering the epoch length N and the model order n, according to

$$f(N,n,\hat{\sigma}_y^2,\hat{\sigma}_n^2) = N \times g(N,n,\hat{\sigma}_n^2) + n \times h(N,n,\hat{\sigma}_y^2,\hat{\sigma}_n^2), \qquad (2)$$

where N was the epoch length, n was the autoregressive-model order, $\hat{\sigma}_n^2$ was the variance of the estimated input signal, $\hat{\sigma}_y^2$ was the output-signal variance estimate computed in the epoch, and the pondering functions were $g(N,n,\hat{\sigma}_n^2)$ and $h(N,n,\hat{\sigma}_y^2,\hat{\sigma}_n^2)$, given by Table I.

The epoch lengths were 12,500 samples, 6,250 samples, 2,500 samples, 1,250 samples or 625 samples for epoch durations of 500ms, 250ms, 100ms, 50ms or 25ms respectively. The model order swept the 1 to 99 range.

TABLE I. Pondering functions for Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and Minimum Description Length (MDL)

	$g(N,n,\hat{\sigma}_n^2)$	$h(N,n,\hat{\sigma}_y^2,\hat{\sigma}_n^2)$
AIC	$\ln(\hat{\sigma}_n^2)$	2
BIC	$\ln(\hat{\sigma}_n^2\,(1-n/N))$	$\ln((1-\hat{\sigma}_y^2/\hat{\sigma}_n^2)\,(1-N/n))$
MDL	$\ln(\hat{\sigma}_n^2)$	$\ln(N)$

* C. Itiki is with the Biomedical Engineering Laboratory, University of Sao Paulo, Sao Paulo, SP 05508-970, Brazil (phone: +55(11)3091-5150; fax: +55(11)3091-5718; e-mail: cinthia@leb.usp.br).

III. RESULTS

Fig.1 shows the mean orders selected by the three criteria, for the five epoch durations. The more compact models are indicated by BIC, closely followed by MDL. On the other hand, AIC provided the highest order estimates, indicating an overestimation of the autoregressive-model order. On average, the selected model order augments with the increase of the epoch duration.

Fig. 2 shows the histograms of the optimal order selected by the Akaike's Information Criterion (AIC) for all the epoch durations. The shape of the histograms varied with the epoch length, as the selected orders were spread over the whole range (1 to 99) for long epochs (250ms to 500ms) and were concentrated in the low orders for short epochs (25ms to 100 ms).

For the Bayesian Information Criterion (BIC), Fig. 3 shows the histograms of selected orders. Unlike the AIC histograms, the BIC histograms show a concentration in the low orders for all epoch lengths (25ms to 500ms).

Fig. 4 shows similar results for the optimal orders selected by the Minimum Description Length (MDL). There is a concentration in the orders below 12 and 25, for short and long epochs respectively.

Table II shows the mean values and standard deviations of the orders that were selected by the three criteria. The standard deviations presented on table II confirmed that the AIC order estimates were more widely spread over the 1 to 99 range, in comparison to BIC and MDL. The mean values also showed that AIC provided the highest order estimates. On the other hand, BIC provided the smallest order estimates, which were also more concentrated around the mean values.

The BIC and MDL results obtained in this work corroborate to the order 20 used by [6] for 500ms epochs, and the order 12 used by [7] for 25,6ms epochs.

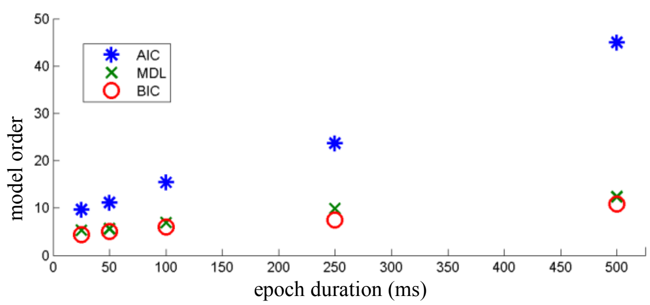

Figure 1. Mean orders selected by AIC (*), BIC (°) and MDL (×) for epoch durations of 25ms, 50ms, 100ms, 250ms and 500ms.

TABLE II. SELECTED ORDER (MEAN VALUE ± STANDARD DEVIATION) OF AUTO-REGRESSIVE MODELS BY AIC, BIC AND MDL.

Epoch duration	AIC	BIC	MDL
25ms	10±9	4±1	5±3
50ms	11±9	5±2	6±2
100ms	15±11	6±3	7±3
250ms	24±18	7±3	10±5
500ms	45±29	11±5	12±8

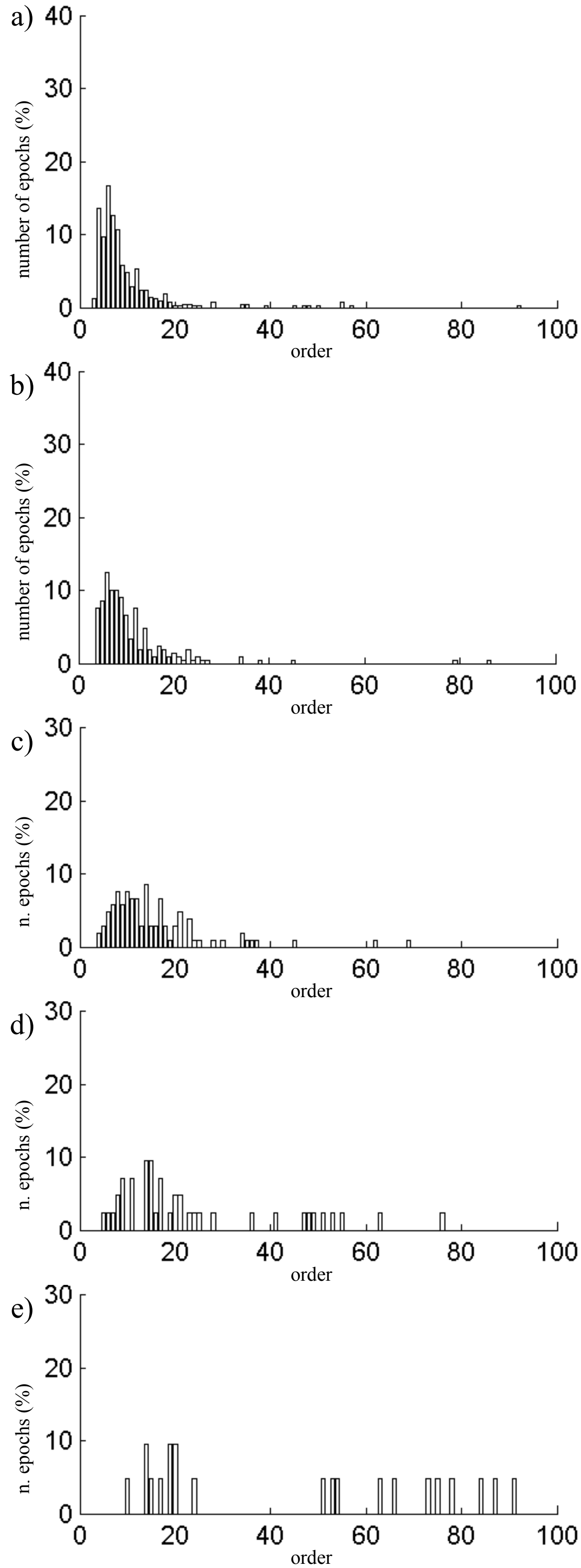

Figure 2. Histogram of the optimal orders selected by AIC for epoch durations of a) 25ms, b) 50ms, c) 100ms, d) 250ms and e) 500ms.

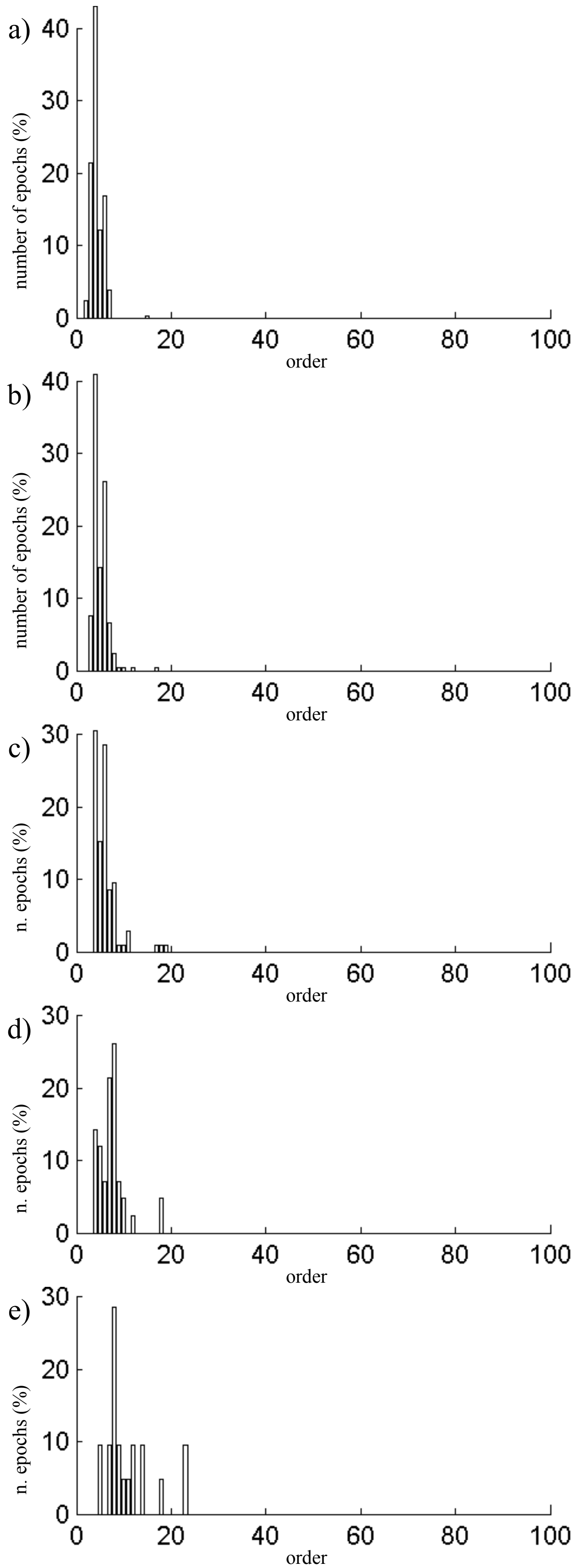

Figure 3. Histogram of the optimal orders selected by BIC for epoch durations of a) 25ms, b) 50ms, c) 100ms, d) 250ms and e) 500ms.

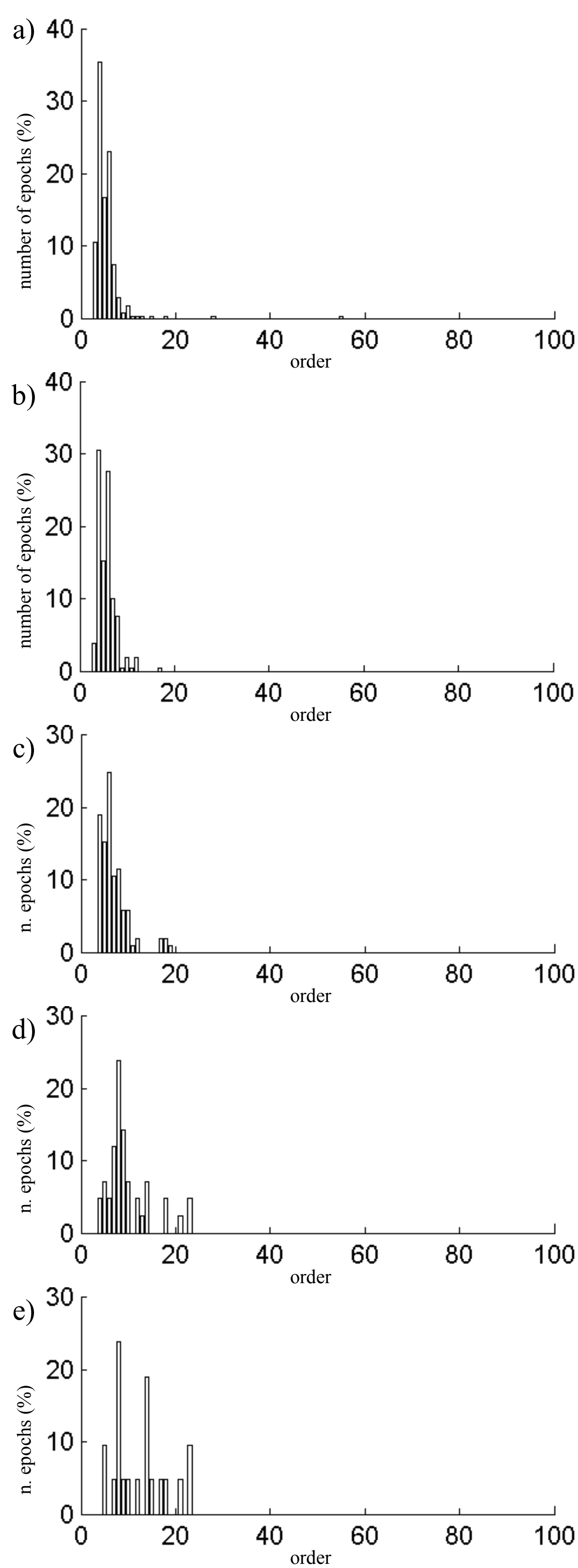

Figure 4. Histogram of the optimal orders selected by MDL for epoch durations of a) 25ms, b) 50ms, c) 100ms, d) 250ms and e) 500ms.

IV. Conclusion

The Akaike Information Criterion was used as a comparison standard, and showed its well-known trend to overestimate the orders of autoregressive models. The Bayesian Information Criterion and the Minimum Description Length indicated orders below twenty five for 500ms epochs, and below twelve for 25ms epochs. These quantitative values should be interpreted in the experimental context and should not be extrapolated to different sampling frequencies, to other contraction forces, to surface electromyography or to pathological subjects. Future works should address these diverse experimental conditions.

However, some qualitative results can be extended to different experimental settings. Indeed, this work shows that short EMG epochs can be described by low-order AR models, which could apply to isolated motor unit action potentials—whose duration is compatible with 25ms epochs. On the other hand, interference patterns are usually analyzed in longer epochs, and as a consequence, they would be better described by higher-order AR models. In this case, the compromising choice of medium-length epochs—between 100ms and 250ms—would reduce the optimal AR-model orders and consequently the computational load as well.

The choice of AR orders should not be based on previous results for different experimental conditions, but it should be confirmed for the specific EMG recording settings. For this purpose, MDL provided intermediate values of order estimates, when compared to AIC and BIC. Furthermore, its pondering functions are simpler than the ones for BIC, guaranteeing a smaller consumption of computing time. So, the results from this work suggest that MDL should be used to estimate the AR-model order for EMG signals.

Appendix

The set of n equations (1), for $t = 1, 2, \ldots, N$, could be represented in matrix notation by

$$\mathbf{y}_N - \mathbf{\Phi}\, \boldsymbol{\theta}_n = \mathbf{e}_N, \tag{3}$$

where $\mathbf{y}_N^T = [\, y(1) \ldots y(N)]$, $\mathbf{e}_N^T = [\, e(1) \ldots e(N)]$, $\boldsymbol{\theta}_n^T = [\, a_1 \ldots a_n \,]$, $\mathbf{\Phi}^T = [\, \varphi(1) \ldots \varphi(N) \,]$, and $\varphi^T(t) = [\, -y(t-1) \ldots -y(t-n) \,]$.

In the least squares method, the parameter estimate $\hat{\boldsymbol{\theta}}_n$ is the minimizing argument of [13]

$$V_N(\boldsymbol{\theta}_n, \mathbf{y}_N) = (\, \mathbf{y}_N - \mathbf{\Phi}\, \boldsymbol{\theta}_n \,)^T (\mathbf{y}_N - \mathbf{\Phi}\, \boldsymbol{\theta}_n \,). \tag{4}$$

An efficient solution for the minimization problem is the QR factorization [13]. An orthonormal $N\times N$ matrix $\mathbf{Q}$ and an upper triangular $N\times(n+1)$ matrix $\mathbf{R}$ are used to represent

$$[\mathbf{\Phi}\ \ \mathbf{y}_N \,] = \mathbf{Q}\, \mathbf{R} \tag{5}$$

where $\mathbf{Q}^T\mathbf{Q}=\mathbf{I}$ and $\mathbf{R} = \begin{bmatrix} \mathbf{R_1} & 0 & 0 \\ \mathbf{r_2} & r_3 & 0 \end{bmatrix}^T$, with a square $n\times n$ matrix $\mathbf{R_1}$, a $n\times 1$ column vector $\mathbf{r_2}$ and a scalar r_3. The left-multiplication of (5) by $\mathbf{Q}^T$ results in

$$\mathbf{Q}^T [\mathbf{\Phi}\ \ \mathbf{y}_N \,] = \mathbf{Q}^T \mathbf{Q}\, \mathbf{R} = \mathbf{R}. \tag{6}$$

Using the orthonormality property of matrix $\mathbf{Q}$, the minimizing function can be rewritten as

$$V_N(\boldsymbol{\theta}_n, \mathbf{y}_N) = (\, \mathbf{y}_N - \mathbf{\Phi}\, \boldsymbol{\theta}_n \,)^T \mathbf{Q}\, \mathbf{Q}^T (\mathbf{y}_N - \mathbf{\Phi}\, \boldsymbol{\theta}_n \,). \tag{7}$$

As a result, the function to be minimized is [13]

$$V_N(\boldsymbol{\theta}_n, \mathbf{y}_N) = [\, \mathbf{r}_2 - \mathbf{R}_1\boldsymbol{\theta}_n \,]^T [\, \mathbf{r}_2 - \mathbf{R}_1\boldsymbol{\theta}_n \,] + r_3 r_3, \tag{8}$$

whose solution is given by a set of n equations [13]

$$\mathbf{R_1}\, \hat{\boldsymbol{\theta}}_n = \mathbf{r_2}\,, \tag{9}$$

and results in the loss function

$$V_N(\hat{\boldsymbol{\theta}}_n, \mathbf{y}_N) = r_3^T r_3\,. \tag{10}$$

Acknowledgment

C. I. thanks Mr. Vicente for the figure editing.

References

[1] H. Huang, H. B. Xie, J. Y. Guo, and H. J. Chen, "Ant colony optimization-based feature selection method for surface electromyography signals classification," *Computers in Biology and Medicine,* vol. 42, no. 1, pp. 30-38, Jan 2012.

[2] A.J. Young, L. J. Hargrove, and T. A. Kuiken, "Improving myoelectric pattern recognition robustness to electrode shift by changing interelectrode distance and electrode configuration," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 3, pp. 645-52, Mar 2012.

[3] Z. G. Zhang, H. T. Liu, S. C. Chan, K. D. Luk, and Y. Hu, "Time-dependent power spectral density estimation of surface electromyography during isometric muscle contraction: methods and comparisons," *Journal of Electromyography and Kinesiology*, vol. 20, no. 1, pp. 89-101, Feb 2010.

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

[5] 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.

[6] 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.

[7] C. S. Pattichis and A. G. Elia, "Autoregressive and cepstral analyses of motor unit action potentials," *Medical Engineering and Physics*, vol. 21, n.6-7, pp. 405–419, 1999.

[8] 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.

[9] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Englewood Cliffs: Prentice- Hall, 1987, pp. 234–237.

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

[11] 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.

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

[13] L. Ljung, *System Identification Theory for the User*, 2nd ed. Upper Saddle River, NJ: PTR Prentice Hall, 1999, ch.10.