

Bias Propagation in the Autocorrelation Method of Linear Prediction

Jan S. Erkelens and Piet M. T. Broersen

Abstract— A time-domain analysis of the autocorrelation method for autoregressive estimation is given. It is shown that a small bias in a reflection coefficient close to one in absolute value is propagated and prohibits an accurate estimation of further reflection coefficients. Tapered data windows largely reduce this effect, but increase the variance of the models.

I. INTRODUCTION

MANY low bit-rate speech coders use the autocorrelation method (ACM) to find a linear prediction model of the speech signal. It is known that a tapered data window improves the performance of the autocorrelation method in speech analysis. Sometimes, the frequency domain interpretation of the ACM is given as the motivation for the use of a tapered window, but it will be shown that the reduction of the edge effects is important in the time-domain. A time-domain analysis of the autocorrelation method will be presented that also gives theoretical support to the use of tapered data windows with the ACM.

For stochastic signals, the behavior of autoregressive estimation methods can be expressed in terms of bias and variance of the model parameters. Estimation methods have been treated extensively in the literature (e.g., [1]–[11]). Some studies express the bias in terms of the poles or the parameters of the autoregressive process [3], [4]. Although the bias is described accurately, the formulas do not give insight in what causes the bias and what are the consequences of it for higher order processes. We show that the edge effects in the ACM may cause a large bias in the residual variance if a reflection coefficient is close to one in absolute value. Propagation of this bias via the residual variance may lead to a large bias in higher order reflection coefficients.

II. THE AUTOCORRELATION METHOD

An autoregressive process $x(n)$ of order K is described by a weighted sum of preceding signal values plus an independent identically distributed (i.i.d.) noise signal $\varepsilon(n)$ with variance σ^2

$$x(n) + \sum_{i=1}^K a_i x(n-i) = \varepsilon(n), \quad (1)$$

Manuscript received December 1, 1995; revised September 24, 1996. This research was supported by the Technology Foundation (STW) under Grant DTN11.2436. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Douglas D. O'Shaughnessy.

The authors are with the Department of Applied Physics, Delft University of Technology, 2628 CJ Delft, The Netherlands (e-mail: broersen@tn.tudelft.nl).
 Publisher Item Identifier S 1063-6676(97)01893-2.

The coefficients a_i are the autoregressive parameters, called LPC parameters in speech coding. The autocorrelation method assumes the signal to be zero outside the interval of observation of length N and estimates the AR parameters of a p th order model by minimizing the residual variance $s_{[p]}^2$, from minus to plus infinity. This residual variance is given by

$$s_{[p]}^2 = \frac{1}{N} \sum_{n=-\infty}^{\infty} \{x_w(n) + a_1 x_w(n-1) + \dots + a_p x_w(n-p)\}^2 \quad (2)$$

where the $x_w(n)$ are samples of the windowed signal.

Because of the infinite sum in (2), the parameters can be found efficiently with the Levinson–Durbin algorithm

$$\hat{k}_m = -\frac{\hat{R}(m) + \sum_{j=1}^{m-1} a_j^{[m-1]} \hat{R}(m-j)}{s_{[m-1]}^2}, \quad m = 1, \dots, p \quad (3)$$

$$s_{[m]}^2 = s_{[m-1]}^2 (1 - k_m^2) \quad (4)$$

$$a_j^{[m]} = a_j^{[m-1]} + k_m a_{m-j}^{[m-1]}, \quad (5)$$

$$a_m^{[m]} = k_m$$

where \hat{k}_m is the m th estimated reflection coefficient, $a_j^{[m]}$ is the j th parameter of an m th order model and $s_{[m]}^2$ is the residual variance for the m th order model. The autocorrelation coefficients $\hat{R}(m)$ of the data are defined by

$$\hat{R}(m) = \frac{1}{N} \sum_{n=-\infty}^{\infty} x_w(n) x_w(n+m), \quad (6)$$

The autocorrelation coefficients $\hat{R}(m)$ are biased estimates of the theoretical autocorrelation coefficients $R(m)$ of the process, because $\hat{R}(m)$ contains only $(N-m)$ nonzero products. This ensures that the resulting model is always stable. There are other estimation methods that ensure a stable model [7]–[10], but the ACM is the most popular one. The bias in the autocorrelation coefficients is caused by the way the autocorrelation method handles the edges. In the next section, we will show that this method yields poor results if some of the reflection coefficients in the process are close to plus or minus one.

III. ANALYSIS OF THE AUTOCORRELATION METHOD

The bias in the sample autocorrelation function (6) also causes bias in the estimated reflection coefficients. But there are other sources of bias as well; the bias consists of different parts. For example, the solution for the first reflection coefficient \hat{k}_1 is given by

$$\hat{k}_1 = -\frac{\hat{R}(1)}{\hat{R}(0)} = -\hat{\rho}(1) \quad (7)$$

where $\hat{\rho}(1)$ is the normalized autocorrelation coefficient of lag one of the data. An expression for the bias in the reflection coefficients can be found by making a second order Taylor expansion and taking the expectation. For example, for (7) the result is

$$\begin{aligned} \mathcal{E}(\hat{k}_1) &\approx -\rho_w(1)\rho(1) - \frac{R_1}{R_0^3} \text{var}[\hat{R}(0)] \\ &\quad + \frac{1}{R_0^2} \text{cov}[\hat{R}(0), \hat{R}(1)] \end{aligned} \quad (8)$$

where $\rho(1)$ is the true normalized autocorrelation coefficient of lag one of the process, $\rho_w(1)$ is the normalized autocorrelation coefficient of lag one of the window, R_m is the expectation of $\hat{R}(m)$. $\text{Var}[\hat{R}(0)]$ and $\text{cov}[\hat{R}(0), \hat{R}(1)]$ are the variance of $\hat{R}(0)$ and the covariance between $\hat{R}(0)$ and $\hat{R}(1)$, respectively. First an analysis of the ACM will be given when a rectangular window is used. In speech applications, almost always a non-rectangular window is used. Yet, the analysis for a rectangular window is of interest, because it explains why the use of a nonrectangular window is necessary and what properties a window must possess when used with the ACM.

The first term on the right hand side of (8) shows the explicit contribution of windowing to the bias in \hat{k}_1 . For a rectangular window, $\rho_w(1)$ equals $(1 - 1/N)$ and since $\rho(1)$ equals $-k_1$ (the true first reflection coefficient of the process), this bias contribution is equal to $-1/N \times k_1$. For all reflection coefficients, the explicit bias contribution due to windowing can be found by multiplying each normalized autocorrelation coefficient of the process with the corresponding normalized autocorrelation coefficient of the window, and transforming the resulting modified normalized autocorrelation function to reflection coefficients. For a rectangular window, $\rho_w(m)$ equals $(1 - m/N)$ and, therefore, this bias contribution will be called *triangular bias* because of the specific form. The triangular bias is only present in the ACM and not in other estimation methods, such as the Burg method or the covariance methods. The second and third terms in (8) are due to the variances of and the covariances between the autocorrelation coefficients in the expressions and will be called *Taylor bias*. Such contributions of Taylor bias to the overall bias are present in the parameter estimates of all estimation methods and contain an implicit contribution due to windowing. Here, only the explicit triangular bias in the ACM is considered. Firstly, a qualitative analysis of the effects associated with the triangular bias will be given by looking at the residual for a first- and second-order model. Next, we will generalize the analysis and give an experimental illustration.

From (2) it is seen that the residual variance for a first order model is given by

$$\begin{aligned} Ns_{[1]}^2 &= x^2(1) + \sum_{n=2}^N \{x(n) + k_1x(n-1)\}^2 \\ &\quad + \{k_1x(N)\}^2. \end{aligned} \quad (9)$$

The value \hat{k}_1 minimizing this equation is the estimate of the first reflection coefficient. The discontinuities at the frame edges cause *incomplete* residual terms in s^2 such as $x^2(1)$ and $\{k_1x(N)\}^2$. The incomplete term $x^2(1)$ has no influence on the solution \hat{k}_1 . The incomplete term $\{k_1x(N)\}^2$ causes \hat{k}_1 to have a small bias that makes it closer to zero in absolute value. The influence of the incomplete term is small because *before* minimization with respect to k_1 , all terms $x^2(n)$ in the residual variance are of the same order of magnitude. The triangular bias in \hat{k}_1 is $1/N \times k_1$, the true first reflection coefficient of the process. Both incomplete terms $x^2(1)$ and $\{k_1x(N)\}^2$ are an important source of bias in the residual variance s^2 . If k_1 is close to one in absolute value, the magnitude of these incomplete terms *after* minimization is large in comparison with the magnitude of the complete terms [after minimization s^2 is reduced with a factor $(1 - k^2)$]. A consequence is that the influence of the incomplete terms on the second reflection coefficient is much larger than on the first one.

The residual variance for a second order model is

$$\begin{aligned} Ns_{[2]}^2 &= x^2(1) + \{x(2) + \hat{k}_1(1 + k_2)x(1)\}^2 \\ &\quad + \sum_{n=3}^N \{x(n) + \hat{k}_1(1 + k_2)x(n-1) \\ &\quad + k_2x(n-2)\}^2 + \{\hat{k}_1(1 + k_2)x(N) \\ &\quad + k_2x(N-1)\}^2 + \{k_2x(N)\}^2. \end{aligned} \quad (10)$$

The value \hat{k}_2 minimizing this expression, given the solution \hat{k}_1 for the first reflection coefficient, is the solution for k_2 . The complete terms in the sum are much smaller than the incomplete terms if k_1 is close to one in absolute value, because for the first order the complete terms have already been minimized with respect to \hat{k}_1 and $s_{[1]}^2$ is reduced with a factor $(1 - \hat{k}_1^2)$. Therefore, the incomplete terms have a much larger influence on \hat{k}_2 than they had on \hat{k}_1 . To put this another way, the incomplete terms cause $s_{[1]}^2$ to be considerably larger than σ^2 . The biased residual variance $s_{[1]}^2$ is propagated via the Levinson–Durbin algorithm (3) to the second reflection coefficient \hat{k}_2 , and this causes \hat{k}_2 to be considerably smaller than the true value k_2 . Higher order reflection coefficients cannot compensate for this large bias because incomplete terms are present for all model orders in $s_{[p]}^2$ and are large in comparison with the complete terms.

This effect will occur after the occurrence of a reflection coefficient close to one in absolute value. For example, consider a process of which the m th reflection coefficient has a distance of m/N from one, i.e., k_m equals $1 - m/N$ in absolute value. For reasons of clarity, it is supposed that all lower order reflection coefficients are equal to zero. The $(m + 1)$ th reflection coefficient may have an arbitrary value. For such a process k_m equals $-R(m)/R(0)$. The autocorrelation method

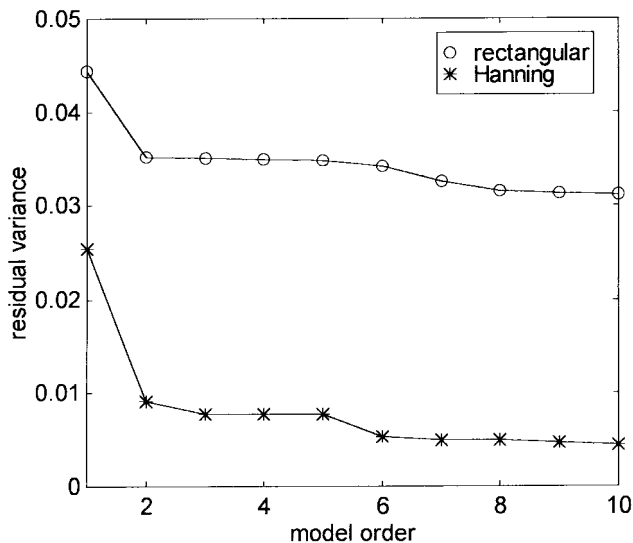


Fig. 1. Normalized residual variance for different model orders obtained with the autocorrelation method, with a rectangular window ("o") and a Hanning window ("*"). Edge effects in the autocorrelation method may cause serious bias in the residual variance and reflection coefficients, when a rectangular window is applied. A tapered window largely reduces this bias.

uses estimates of $R(m)$ that are biased with a factor $(1 - m/N)$, and, therefore, the reflection coefficient corresponding to this biased autocorrelation will be $(1 - m/N)k_m = (1 - m/N)^2 \approx (1 - 2m/N)$; a bias of only order $1/N$. If the biased value $(1 - 2m/N)$ for k_m is used in (4), $s_{[m]}^2$ will get a value *twice* the value it should have had. This leads to a value of k_{m+1} with (3) which is one half the true value and the result is a large bias of order 1 (as compared to the bias in $\hat{R}(m)$, which is of order $1/N$).

This example is of practical interest because the first reflection coefficient of sampled bandlimited data can be close to -1 . The first reflection coefficient equals $-R(1)/R(0)$ and this quotient approaches -1 if the continuous signal is sampled with a very high sampling frequency. In fact, the first estimated reflection coefficient often is close to -1 in voiced speech (and the second one close to $+1$). The conclusion to be drawn is that the autocorrelation method in its basic form without windowing is not a useful method for parameter estimation.

A tapered data window smoothly brings the amplitude of the signal down at the edges. Such windows reduce the bias in the autocorrelation function. Only the normalized autocorrelation coefficients of the window up to the LPC analysis order are of importance. For the rectangular window, these are equal to $1 - m/N$; for other windows such as the popular

Hamming and Hanning windows, these are much closer to one. This means that severe bias propagation will only occur with these windows if the reflection coefficients of the process are much closer to one in absolute value than m/N . The application of a tapered data window also has a disadvantage: When a data window is used, effectively, some information is thrown away. The result is that the variance of the estimated parameters is increased. We have observed that the increase in variance as a result of windowing corresponds roughly to the decrease in effective number of observations as defined in [11], which makes sense because the theoretical variances of LPC coefficients are inversely proportional to the number of observations.

At this point, it is interesting to compare the bias in the reflection coefficients with that in the log area ratios. An estimated model has a certain spectrum distortion with respect to the original process. It has been shown [12] that the reflection coefficients are more sensitive to quantization than the log area ratios if their absolute value is close to one. This means that the bias in the log area ratios is larger than the bias in the reflection coefficients, if the latter have an absolute value close to one. If the reflection coefficients are small in absolute value, they are approximately linearly related to the log area ratios [12]. In that case, the log area ratios have bias properties similar to the reflection coefficients, e.g., if the bias in a reflection coefficient is a factor one half, it also is approximately a factor one half in the corresponding log area ratio.

The bias propagation can be illustrated by a simulation experiment. We have generated 5000 realizations of length $N = 100$ samples of a tenth-order autoregressive process with the reflection coefficient vector in (11), shown at the bottom of the page.

The first reflection coefficient is equal to $-1 + 1/N$. From these realizations we have estimated LPC models with a rectangular window and with a Hanning window. The average reflection coefficient vectors for these 5000 realizations are, respectively, for the rectangular window and the Hanning window as shown in (12) and (13), at the bottom of the page.

In Fig. 1, the product of $(1 - k_j^2)$, $j = 1, \dots, i$, for orders $i = 1, \dots, 10$, is shown, which is the normalized residual variance. Results for the rectangular window are denoted by "o," results for the Hanning window are denoted by "*." From this figure, and from the average reflection coefficient vectors, the phenomenon of bias propagation can be clearly seen. The bias in the first estimated reflection coefficient is small, but the residual variance for the rectangular window is close to

$$[1 \quad -0.99 \quad 0.83 \quad -0.47 \quad 0.11 \quad 0.10 \quad 0.71 \quad 0.25 \quad -0.00 \quad -0.37 \quad 0.52] \quad (11)$$

$$[1 \quad -0.98 \quad 0.46 \quad 0.06 \quad -0.03 \quad -0.05 \quad 0.14 \quad 0.22 \quad 0.18 \quad 0.07 \quad 0.06] \quad (12)$$

$$[1 \quad -0.99 \quad 0.82 \quad -0.47 \quad 0.06 \quad 0.06 \quad 0.64 \quad 0.10 \quad -0.09 \quad -0.33 \quad 0.41] \quad (13)$$

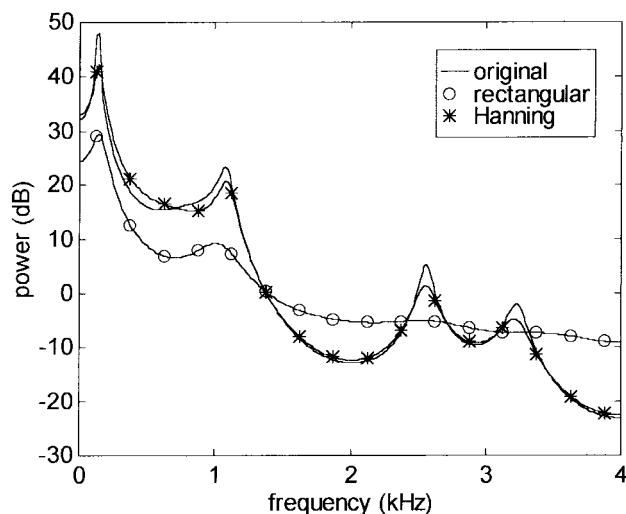


Fig. 2. Original spectrum (solid unmarked line) and spectra obtained with the autocorrelation method, with a rectangular window ("o") and a Hanning window ("*").

a factor 2 larger than for the Hanning window for this order, due to the influence of the triangular bias. The bias in the second estimated reflection coefficient consequently is large: for the rectangular window the second reflection coefficient is close to one half that obtained with the Hanning window. The decrease in residual variance for the rectangular window is therefore smaller than for the Hanning window.

The spectrum of the process that was used in the simulations is the unmarked solid line in Fig. 2. The spectra obtained with the rectangular and Hanning window are marked by "o" and "*", respectively. It is clear that the Hanning window greatly reduces the bias.

We have explained the influence on the reflection coefficients due to the edge effects in the ACM. It is difficult to relate this bias in the reflection coefficients quantitatively to the bias in individual formant frequencies and bandwidths. However, the bias always tends to make the reflection coefficients smaller in absolute value, and this means that in general the poles will stay away further from the unit circle. Therefore, an important effect of the bias propagation is an increase in the bandwidth of spectral peaks.

IV. CONCLUSIONS

It is shown that the autocorrelation method of autoregressive estimation is not suitable if reflection coefficients are close to $+1$ or -1 . This result is of practical interest because for sampled continuous signals the first reflection coefficient is often close to -1 unless the sample frequency is low. The poor performance of the autocorrelation method is due to edge effects; incomplete terms in the residual cause a large bias in the residual variance, this bias is propagated via the

denominator of the Levinson–Durbin recursion and causes higher order reflection coefficients to be seriously biased as well. A tapered data window decreases the edge effects, but increases the variance of estimated models, because effectively the number of data available for estimation is decreased.

REFERENCES

- [1] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*. New York: Springer-Verlag, 1976.
- [2] P. M. T. Broersen and H. E. Wensink, "On finite sample theory for autoregressive model order selection," *IEEE Trans. Signal Processing*, vol. 41, pp. 194–204, 1993.
- [3] D. Tjøstheim and J. Paulsen, "Bias of some commonly used time series estimates," *Biometrika* 70, pp. 389–399, 1983; correction in *Biometrika*, 71, p. 656.
- [4] J. Paulsen and D. Tjøstheim, "On the estimation of residual variance and order in autoregressive time series," *J. Roy. Stat. Soc. B*, vol. 47, pp. 216–228, 1985.
- [5] A. H. Gray and D. Y. Wong, "The Burg algorithm for LPC speech analysis/synthesis," *IEEE Trans. Signal Processing*, vol. 28, pp. 609–615, 1980.
- [6] L. R. Rabiner and R. W. Shafer, *Digital Processing of Speech Signals*. Prentice-Hall, 1978.
- [7] S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [8] B. S. Atal, "Predictive coding of speech at low bit rates," *IEEE Trans. Commun.*, vol. 30, pp. 600–614, 1982.
- [9] J. Makhoul, "Stable and efficient lattice methods for linear prediction," in *Modern Spectrum Analysis*, D. G. Childers, Ed. IEEE Press, 1978, pp. 256–261.
- [10] B. W. Dickinson, "Autoregressive estimation using residual energy ratios," *IEEE Trans. Inform. Theory*, vol. 24, pp. 503–506, 1978.
- [11] J. Makhoul, "Linear prediction: A tutorial review," *Modern Spectrum Analysis*, D. G. Childers, Ed. IEEE Press, 1978, pp. 99–118.
- [12] V. R. Viswanathan and J. Makhoul, "Quantization properties of transmission parameters in linear predictive systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 23, pp. 309–312, 1975.



Jan S. Erkelens was born in Maassluis, The Netherlands, in 1969. He received the M.Sc. degree in applied physics and the Ph.D. degree in 1996, both from the Delft University of Technology, The Netherlands. His Ph.D. thesis focuses on linear-prediction-based speech coding.

Currently, he holds a post-doctoral position at the International Research Centre of Telecommunications-Transmission and Radar IRCTR, Delft, where he is studying the micro-physical properties of clouds by means of radar measurements.



Piet M. T. Broersen was born in Zijdewind, The Netherlands, in 1944. He received the M.Sc. degree in applied physics and the Ph.D. degree in 1976, both from the Delft University of Technology, The Netherlands.

He is with the Department of Applied Physics of the Delft University of Technology. His research interests are the application of signal processing, time series analysis, and parameter estimation to model building.