Pole-Zero Modeling and its Applications to Speech Processing

by

Mohammad Ali Atashroo

August 1976

UTEC-CSc-76-271

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. DAHC15-73-C-0363.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Autocorrelation Pole-Zero modeling identifies the parameters of a rational transfer function H(z) whose short time-lag autocorrelations either exactly match (Autocorrelation Partial Realization) or closely approximate(Autocorrelation Prediction) those of a given spectrum. As a result, the spectrum of the H(z) obtained from either method approximates the gross structure of the given spectrum. Autocorrelation Partial Realization (APR) uses the Pade approximation to determine the denominator coefficients of H(z). To compute the numerator coefficients of H(z), APR uses an iterative algorithm such as Fejer's or Newton-Raphson's. In contrast, Autocorrelation Prediction (AP) uses only Linear Prediction (LP) to determine both the denominator and numerator coefficients. Therefore, once the autocorrelation function of the given spectrum is known, AP uses only linear operations and no Fourier Transformations to determine the parameters of H(z). Moreover, the resulting rational transfer function is guaranteed to be minimum phase and consequently stable. AP can also automatically determine the least (parsimonious) denominator and numerator orders required to model efficiently a given spectral envelope.

A dynamic filtering process, based on Wiener filtering and Autocorrelation Prediction, was developed to suppress the background noise from degraded speech. More important, using AP, a Linear

Predictive Vocoder was integrated into the so called "Pole-Zero Vocoder"(PZV). Computer simulations of both, the dynamic filtering process and the PZV were successfully used in speech processing.

CHAPTER 1

INTRODUCTION

1.1 Problem Presentation

Spectral Pole-Zero modeling has been the focus of many research efforts in recent years. The success of Linear Prediction [27,29] in All-Pole modeling of the spectral envelope [21,23] has encouraged the search for a technique of comparable success in spectral envelope Pole-Zero modeling. On the other hand, a high order All-Pole model is required to model the spectral envelope having deep valleys. This is so because a large number of poles are required to approximate a small number of zeros represented by these valleys. Therefore, to model a spectral envelope having deep valleys, the Pole-Zero model requires fewer parameters than the All-Pole model. This advantage of the Pole-Zero model over the All-Pole model is appreciated in data compression.

Different methods of Pole-Zero modeling have been proposed for matching the envelope or some smoothed version of a given spectrum. These methods may be divided into two groups. The first group encompasses those which estimate the pole and zero parameters simultaneously. In contrast, the methods encompassed in the second group determine the pole and zero parameters separately. The methods in the first group commonly face the problem of solving a system of non-linear equations in terms of the sought-for parameters. Whereas, those in the second group commonly use variations of Linear

Prediction to determine the pole and/or zero parameters. Linear Prediction itself leads to a system of linear equations solvable by fast recursive algorithms [27,29,Appendix A].

Some recent frequency domain methods of the second group are mentioned in the following. Cepstral Prediction [6,33,40] and Homomorphic Prediction [20,21] both use Homomorphic deconvolution [31] to initially smooth the given spectrum . Then each of them uses its own approach to model the cepstrally smoothed spectrum with a Pole-Zero model. Cepstral Prediction is successful when the resulting smoothed spectrum is some rational spectrum. Homomorphic Prediction was reported successful when applied to the short-time spectrum of natural speech [20,21]. Makhoul proposed a method [27] in which a variation of Linear Prediction [10,41] is used to determine the pole parameters. The zero parameters are then estimated by Inverse Linear Prediction [27]. This procedure basically applies Linear Prediction to reciprocal of the smoothed ratio of the given spectrum to the computed All-Pole spectrum.

Autocorrelation Partial Realization and Autocorrelation Prediction, described in chapter 3, are new techniques; modeling the envelope of a given spectrum with a Pole-Zero model. Both techniques determine the pole and zero parameters separately. Also, both are autocorrelation domain methods. In other words, once the autocorrelation function of the given spectrum is known, all the operations performed are in the autocorrelation domain and no Fourier Transformation is required.

Autocorrelation Prediction was successfully used for Pole-Zero

modeling in two applications: Wiener filter spectral matching and natural speech short-time spectral matching.

1.2 Chapter Summaries

The main body of the dissertation is devoted to development of Pole-Zero modeling techniques and their applications. The first part of the main body develops new techniques for spectral Pole-Zero modeling while the last part applies these techniques to natural speech processing. To make the dissertation more self-contained, a collection of Appendices is also added. These Appendices provide the basic algorithms and concepts used in developing the above techniques.

Chapters 2 and 3 comprise the first part, with chapter 2 providing the background needed for chapter 3. The new Pole-Zero modeling techniques, APR and AP, are derived in chapter 3. Chapters 4 and 5 comprise the last part, presenting the applications of these techniques. A dynamic filtering process that suppresses the background noise from degraded natural speech is described in chapter 4. Chapter 5 presents the "Pole-Zero Vocoder" (PZV); an analysis-synthesis process of natural speech based on Autocorrelation Prediction.

The author's majors contribution are: the Pole-Zero modeling techniques derived in chapter 3, the improved dynamic filtering process described in chapter 4, and the integration of the Linear Prediction Vocoder into the Pole-Zero Vocoder presented in chapter 5. Introducing a new spectral flatness measure, section 2.2, and preliminary work on a modified Pole-Zero Vocoder to account for

background noise in degraded natural speech, section 5.5, are some of the author´s minor contributions.

CHAPTER 2

PARAMETRIC MODELING OF SPECTRA

2.1 Introduction

The identification of a parametric model whose spectrum approximates a given spectrum by minimizing some distance measure is the general theme of the chapter. A new distance measure having an upper bound proportional to a well-known distance measure E is defined. Two types of parametric models, namely All-Pole and Pole-Zero, are focused upon. To show the effect of the order of the All-Pole model on the approximation, the relationship between the autocorrelation functions of the model and the given spectra is derived. Also, the dependence of the minimum of the distance measure E on the order of the All-Pole model is analyzed. Finally, it is shown that minimizing E to estimate the parameters of the Pole-Zero model leads to a system of non-linear equations, in contrast to the system of linear equations for the All-Pole model.

2.2 Spectral Matching by Inverse Filtering

The spectrum of the scaled model G H(z) can be matched to a given spectrum $|S(\omega)|^2$ by requiring that when S(z) is filtered by the inverse model $H^{-1}(z)$, the spectrum of the output E(z) is flat, Figure 2-1.

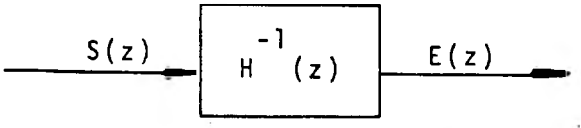


Figure 2-1. Inverse filtering.

The model scale factor squared G^2 is equal to the <u>average</u> of the inverse filter output spectrum E, i.e.

$$G^2 = E \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} |E(\omega)|^2 \, d\omega, \tag{2-1}$$

where, from Figure 2-1, the output spectrum $|E(\omega)|^2$ is equal to

$$|E(\omega)|^2 = |S(\omega)|^2 |H^{-1}(\omega)|^2. \tag{2-2}$$

A distance measure for the flatness of the output spectrum $|E(\omega)|^2$ can be quantified as: The average distance of the output spectrum $|E(\omega)|^2$ from its average E, i.e.

$$\varepsilon \triangleq \frac{1}{2\pi}\int_{-\pi}^{\pi} \Big| |E(\omega)|^2 - E \Big| d\omega. \tag{2-3a}$$

The smaller the distance measure ε, the flatter the output spectrum $|E(\omega)|^2$, and consequently there is a closer match between the scaled model spectrum, $|G\,H(\omega)|^2$, and the given spectrum $|S(\omega)|^2$. This non-negative distance measure has zero value when the output spectrum is constant, that is $E(\omega)=E$ for $-\pi<\omega<\pi$.

Properties of the Distance Measure ε. Using (2-2), and (2-1) in (2-3a), gives another expression for ε, i.e.

$$\varepsilon = \frac{1}{2}\int_{-\pi}^{\pi} E\left|S(\omega)\right|^2 \left| \left|\frac{1}{G} H^{-1}(\omega)\right|^2 - \left|S^{-1}(\omega)\right|^2 \right| d\omega. \qquad (2\text{-}3b)$$

The relation (2-3b) shows that, for a given spectrum $|S(\omega)|^2$, the optimum inverse model $H^{-1}(\omega)$ minimizes ε in weighted least mean amplitude sense. The weighting function is proportional to the given spectrum $|S(\omega)|^2$. Hence, from (2-3b), the scaled inverse model spectrum $|\frac{1}{G} H^{-1}(\omega)|^2$ approximates $|S^{-1}(\omega)|^2$ more accurately at those frequencies where the given spectrum $|S(\omega)|^2$ has its peaks rather than valleys.

An upper bound for ε may be found by using the fact that the absolute value of the difference has the least upper bound equal to the sum of the absolute values, i.e.

$$\left| |E(\omega)|^2 - E \right| \leq |E(\omega)|^2 + E. \qquad (2\text{-}4)$$

From (2-1) it is clear that E is always positive except for the trivial case, $E(\omega) = 0$ for all ω in which case $E = 0$. Thus for non-trivial $E(\omega)$, the equality in (2-4) holds only at those frequencies where $E(\omega)=0$. For non-trivial $E(\omega)$, from (2-4) and (2-3a), we obtain

$$\varepsilon < \frac{1}{2\pi}\int_{-\pi}^{\pi}[|E(\omega)|^2 + E]d\omega. \qquad (2\text{-}5a)$$

Performing the integration in (2-5a) and using (2-1) gives the following upper bound for ε.

$$\varepsilon < 2E. \qquad (2\text{-}5b)$$

Thus the flatness distance measure ε has an upper bound proportional to E, the average of the inverse filter output spectrum. Note that the model $H_0(z)$ which minimizes the upper bound 2E <u>does not</u> necessarily minimize the flatness distance measure ε. It is guaranteed, nevertheless, that the flatness distance measure ε_0 for $H_0(z)$ <u>does not exceed</u> the smallest upper bound $2E_{min}$. Since identification of the model which minimizes ε is complicated, we seek only the optimal model which minimizes the <u>upper bound</u>, 2E, of the distance measure ε.

<u>Other Interpretations For E</u>. Using (2-2) in (2-1) gives another interpretation for E, i.e.

$$E = \frac{1}{2\pi}\int_{-\pi}^{\pi}|S(\omega)|^2 \Big/ |H(\omega)|^2 d\omega. \qquad (2\text{-}6)$$

Thus E is the average of the ratio of the given spectrum to the model spectrum. Hence, for the optimal model H(z), the average of the ratio of the given spectrum to the model spectrum is minimum.

According to Parseval's Theorem [32,36], E given by (2-1), is also the energy in the inverse filter output signal E(z).

2.3 All-Pole Spectral Matching

Let H(z) in Figure (2-1) of section 2.2 be a stable All-Pole model defined as

$$H(z) = \frac{1}{A(z)} = \frac{1}{\sum_{k=0}^{M} a(k)z^{-k}}, \qquad a(0) = 1. \tag{2-7}$$

To match the spectrum of the scaled H(z) to the given spectrum $|S(\omega)|^2$, the prediction coefficients $\{a(k)\}$ are computed to flatten the output spectrum $|E(\omega)|^2$ by minimizing the output energy E, [24,25]. The spectrum of the resulting scaled H(z) is called the Linear Prediction (LP) Spectrum.

Using (2-7) in (2-6) gives

$$E = \frac{1}{2\pi}\int_{-\pi}^{\pi} |S(\omega)|^2 |A(\omega)|^2 d\omega. \tag{2-8}$$

E is minimized by setting

$$\frac{\partial E}{\partial a(i)} = 0, \qquad 1 \le i \le M. \tag{2-9}$$

On the other hand, the following relation holds

$$\frac{\partial}{\partial a(i)} |A(\omega)|^2 = \frac{\partial}{\partial a(i)} [A(e^{jk\omega})A(e^{-jk\omega})]$$

$$= [e^{-ji\omega}(\sum_{k=0}^{M} a(k)e^{jk\omega}) + e^{ji\omega}(\sum_{k=0}^{M} a(k)e^{-jk\omega})],$$

$$1 \le i \le M,$$

or

$$\frac{\partial}{\partial a(i)}|A(\omega)|^2 = 2\sum_{k=0}^{M} a(k)\mathrm{Cos}(i-k)\omega, \qquad 1 \leq i \leq M. \tag{2-10}$$

From (2-8), (2-9) and (2-10) and interchanging of summation and integration, one obtain

$$\sum_{k=0}^{M} a(k)\,\frac{1}{2\pi}\int_{-\pi}^{\pi} |S(\omega)|^2 \mathrm{Cos}(i-k)\omega\, d\omega = 0, \qquad 1 \leq i \leq M. \tag{2-11}$$

We know that the autocorrelation function R(k) is the Inverse Fourier Transform of the spectrum $|S(\omega)|^2$, i.e.

$$R(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |S(\omega)|^2 e^{-jk\omega} d\omega, \tag{2-12a}$$

or in a more simplified form

$$R(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |S(\omega)|^2 \mathrm{Cos}(k\omega) d\omega. \tag{2-12b}$$

Substituting (2-12b) in (2-11) results in the well-known autocorrelation normal equations

$$\sum_{k=0}^{M} a(k)R(i-k) = 0, \qquad 1 \leq i \leq M. \tag{2-13}$$

The optimum predictor coefficients {a(k)} are obtained by solving the system of linear equations (2-13) using Levinson's [37] or Durbin's [27] recursive algorithms [Appendix A]. Using (2-13), (2-12) in (2-8) and some simplification gives the minimum output energy E_M

$$E_M = G_A^2 = \sum_{k=0}^{M} a(k)R(k). \tag{2-14}$$

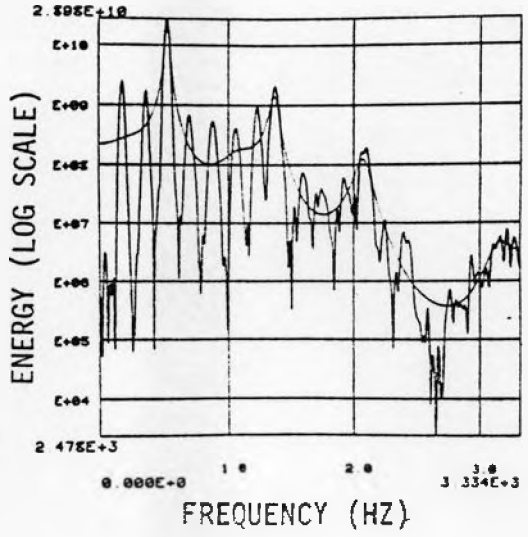
Therefore, from Section 2.2, the spectrum of the scaled H(z), i.e.

$$|G_A H(e^{jk\omega})|^2 = \frac{G_A^2}{|A(\omega)|^2} = \frac{G_A^2}{\left|\sum_{k=0}^{M} a(k)e^{-jk\omega}\right|^2}, \tag{2-15}$$

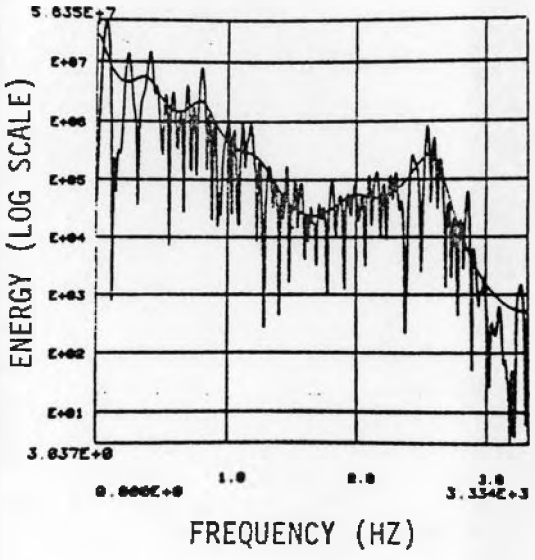
is the optimal All-Pole spectral match to the given spectrum $|S(\omega)|^2$, Figure 2-2. The predictor coefficients {a(k)} in (2-15) and the scale factor squared G_A^2 are computed from (2-13) and (2-14), respectively.

2.4 Autocorrelation Function of the Optimal All-Pole Spectrum

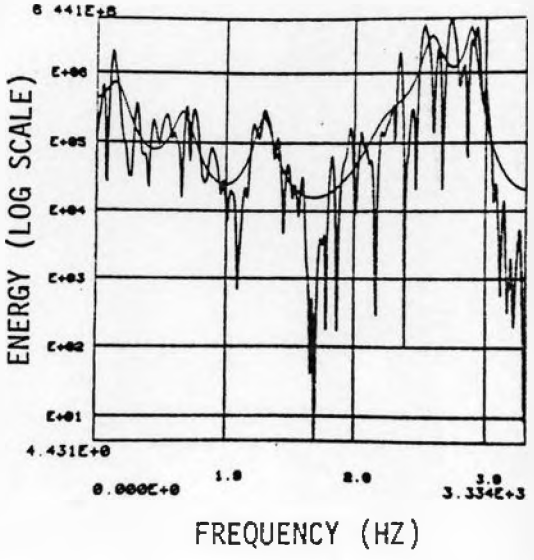
The relationship between the autocorrelation functions of the optimal All-Pole spectrum $G_A^2/|A(\omega)|^2$ and the given spectrum $|S(\omega)|^2$ is discussed here. Rearranging (2-15) gives



(a)



(b)



(c)

Figure 2-2. 14-Pole LP spectra superimposing :

(a) The given short-time spectrum of the voiced sound [O].

(b) The given short-time spectrum of the unvoiced sound [P].

(c) The given short-time spectrum of the unvoiced sound [f].

$$|G_A H(e^{jk\omega})|^2 \cdot A(\omega) = G_A^2 \cdot \frac{1}{A^*(\omega)}, \qquad (2\text{-}16)$$

Where "*" indicates the complex conjugate. Taking the Inverse Fourier Transform of both sides of (2-16) results in the difference equation that the autocorrelation function of the scaled optimal All-Pole model obeys, i.e.

$$\sum_{k=0}^{M} a(k)\hat{R}(i-k) = G_A^2 h(-i), \qquad (2\text{-}17)$$

where

$$\hat{R}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{G_A^2}{|A(\omega)|^2}\, e^{-jk\omega} d\omega, \qquad (2\text{-}18a)$$

or

$$\hat{R}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{G_A^2}{|A(\omega)|^2}\, \mathrm{Cos}(k\omega)\, d\omega, \qquad (2\text{-}18b)$$

and

$$h(-i) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{A^*(\omega)}\, e^{-ji\omega} d\omega. \qquad (2\text{-}18c)$$

Since h(i) is casual and from (2-7)

$$h(0) = 1, \tag{2-18d}$$

then for $i \geq 0$ (2-17) reduces to

$$\sum_{k=0}^{M} a(k)\hat{R}(i-k) = 0, \qquad i > 0, \tag{2-19}$$

$$\sum_{k=0}^{M} a(k)\hat{R}(k) = G_A^2, \qquad i = 0 \tag{2-20}$$

The system of equations given by (2-19) for $1 \leq i \leq M$ and (2-20) has the same form as the system of equations given by (2-13) and (2-20). Therefore, these two systems of equations have identical solutions, i.e.

$$\hat{R}(|i|) = R(|i|). \qquad 0 \leq i \leq M \tag{2-21}$$

Thus, matching the spectrum of a scaled All-Pole model of order M to a given spectrum is equivalent to finding an All-Pole model whose scaled autocorrelation function exactly matches that of the given spectrum for the first M+1 time-lags.

Equations (2-21) show that the first M+1 autocorrelations of the matching All-Pole spectrum $G_A^2/|A(\omega)|^2$ are <u>exactly</u> the same as those of the given spectrum $|S(\omega)|^2$. The rest of the autocorrelation $\hat{R}(k)$ for $|k|>M$ are the extrapolation of $R(k)$ for $1 \leq k \leq M$, and are determined recursively from (2-19) using the optimum predictors $\{a(k)\}$. $|S(\omega)|^2$ and $G_A^2/|A(\omega)|^2$ are the Fourier Transforms of the

autocorrelation functions R(k) and $\hat{R}(k)$, respectively. Therefore, increasing the order M of the matching All-Pole spectrum increases the range of time-lags over which (2-21) holds, resulting in a closer match of the $G_A^2/|A(\omega)|^2$ to $|S(\omega)|^2$. Hence, as $M\to\infty$ we obtain

$$\frac{G_\infty^2}{|A_\infty(\omega)|^2} \triangleq \frac{G_\infty^2}{\sum_{k=0}^{\infty} a(i)e^{-jk\omega}} = |S(\omega)|^2. \qquad (2\text{-}22)$$

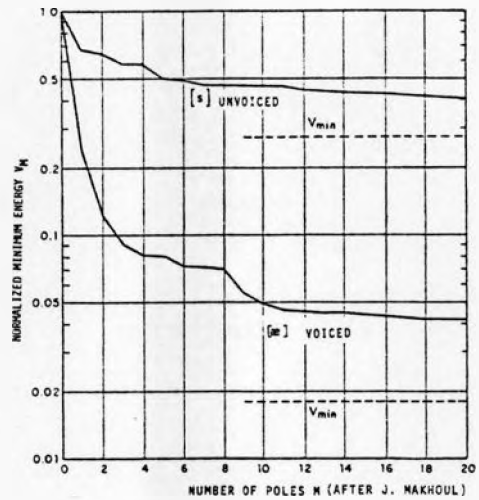
The $G_\infty^2/|A_\infty(\omega)|^2$ can be considered the All-Pole representation of the given spectrum $|S(\omega)|^2$ [35].

2.5 Analysis of the Minimum Energy E_M

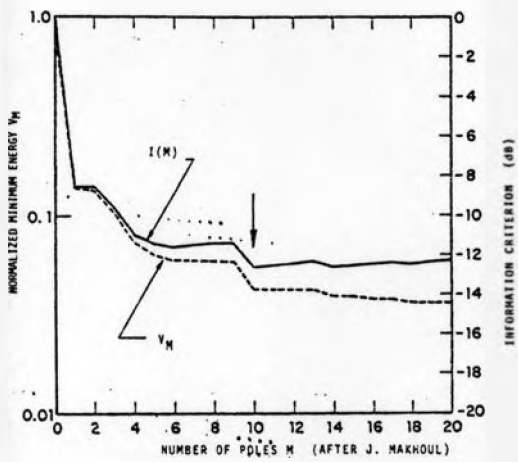
The minimum energy E_M, given by (2-14), of the inverse filter output is a monotonically decreasing function of the All-Pole model order M. Figure 2-3 shows the normalized minimum energy

$$v_M = E_M/R(0), \qquad (2\text{-}23)$$

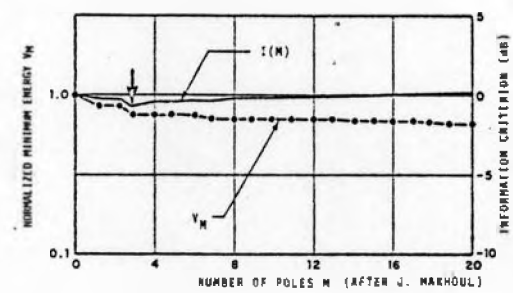
as a function of the order M. It can be shown that $0 \le v_M \le 1$ for all M [25,42]. The minimum energy decreases sharply as the order M is incresed up to some order M_P. For increases of the All-Pole model order beyond M_P, the minimum energy E_M decreases slightly. M_P is referred to as the parsimonious (most economical) order. By the time the order M has reached M_P, the scaled All-Pole model spectrum matches all the spectral envelope peaks of the given spectrum. The parsimonious number of poles M_P, therefore, depends on the number of spectral envelope peaks of the given spectrum.



(a)



(b)



(c)

Figure 2-3. (a) Normalized minimum energy curves for speech short-time spectra (voiced and unvoiced).
(b) Akaike's Information Criterion I(M) for voiced speech. The parsimonious order M_p occurs at the global minimum of I(M), shown by the arrow at M=10.
(c) Akaike's Information Criterion I(M) for unvoiced speech. The parsimonious order M_p occurs at the global minimum of I(M), shown by the arrow at M=3.

To estimate M_P for a given spectrum, as the order M of the All-Pole model increases, one might monitor the normalized minimum energy given by (2-23). This monitoring can be done automatically using Durbin's recursive algorithm, a simplification of Levinson's algorithm [Appendix A]. Durbin's recursive algorithm computes the minimum energy E_M for successive orders as a by-product of the calculation of the predictor coefficients.

Another approach to estimate M_P, when the length of the time sequence that generates the given spectrum is known, was proposed by Akaike [1,2,3,4,15]. Makhoul applied this approach to estimate M_P in speech spectral matching [26]. In this approach, the so-called Akaike's Information Criterion I(M), given by (2-24) below, has its global minimum at M_P.

$$I(M) = \text{Log } v_M + \frac{2M}{cW} . \qquad (2\text{-}24)$$

The W in (2-24) is the time domain window length and the constant c, $0<c\leq 1$, accounts for the effective length of the time domain window (Makhoul reports c = o.4 for a Hamming window), Figure 2-3.

To match the spectral envelope peaks and valleys of the given spectrum, a large increase in the order M beyond M_P is required. This is so, because the spectral zeros represented by the deep valleys require a large number of poles to approximate them. To avoid the large increase in M beyond M_P, zeros are introduced explicitly into the model. In other words, the given spectrum is matched with the spectrum of a Pole-Zero model rather than an All-Pole model.

2.6 Pole-Zero Spectral Matching [27]

Let H(z) in figure (2-1) of section 2.2 be a Pole-Zero model defined as

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{i=0}^{L} b(i)z^{-i}}{\sum_{i=0}^{M} a(i)z^{-i}}, \qquad \begin{matrix} a_0 = 1, \\ b_0 = 1. \end{matrix} \qquad (2\text{-}25)$$

To match the spectrum of scaled H(z) to the given spectrum $|S(\omega)|^2$, the optimal pole predictors $\{a(k)\}$ and the zero predictors $\{b(k)\}$ are computed by minimizing the output energy E, [27]. Using (2-25) in (2-6) gives

$$E = \frac{1}{2\pi}\int_{-\pi}^{\pi} |S(\omega)|^2 \frac{|A(\omega)|^2}{|B(\omega)|^2}\, d\omega. \qquad (2\text{-}26)$$

E is minimized by setting

$$\frac{\partial E}{\partial a(i)} = 0, \qquad 1 \le i \le M, \qquad (2\text{-}27a)$$

$$\frac{\partial E}{\partial b(i)} = 0, \qquad 1 \le i \le L. \qquad (2\text{-}27b)$$

On the other hand, similar to (2-10), the following relation holds:

$$\frac{\partial}{\partial b(i)} |B(\omega)|^2 = 2\sum_{k=0}^{L} b(k)\mathrm{Cos}(i-k)\omega, \quad 1 \le i \le L. \qquad (2\text{-}28)$$

We also define

$$R_{\alpha\beta}(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |S(\omega)|^2 \frac{|A(\omega)|^{2\alpha}}{|B(\omega)|^{2\beta}} \cos(k\omega)\,d\omega, \qquad (2\text{-}29)$$

where α and β are positive integers. Using (2-26), (2-10) and (2-29) in (2-27a) and some simplification results in

$$\sum_{k=0}^{M} a(k)R_{01}(i-k) = 0, \qquad 1 \le i \le M. \qquad (2\text{-}30a)$$

Similarly, using (2-26), (2-28) and (2-29) in (2-27b) and some simplifying results in

$$\sum_{k=0}^{L} b(k)R_{12}(i-k) = 0, \qquad 1 \le i \le L. \qquad (2\text{-}30b)$$

From (2-29) it is clear that $R_{01}(k)$ is a function of the $\{b(i)\}$ whereas $R_{12}(k)$ is a function of both $\{a(k)\}$ and $\{b(k)\}$. Consequently, the system consisting of equations (2-30a) and (2-30b) is non-linear in terms of $\{a(k)\}$ and $\{b(k)\}$. To solve this system of non-linear equations, an iterative scheme can be used [27,36]. An iterative scheme, however, brings about its own problems, such as convergence, stability, and high rates of computation.

To avoid these problems, in the next chapter we seek suboptimal Pole-Zero models whose parameters are partially or totally solutions to systems of linear equations.

CHAPTER 3

AUTOCORRELATION FUNCTION POLE-ZERO MODELING

3.1 Introduction

Identification of a Pole-Zero model whose autocorrelation function approximates that of a given spectrum is the main issue of this chapter. To see the relationship between the parameters and the autocorrelation function of a Pole-Zero model, the difference equation governing the autocorrelation function of the Pole-Zero model is derived. Two different techniques, "Autocorrelation Partial Realization" (APR) and "Autocorrelation Prediction" (AP), are developed to estimate the parameters of a Pole-Zero model whose autocorrelation function approximates that of a given spectrum. In either technique, the pole parameters and the zero parameters are estimated separately .

APR uses the Pade´ approximation to estimate the pole parameters and an iterative method to estimate the zero parameters by solving a system of non-linear equations. In contrast, AP uses Linear Prediction to estimate both the pole and the zero parameters. The spectral interpretation of AP is given and the selection of orders of the Pole-Zero model are discussed. Finally, APR and AP are compared.

3.2 Autocorrelation Function of Pole-Zero Model

To understand the relation between the autocorrelations of a stable Pole-Zero model, the difference equation that governs the

Pole-Zero model is derived.

Consider the Pole-zero model:

$$H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{i=0}^{L} b_i z^{-i}}{\sum_{i=0}^{M} a_i z^{-i}}, \qquad a_0 = 1, \tag{3-1a}$$

whose power series representation is:

$$H(z) = \sum_{k=0}^{\infty} h(k) z^{-k}, \qquad |z| > r, \quad r < 1, \tag{3-1b}$$

where, from (3-1a) and (3-1b), the coefficients h(k) are obtained from the following recursive formula:

$$h(k) = \sum_{i=0}^{L} b(i)\delta(k-i) - \sum_{i=1}^{M} a(i)h(k-i). \tag{3-2}$$

The $\{a_i\}$ and $\{b_i\}$ are referred to as pole predictors and zero predictors, respectively. Multiplying both sides of (3-1a) by H(1/z)A(z) gives:

$$[H(z)H(1/z)]A(z) = B(z)H(1/z). \tag{3-3}$$

Taking the Inverse Z-transform of both sides of (3-3) and using (3-1a) and (3-1b) results in the difference equation:

$$\sum_{i=0}^{M} a(i)\, \hat{R}(k-i) = \sum_{i=0}^{L} b(i)\, h(i-k), \tag{3-4a}$$

or

$$\hat{R}(k) = \sum_{i=0}^{L} b(i)h(i-k) - \sum_{i=1}^{M} a(i)\hat{R}(k-i)\,, \qquad (3\text{-}4b)$$

where the model autocorrelation function $\hat{R}(k)$ is defined by:

$$\sum_{k=-\infty}^{+\infty} \hat{R}(k)z^{-k} \triangleq H(z)H(1/z)\,, \qquad r < |z| < 1/r\,, \qquad (3\text{-}5)$$

and from (3-1b) the power series representation of $H(1/z)$ is:

$$H(1/z) = \sum_{k=0}^{\infty} h(k)z^{k} = \sum_{k=-\infty}^{0} h(-k)z^{-k}\,, \qquad |z| < \frac{1}{r}\,. \qquad (3\text{-}6)$$

Since $h(k)$ is causal, then the difference equation (3-4) is further simplified for $k > L$. That is,

$$\sum_{i=0}^{M} a(i)\hat{R}(k-i) = 0\,, \qquad k > L\,. \qquad (3\text{-}7)$$

Comparing (3-4b) and (3-2) reveals that for $k > L$ the same pole predictors $\{a_i\}$ predict both $\hat{R}(k)$ and $h(k)$ from their corresponding last values.

3.3 Autocorrelation Partial Realization

The idea is to find the Pole-Zero model $H(z)$, of the form (3-1a), whose autocorrelation function $\hat{R}_k$ exactly matches the symmetric autocorrelation function R_k, of a given spectrum $S(z)S(1/z)$, for the first N time-lags. That is,

$$\hat{R}_{|k|} = R_{|k|} , \qquad k = 0,1,2,\ldots N-1 , \qquad (3-8)$$

The autocorrelation function R_k is the Inverse Z-transform of the given spectrum $S(z)S(1/z)$, i.e.

$$\sum_{k=-\infty}^{+\infty} R_k z^{-k} = S(z)S(1/z) , \qquad r_1 < |z| < \frac{1}{r_1} , \qquad (3-9)$$

$$r_1 < 1 .$$

and similarly, $\hat{R}_k$ is the Inverse Z-transform of the model spectrum $H(z)H(1/z)$:

$$\sum_{k=-\infty}^{+\infty} \hat{R}_k z^{-k} = H(z)H(1/z) = \frac{B(z)B(1/z)}{A(z)A(1/z)} , \quad r_2 < |z| < \frac{1}{r_2} , \qquad (3-10)$$

$$r_2 < 1 .$$

The idea is realized in three steps, Figure 3-1:

i. We show that using the Pade´ approximation [Appendix B] on the right half of the autocorrelation function R_k leads to a rational function $C(z)/D(z)$ whose denominator is equal to the sought for denominator $A(z)$. Furthermore, $P(z)$, the numerator of the two-sided rational function $C(z)/D(z)+C(1/z)/D(1/z)$ is shown to be the spectrum of the desired numerator $B(z)$.

ii. Some properties of the polynomial $P(z)$ are discussed, and a direct method for computing its coefficients is derived.

iii. Some iterative algorithm such as Fejer´s [37] or Newton-Raphson´s [43, Appendix D] is used to decompose the $P(z)$ into

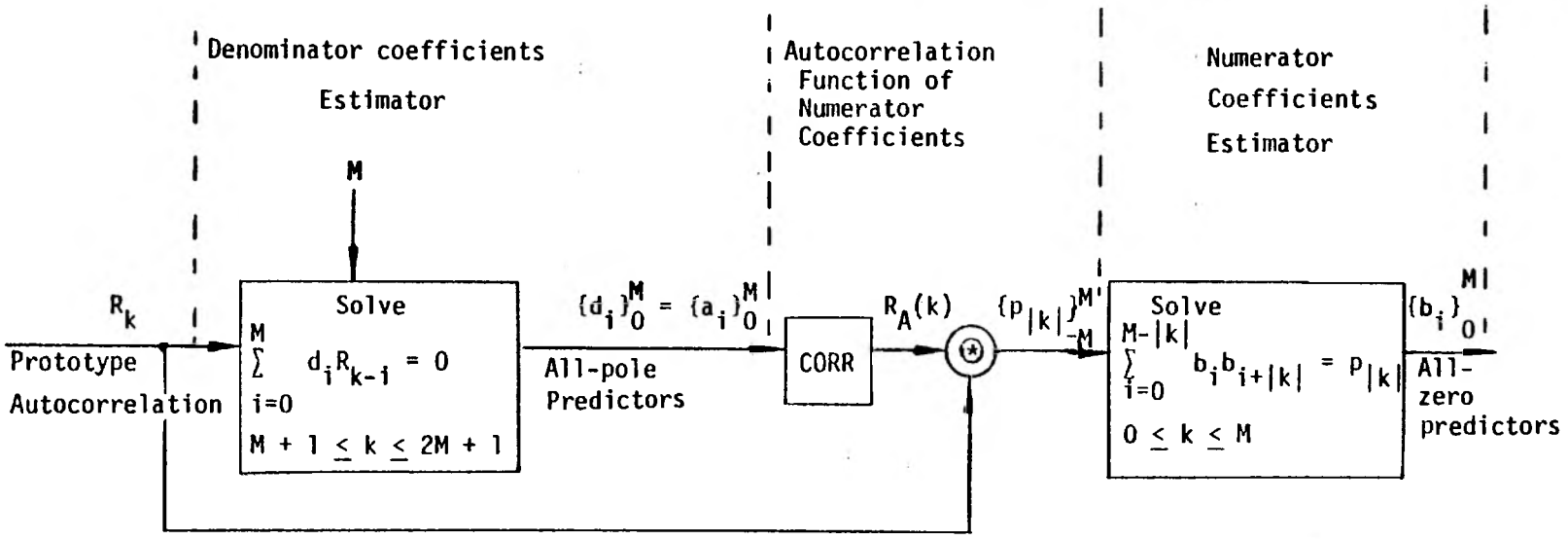


Figure 3-1. Autocorrelation Partial Realization block diagram.

$B(z)B(1/z)$.

Finally some issues invloved in Autocorrelation Partial Realiztion are discussed.

i. Pade Approximation. Consider the one-sided sequence x_k defined as:

$$x_k = \begin{cases} \frac{1}{2} R_k & k = 0 \\ R_k & k > 0 . \end{cases} \qquad (3\text{-}11)$$

Then the two-side power series $\sum_{k=-\infty}^{+\infty} R_k z^{-k}$ can be decomposed into the sum of the following two, one-sided power series:

$$\sum_{k=-\infty}^{+\infty} R_k z^{-k} = \sum_{k=0}^{\infty} x_k z^{k} + \sum_{k=0}^{\infty} x_k z^{-k}, \qquad r_1 < |z| < \frac{1}{r_1}, \qquad (3\text{-}12)$$

$$r_1 < 1 .$$

Furthermore, the right most power series in (3-12) is denoted by $X(z)$. That is,

$$X(z) = \sum_{k=0}^{\infty} x_k z^{-k}, \qquad |z| > r_1, \qquad (3\text{-}13)$$

$$r_1 < 1 .$$

Now , using the Pade´ approximation [Appendix B], the one-sided power series $X(z)$ is approximated with the stable rational function $\hat{X}(z)$:

$$\hat{X}(z) \triangleq \frac{C(z)}{D(z)} = \frac{\sum_{k=0}^{M} c_i z^{-i}}{\sum_{i=0}^{M} d_i z^{-i}} = \sum_{k=0}^{\infty} \hat{x}_k z^{-k}, \qquad d_0 = 1, \quad |z| > r_2, \quad r_2 < 1, \tag{3-14}$$

where, from (3-14), the coefficients $\hat{x}_k$ of the C(z)/D(z) power series representation are given by

$$\hat{x}_k = \sum_{i=0}^{M} c_i \delta(k-i) - \sum_{i=1}^{M} d_i \hat{x}_{k-i} \, . \tag{3-15}$$

The Berlekamp-Massey [7,30] or Trench [39,41] recursive algorithms [Appendix C] can be used to perform the Pade´ approximation, leading to the rational function $\hat{X}(z)=C(z)/D(z)$. As shown in Appendix C, the Pade´ approximation $\hat{X}(z)$ for the power series X(z) has the following property:

$$\hat{x}_k = x_k \, , \qquad k = 0, 1, 2, \ldots N-1 \, , \tag{3-16}$$

where

$$N = 2M + 1. \tag{3-17}$$

Also from (3-17) and (3-15) one concludes that

$$\hat{x}_k = - \sum_{i=1}^{M} d_i \hat{x}_{k-i} \, , \qquad k \geq N \, . \tag{3-18}$$

Therefore, $\hat{X}(z)=C(z)/D(z)$ approximates the given power series $X(z)$, satisfying the equations (3-16) and (3-18). In other words, the first N coefficients of the power series $\hat{X}(z)$ are <u>equal</u> to those of the power series $X(z)$ and the rest are recursive extrapolations of these first N coefficients. Thus,

$$\sum_{k=0}^{\infty} x_k z^{-k} \simeq \sum_{k=0}^{\infty} \hat{x}_k z^{-k} = \frac{C(z)}{D(z)}, \qquad |z| > \mathrm{Max}(r_1,r_2), \quad \mathrm{Max}(r_1,r_2) < 1. \tag{3-19}$$

We now find a two-sided rational spectrum whose Inverse Z-transform satisfies the equation (3-8). Using (3-19) in (3-12) and some simplifying operations results in

$$\sum_{k=-\infty}^{+\infty} R_k z^{-k} \simeq \sum_{k=0}^{\infty} \hat{x}_k z^{k} + \sum_{k=0}^{\infty} \hat{x}_k z^{-k} = \frac{C(1/z)D(z)+C(z)D(1/z)}{D(z)D(1/z)} = \frac{P(z)}{D(z)D(1/z)}, \tag{3-20a}$$

$$\mathrm{Max}(r_1,r_2) < |z| < 1/\mathrm{Max}(r_1,r_2), \quad \mathrm{Max}(r_1,r_2) < 1,$$

or

$$\sum_{k=-\infty}^{\infty} R_k z^{-k} \simeq \sum_{k=-\infty}^{\infty} \hat{R}_k z^{-k} = \frac{C(1/z)D(z)+C(z)D(1/z)}{D(z)D(1/z)} = \frac{P(z)}{D(z)D(1/z)},$$

$$\mathrm{Max}(r_1,r_2) < |z| < 1/\mathrm{Max}(r_1,r_2), \quad \mathrm{Max}(r_1,r_2) < 1, \tag{3-20b}$$

where

$$\hat{R}_{|k|} \triangleq \begin{cases} 2\hat{x}_k & k = 0 \\ \hat{x}_k & k \neq 0 . \end{cases} \qquad (3\text{-}21)$$

The two-sided rational spectrum $P(z)/D(z)D(1/z)$, given in (3-20b), has the Inverse Z-transform $\hat{R}_k$ for which, from (3-21), (3-16) and (3-11), the relation (3-8) holds. Since $P(z)/D(z)D(1/z)$ is the spectrum of the desired Pole-Zero model $H(z)$, then:

$$\frac{B(z)B(1/z)}{A(z)A(1/z)} = \frac{P(z)}{D(z)D(1/z)} \cdot \qquad (3\text{-}22)$$

From (3-22) the model denominator $A(z)$ is equal to

$$A(z) = D(z) , \qquad (3\text{-}23)$$

while, from (3-22) and (3-20) the spectrum of the model numerator $B(z)$ is

$$B(z)B(1/z) = P(z) \triangleq C(1/z)D(z) + C(z)D(1/z) \cdot \qquad (3\text{-}24)$$

ii. Properties of the $P(z)$. Before attempting to decompose the polynomial $P(z)$, we examine some properties of $P(z)$ and give a direct method for computing $P(z)$. From (3-24) and (3-14), it can be shown that $P(z)$ is a symmetric two-sided polynomial with real coefficients. That is,

$$P(z) = \sum_{k=-M}^{M} p_{|k|} z^{-k}, \tag{3-25}$$

where

$$p_{|k|} = \sum_{i=0}^{M-k} c_i d_{i+k} + \sum_{i=0}^{M-k} d_i c_{i+k}, \qquad k = 0, 1, \ldots M. \tag{3-26}$$

Therefore P(z) has real values on the unit circle in the z plane.

To compute the polynomial P(z) directly, consider the symmetric two-sided power series U(z) defined as

$$U(z) = D(z)D(1/z) \sum_{k=-\infty}^{+\infty} R_k z^{-k} = \sum_{k=-\infty}^{+\infty} U_k z^{-k}, \tag{3-27}$$

$$r_1 < |z| < 1/r_1,$$

$$r_1 < 1.$$

From (3-20b) we also have

$$P(z) = D(z)D(1/z) \sum_{k=-\infty}^{+\infty} \hat{R}_k z^{-k}, \qquad r_2 < |z| < 1/r_2, \tag{3-28}$$

$$r_2 < 1.$$

Similar to P(z), D(z)D(1/z) is also a symmetric two-sided polynomial of order M. From this property of D(z)D(1/z) and using the equations (3-8) and (3-17) in comparing (3-27) with (3-28), one concludes that

$$p_{|k|} = u_{|k|}, \qquad k = 0, 1, \ldots M. \tag{3-29}$$

(3-29) shows P(z), given by (3-25), is obtained by truncating the power series U(z), i.e.

$$P(z) = \sum_{k=-M}^{M} p_{|k|} z^{k} = \sum_{k=-M}^{M} u_{k} z^{-k} \tag{3-30}$$

Thus, the polynomial coefficients $p_{|k|}$ can be computed directly by convolving the autocorrelation function of the A(z) coefficients with that of the given spectrum.

iii. Decomposition of the P(z). To decompose P(z) into B(z)B(1/z), Fejer's [37] or Newton-Raphson's iterative algorithm [Appendix D] can be used. Fejer's algorithm finds the 2M roots of the symmetric polynomial P(z) and properly chooses M of the them to construct the polynomial B(z). The latter algorithm finds the minimum phase polynomial B(z) whose spectrum approximates P(z) with desired accuracy.

Discussion. A few issues concerning Autocorrelation Partial Realization deserve attention. These are: a. the stability of the Pole-Zero model B(z)/A(z), b. the condition that P(z) should meet to make its decomposition into B(z)B(1/z) possible, and finally, c. the non-linear methods required to perform the decomposition in b.

a. To have an stable Pole-Zero model B(z)/A(z), the rational

function C(z)/D(z) resulting from the Pade approximation should be stable . We used this assumption to show that equation (3-8) holds for the Inverse Z-transform of P(z)/D(z)D(1/z). For a given N, the rational function C(z)/D(z), however, is not guaranteed to be stable. Because of the equations (3-11) and (3-8), one hopes that by increasing N and consequently M, one finally finds a stable C(z)/D(z) and as a result a stable Pole-Zero model B(z)/A(z).

b. To decompose the P(z) into B(z)B(1/z), the polynomial P(z) should be non-negative on the unit circle. Why this condition should be met is verified easily from (3-24) where B(z)B(1/z) is non-negative on the unit circle. On the other hand, P(z) is equal to the truncated U(z). Though U(z) is non-negative on the unit circle, this is not necessarily true for the truncated U(z). Because of (3-27) and (3-30), the P(z) can be made non-negative on the unit circle, however, by choosing some high order M.

c. Due to the non-linear equation (3-24), the composition in (b) requires an iterative method. Consequently, the problems of non-negativeness, convergence and the high rates of computation should be dealt with.

3.4 Autocorrelation Prediction

The idea is to find a minimum phase Pole-Zero model of the form

$$H(z) = G\,\frac{B(z)}{A(z)} = G\,\frac{\sum_{i=0}^{L} b_i z^{-i}}{\sum_{i=0}^{M} a_i z^{-i}}\,, \qquad a_0 = b_0 = 1\,, \quad L \lesssim M\,, \tag{3-31}$$

whose autocorrelation function R(k) approximates the autocorrelation function R(k) of a given spectrum S(z)S(1/z), using only linear

operations, i.e.

$$\hat{R}(k) \simeq R(k), \tag{3-32}$$

where $R(k)$ is defined by (3-9) and $\hat{R}(k)$ is defined by :

$$\sum_{k=-\infty}^{+\infty} \hat{R}(k) z^{-k} \triangleq G^2 \frac{B(z)B(1/z)}{A(z)A(1/z)} = \frac{\sum_{i=-L}^{L} R_B(k) z^{-k}}{\sum_{i=-M}^{M} R_A(k) z^{-k}}, \tag{3-33}$$

$$r_2 < |z| < 1/r_2 ,$$

$$r_2 < 1.$$

Also, using (3-31), the autocorrelation functions $R_B(k)$ and $R_A(k)$ in (3-33) are obtained from

$$R_B(k) = \sum_{i=0}^{L-|k|} b_i \, b_{i+|k|}. \tag{3-34}$$

$$R_A(k) = \sum_{i=0}^{M-|k|} a_i \, a_{i+|k|}. \tag{3-35}$$

The above idea is realized in three steps: i. Estimation of the minimum phase polynomial $A(z)$, ii. Computation of the "Residual Autocorrelation Function" defined as:

$$R_D(k) \triangleq R(k) \circledast R_A(k) , \tag{3-36}$$

where " $\circledast$ " indicates linear convolution, iii. Estimation of the minimum phase All-Zero model B(z) and the gain G, Figure 3-2. At the end, the approximation error, and how Autocorrelation Prediction compares with Autocorrelation Partial Realization are addressed.

i. Estimation of the A(z). The minimum phase polynomial A(z) is obtained simply by finding the All-Pole model $G_A/A(z)$ whose autocorrelation function $G_A^2 R_{1/A}(k)$ exactly matches that of the given spectrum S(z)S(1/z) for the first M+1 time-lags, i.e.

$$R(k) = G_A^2 \, R_{1/A}(k), \qquad 0 \leq k \leq M, \tag{3-37}$$

where the autocorrelation function $G_A^2 R_{1/A}(k)$ is defined as:

$$G_A^2 \sum_{k=-\infty}^{+\infty} R_{1/A}(k) z^{-k} \triangleq \frac{G_A^2}{A(z)A(1/z)} = \frac{G_A^2}{\sum_{k=-M}^{M} R_A(k) z^{-k}}, \tag{3-38}$$

$$r_2 < |z| < 1/r_2,$$

$$r_2 < 1.$$

As is shown in section 2.3 and 2.4, the pole predictors $\{a_i\}$ are the solution to the system of linear equations (2-13) which also guarantees that A(z) is minimum phase. The gain squared G_A^2 is obtained from (2-14).

ii. Residual Autocorrelation Function. After finding the pole predictors $\{a_i\}$, the autocorrelation function $R_A(k)$ is computed according to the definition (3-35). Then, from (3-36), the Residual

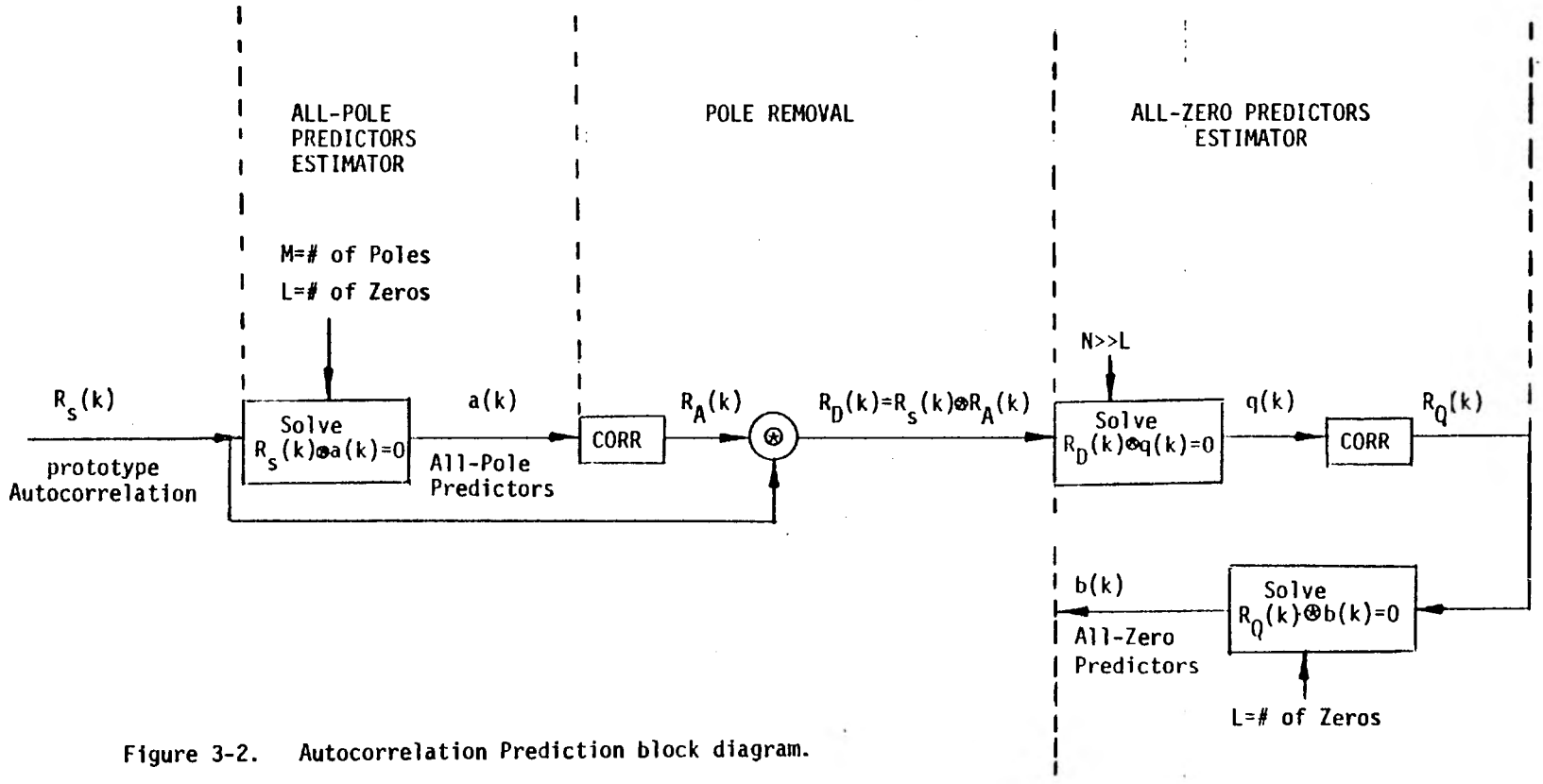


Figure 3-2. Autocorrelation Prediction block diagram.

Autocorrelation function $R_D(k)$ is obtained by simply convolving the finite length autocorrelation function $R_A(k)$ with the autocorrelation function R(k).

iii. Estimation of the B(z) and G. The estimation of the All-Zero model B(z) and the gain G is performed in steps a and b.

a. We find a high order All-Pole model of the form

$$\frac{G_Q}{Q(z)} = \frac{G_Q}{\sum_{i=0}^{N} q_i z^{-i}}, \qquad q_0 = 1, \quad N >> L, \tag{3-39}$$

whose autocorrelation function $G_Q^2 R_{1/Q}(k)$ exactly matches the Residual Autocorrelation function $R_D(k)$ for the first N+1 time-lags. That is,

$$G_Q^2 R_{1/Q}(k) = R_D(k), \qquad 0 \leq |k| \leq N, \tag{3-40}$$

where the autocorrelation function $G_Q^2 R_{1/Q}(k)$ is defined as

$$G_Q^2 \sum_{k=-\infty}^{+\infty} R_{1/Q}(k) \triangleq \frac{G_Q^2}{Q(z)Q(1/z)} = \frac{G_Q^2}{\sum_{k=-N}^{N} R_Q(k) z^{-k}}. \tag{3-41}$$

Similar to the estimation of the $\{a_i\}$, the coefficients $\{q_i\}$ are the solution to the system of linear equations (2-13) after replacing $\{a_i\}$ by $\{q_i\}$, R(k) for $0 \leq |k| \leq N$ by $R_D(k)$ for $0 \leq |k| \leq N$, and M by N, i.e.

$$\sum_{i=0}^{N} q_i R_D(k-i) = 0 , \qquad 1 \leq k \leq N. \tag{3-42}$$

Also, the gain squared G_Q^2 is obtained from (2-14) after the same replacement. That is,

$$G_Q^2 = \sum_{i=0}^{N} q_i R_D(i). \tag{3-43}$$

Using (3-39) in (3-41) gives the expression for the autocorrelation function $R_Q(k)$ as

$$R_Q(k) = \sum_{i=0}^{N-|k|} q_i q_{i+k} , \qquad 0 \leq |k| \leq N. \tag{3-44}$$

After computing the coefficients $\{q_i\}$ from (3-42), the equation (3-44) is used to calculate $R_Q(k)$.

b. Finally, we find the All-Pole model $G_B/B(z)$ whose autocorrelation function $G_B^2 R_{1/B}(k)$ exactly matches the autocorrelation function $R_Q(k)$ for the first L+1 time-lags, i.e.

$$G_B^2 R_{1/B}(k) = R_Q(k), \qquad 0 \leq |k| \leq L, \tag{3-45}$$

where the $G_B^2 R_{1/B}(k)$ is defined as

$$G_B^2 \sum_{k=-\infty}^{+\infty} R_{1/B}(k)z^{-k} \triangleq \frac{G_B^2}{B(z)B(1/z)} = \frac{G_B^2}{\sum_{k=-L}^{L} R_B(z)} . \qquad (3\text{-}46)$$

Again, similar to the estimation of the $\{a_i\}$, the zero predictors $\{b_i\}$ are the solution to the system of linear equations (2-13) after replacing $\{a_i\}$ by $\{b_i\}$, R(k) for $0 \leq |k| \leq M$ by $R_Q(k)$ for $0 \leq |k| \leq 1$, and M by L. That is,

$$\sum_{i=0}^{L} b_i R_Q(k-i) = 0, \qquad 1 \leq k \leq L. \qquad (3\text{-}47)$$

Similar to (2-13), the solution of the system of linear equations (3-47) guarantees that B(z) is minimum phase. Also, the gain squared G_B^2 is obtained from (2-14) after the same replacement.

$$G_B^2 = \sum_{i=0}^{L} b_i R_Q(i). \qquad (3\text{-}48)$$

It is shown in the next section that the gain G is equal to

$$G = \frac{G_Q}{G_B} = \frac{\sum_{i=0}^{N} q_i R_D(i)}{\sum_{i=0}^{L} b_i R_Q(i)} . \qquad (3\text{-}49)$$

The identification process of the Pole-Zero model G B(z)/A(z) is summarized in Figure 3-2. Figure 3-3 shows a speech short-time spectrum superimposed by the All-Pole and Pole-Zero model spectra for comparison. Note the advantage of the Pole-Zero model over the All-Pole model in matching the spectral envelope valley of the given speech short-time spectrum.

A variation for AP is obtained by computing the pole predictors from a non-symmetric, rather than a symmetric, toepltiz matrix. In this way, the zeros in the model are explicitly accounted for. For this variation, however, the stability is no longer gauranteed.

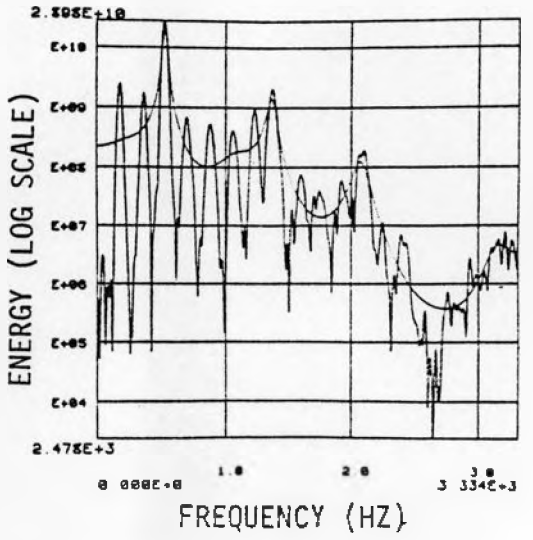
Approximation Error. To show how close the autocorrelation of the resulting Pole-Zero model G B(z)/A(z) approximates the autocorrelation function R(k), the approximation error is derived. From (3-38), (3-41) and (3-46), the following similar equations are obtained

$$G_A^2 \, R_{1/A}(k) \circledast R_A(k) = G_A^2 \delta(k), \tag{3-50a}$$

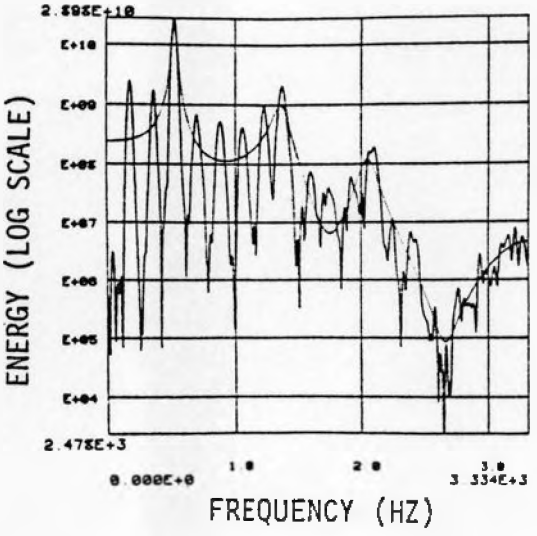
$$G_Q^2 \, R_{1/Q}(k) \circledast R_Q(k) = G_Q^2 \delta(k), \tag{3-50b}$$

$$G_B^2 \, R_{1/B}(k) \circledast R_B(k) = G_B^2 \delta(k). \tag{3-50c}$$

Using (3-40), we define the error autocorrelation function $\Delta R_D(k)$ to be



(a)



(b)

Figure 3-3. (a) Speech short-time spectrum superimposed with 14-Pole model spectrum (Linear Prediction).
(b) Speech short-time spectrum superimposed with 8-Pole and 6-Zero model spectrum (Autocorrelation Prediction).

$$\Delta R_D(k) = \begin{cases} 0 & 0 \leq |k| \leq N \\ R_D(k) - G_Q^2 R_{1/Q}(k) & |k| > N. \end{cases} \tag{3-51}$$

Thus we have

$$R_D(k) = G_Q^2 R_{1/Q}(k) + \Delta R_D(k). \tag{3-52}$$

Using (3-50b) in (3-52) gives

$$R_Q(k) \circledast R_D(k) = G_Q^2 \delta(k) + \Delta R_D(k) \circledast R_Q(k). \tag{3-53}$$

Similarly, using (3-45), we define the error autocorrelation function $\Delta R_Q(k)$ to be

$$\Delta R_Q(k) = \begin{cases} 0 & 0 \leq k \leq L, \\ R_Q(k) - G_B^2 R_{1/B}(k), & |k| > L. \end{cases} \tag{3-54}$$

Thus we have

$$R_Q(k) = G_B^2 R_{1/B}(k) + \Delta R_Q(k). \tag{3-55}$$

Using (3-55) and (3-36) in (3-53) and some rearranging of terms results in

$$G_B^2 R_{1/B}(k) \circledast R_A(k) \circledast R(k) = G_Q^2 \delta(k) + [\Delta R_D(k) \circledast R_Q(k) - \Delta R_Q(k) \circledast R_D(k)]. \quad (3\text{-}56)$$

Finally, using (3-50a), (3-50c) in (3-56) gives

$$R(k) = \frac{G_Q^2}{G_B^2} R_B(k) \circledast R_{1/A}(k) + \Delta R(k), \quad (3\text{-}57)$$

where the approximation error is

$$\Delta R(k) \triangleq \frac{1}{G_B^2} R_B(k) \circledast R_{1/A}(k) \circledast [\Delta R_D(k) \circledast R_Q(k) - \Delta R_Q(k) \circledast R_D(k)]. \quad (3\text{-}58)$$

For proper selection, discussed below, of the orders M, L, and N the function in the square brackets in (3-56) or (3-58) has values relatively close to zero for the short time-lags. Therefore,

$$R(k) \simeq \frac{G_Q^2}{G_B^2} R_B(k) \circledast R_{1/A}(k). \quad (3\text{-}59)$$

Taking Z-transform of (3-59) and using (3-9)and (3-38) results in

$$\sum_{k=-\infty}^{+\infty} R(k) z^{-k} = S(z)S(1/z) \simeq \left(\frac{G_A}{G_B}\right)^2 \frac{B(z)B(1/z)}{A(z)A1/z)} = \sum_{k=-\infty}^{+\infty} \hat{R}(k) z^{-k}. \quad (3\text{-}60)$$

Hence, the autocorrelation function of the Pole-Zero model approximates the R(k) at short time-lags and the approximation error is given by (3-58). Note (3-56) shows that the autocorrelation

function of the inverse filter output, for the obtained Pole-Zero model, differs from the impulse response $G\delta(k)$ by $1/G_B^2$ times the "small" function in the square brackets

Comparison of APR and AP. There is a trade-off between the Autocorrelation (AP) Prediction and Autocorrelation . Partial Realization(APR) . In AP all the operations are linear and the stability is guaranteed, but the short time-lag autocorrelations of the given spectrum are approximately matched. In contrast, APR has partially non-linear operations and the stability is not guaranteed, but the short time-lag autocorrelations are exactly matched. The APR is also theoretically more appealing. AP and APR have, however, some common properties. Once the autocorrelation of the given spectrum is known, no Fourier transformation is required to estimate the model parameters,using either technique. More important, both are well suited for matching the spectral envelope of a given spectrum having fine structure, such as speech short-time spectrum. This is possible because in both techniques the model parameters are computed such that the short time-lag autocorrelations, representing the gross structure of the given spectrum, are either exactly matched or closely approximated.

3.5 Frequency Interpretation of AP

The frequency domain interpretation of the steps taken in Autocorrelation Prediction, described in section 3.4, and selection of the orders M, L and N are addressed here, Figure 3-4.

The given spectrum $S(z)S(1/z)$ can be thought of as being

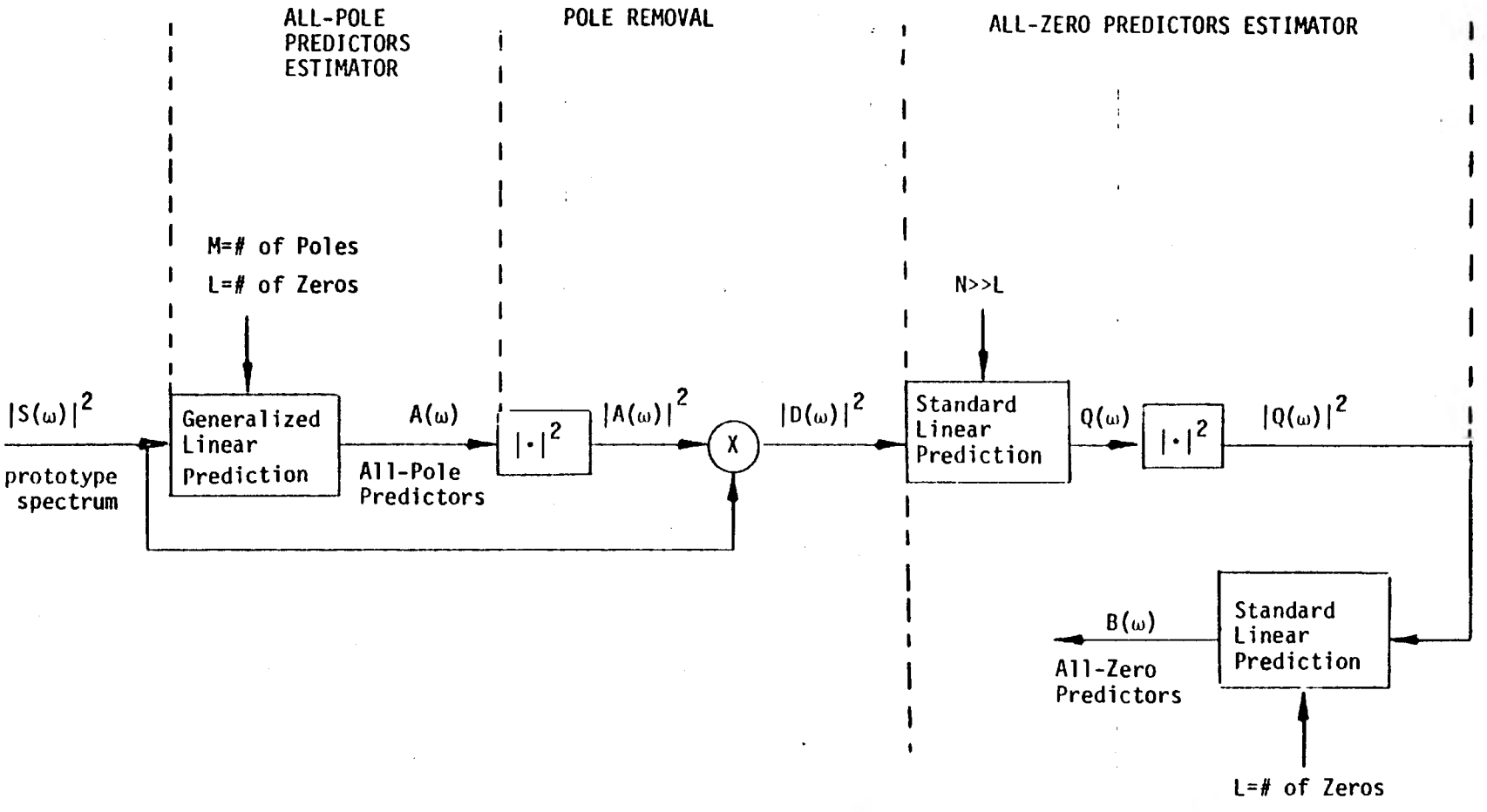


Figure 3-4. Frequency domain equivalent of Autocorrelation Prediction block diagram.

composed of spectral poles and zeros represented by the spectrum peaks and deep valleys. As disscussed in sections 2.3 asd 2.5, fitting the All-Pole spectrum $G_A^2/A(z)A(1/z)$ to the given spectrum $S(z)S(1/z)$, matches the peaks rather the valleys of the spectral envelope. By selecting the order of the All-Pole model equal to the parsimonious order M_P, those peaks having higher amplitudes are matched, while the deep valleys are left unmatched. The residual spectrum $D(z)D(1/z)$, which is the given spectrum after removing the estimated poles,contains primarily the spectral zeros represented by the deep valleys in the residual spectrum. To fit the All-Zero spectrum $B(z)B(1/z)$ to the residual spectrum $D(z)D(1/z)$, the All-Pole spectrum $G_B^2/B(z)B(1/z)$ is fitted to the $Q(z)Q(1/z)$; an approximation for the reciprocal of the residual spectrum envelope. To obtain $Q(z)Q(1/z)$, the high order All-Pole spectrum $G_Q^2/Q(z)Q(1/z)$ is matched to the residual spectrum $D(z)D(1/z)$. The high order N is required because each of the spectral zeros of the residual spectrum is approximated with a large number of poles in the model $G_Q/Q(z)$. The lower bound for the order N, therefore, depends on the number of spectral zeros in the residual spectrum or, equivalently, in the given spectrum $S(z)S(1/z)$, and how close these spectral zeros are to the unit circle. Similar to the order M, the order L may also be equal to the parsimonious order L_P of the All-Pole spectrum match to the spectrum $Q(z)Q(1/z)$.

To reduce the computation of the parameter estimation when the given spectral envelope has more deep valleys than peaks, the Autocorrelation Prediction is applied to the reciprocal of the given

spectrum and then reciprocal of the resulting Pole-Zero model is used. In chapter 4 this approach is used in Pole-Zero modeling of Wiener filter spectrum.

3.6 Realization of Pole-Zero Model

The Pole-Zero model may be realized using any of a variety of methods [16,17]. We propose a method which is more suitable for realization of the Pole-Zero model obtained by AP or APR.

Cascade of Two Lattice Filters. The Pole-Zero model G B(z)/A(z) given by (3-31) can be realized as the cascade of the All-Zero filter B(z) and the All-Pole filter 1/A(z), having the overall gain G. Either of these filter is implemented by a lattice filter [19] which is identified with the so-called parcor parameters, Figure 3-5.

There is a unique set of parcor parameters $\{K_i\}$ for the set of predictor coefficients $\{a_i\}$ and vice versa [Appendix A]. Similarly, there is unique set of parcor parameters $\{K'_i\}$ for the zero predictors $\{b_i\}$ and vice versa. Both sets of parcor parameters $\{K_i\}$ and $\{K'_i\}$ are obtained as a by-product using AP, but both should be computed [Appendix A] when the Pole-Zero model is identified using APR.

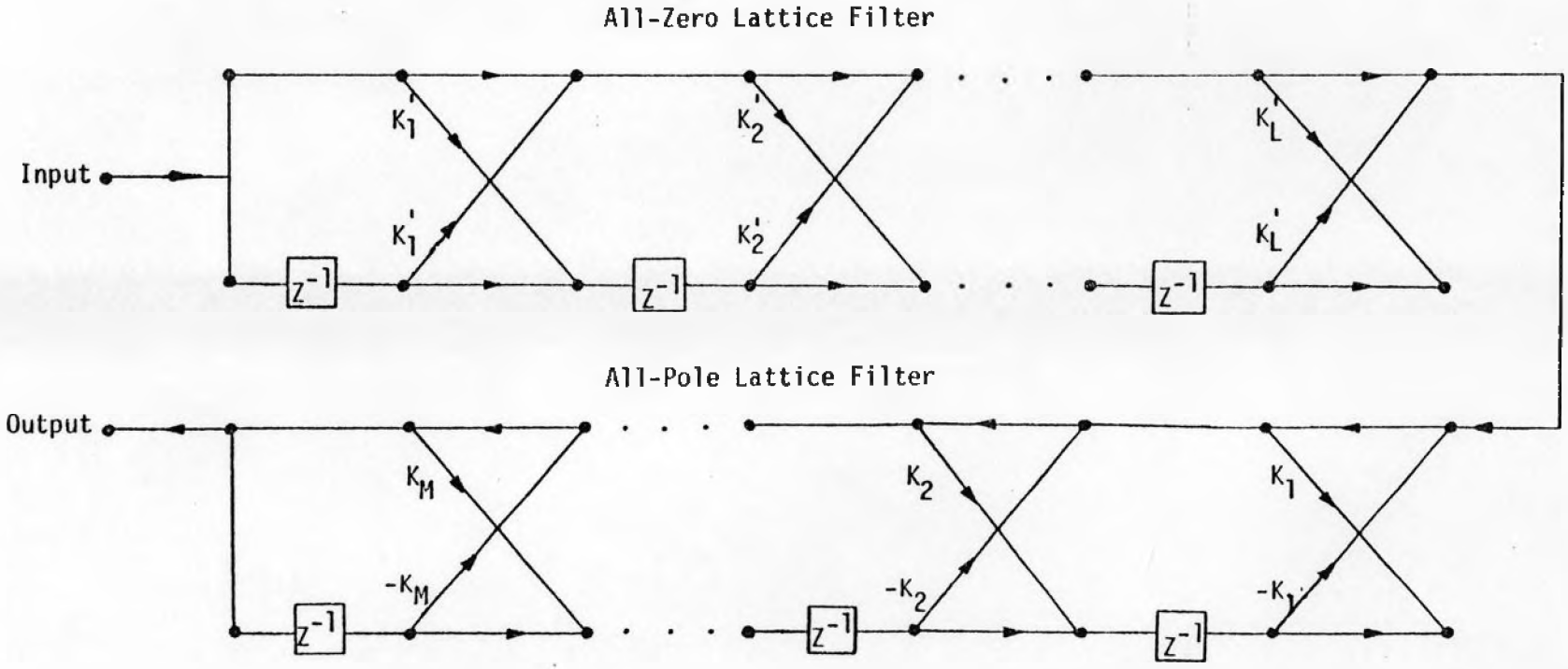


Figure 3-5. Cascade of two lattice filters as a realization of Pole-Zero model.

CHAPTER 4

DYNAMIC POLE-ZERO FILTERING

4.1 Introduction

Quite often speech is degraded with additive independent background noise, such as interference of helicopter noise with pilot's speech and surface noise with singing in the play back of old recordings. To enhance the speech, an input controlled time-varying filter is devised that suppresses the noise and passes through the speech.

Speech and singing may be assumed stationary during short intervals, either T =30 msec for speech or T =100 msec for singing. The noise is assumed stationary during a longer period of interest. During this period of interest, therefore, the degraded speech may be thought of as a concatenation of stationary processes, short speech intervals degraded with additive stationary noise. The noise spectrum generally has a common range of frequencies with the speech short interval spectrum. To suppress the noise, therefore, a Wiener filter can be used, Figure 4-1. The Wiener filter smoothed spectrum $|W(kT,\omega)|^2$ is estimated for each short interval T. The spectrum of a rational filter $H(kT,\omega)$ is then matched to $1/|W(kT,\omega)|^2$, using Autocorrelation Prediction described in sections 3.4 and 3.5.

Finally, the rational filter $1/H(kT,\omega)$ is used to filter the corresponding interval of the degraded speech, Figure 4-2.

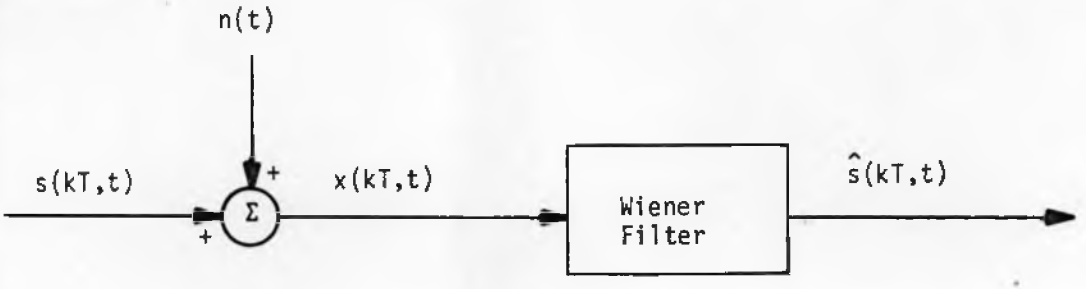


Figure 4-1. Short interval Wiener filtering.

4.2 Estimation of Wiener Filter Smoothed Spectrum

The Wiener filter smoothed spectrum $|W(kT,\omega)|^2$ for each interval T is estimated from the following equation *

$$W(kT,\omega) = 1 - \frac{\hat{\Phi}_N(\omega)}{\hat{\Phi}_{S+N}(kT,\omega)} \qquad (4\text{-}1)$$

where $\hat{\Phi}_N(\omega)$ is an estimate for the smoothed spectrum of the stationary noise and $\hat{\Phi}_{S+N}(kT,\omega)$ is an estimate for the smoothed spectrum of the degraded speech during the k-th interval.

Since we are interested in filtering the gross structure of degraded speech and leaving alone the fine structure, mainly caused by the speech excitation function, the finite Linear Prediction (LP) spectrum is used rather than the standard Fourier Spectrum. Thus, $\hat{\Phi}_{S+N}(kT,\omega)$ is obtained by averaging the short-time LP spectrum of the degraded speech during the k-th interval. To estimate the noise

---

* The equation for Wiener filter spectrum is derived in section 5.5.

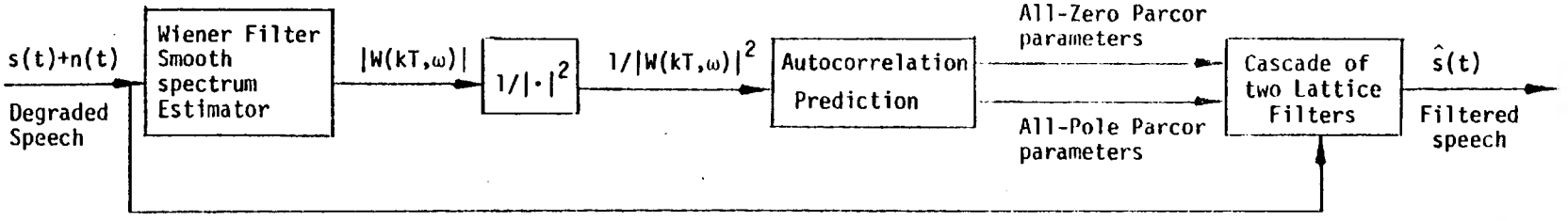


Figure 4-2. Dynamic Pole-Zero Filtering block diagram.

spectrum, speechless portions of the degraded speech are used. Thus, $\Phi_N(\omega)$ is obtained by averaging the short-time LP spectrum during these speechless portions.

4.3 Pole-Zero Modeling Of Wiener Filter Smoothed Spectrum

Equation (4.1) shows that the Wiener filter smoothed spectrum $|W(kT,\omega)|^2$ has flat pass bands for those frequencies where the speech dominates the noise, and has deep stop bands at those frequencies where the noise dominates the speech, Figure 4-3. In the case where speech dominates the noise at most of the frequencies, the above remark suggests modeling $|W(kT,\omega)|^2$ at the stop bands first. This can be accomplished by matching the spectrum of a rational filter $H(kT,\omega)$ to $1/|W(kT,\omega)|^2$, using Autocorrelation Prediction described in Section 3.4. Then, the rational filter $1/H(kT,\omega)$ is used to filter the k-th interval of the degraded speech, Figure 4-2. In this way, the parameters of the model $H(kT,\omega)$ are reduced in number considerably compared with an All-Pole match to $|W(kT,\omega)|^2$. Figure 4-4 compares a Pole-Zero model and an All-Pole model spectral match to the same Wiener filter spectrum in Figure 4-3. Note that with equal number of parameters, the Pole-Zero model matches the deep valleys much more accurately than the All-Pole model.

Since $H(kT,\omega)$ is minimum phase, the stability of the rational filter $1/H(kT,\omega)$ is also guaranteed.

4.4 Implementation And Results

The rational filter $1/H(kT,\omega)$ is identified by two sets of parcor parameters obtained from autocorrelation prediction. The

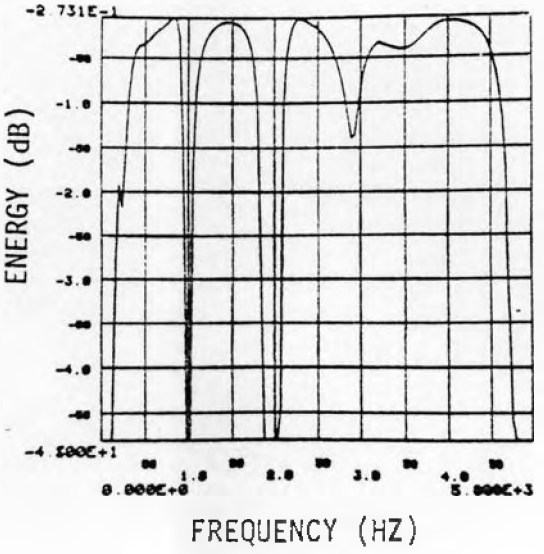
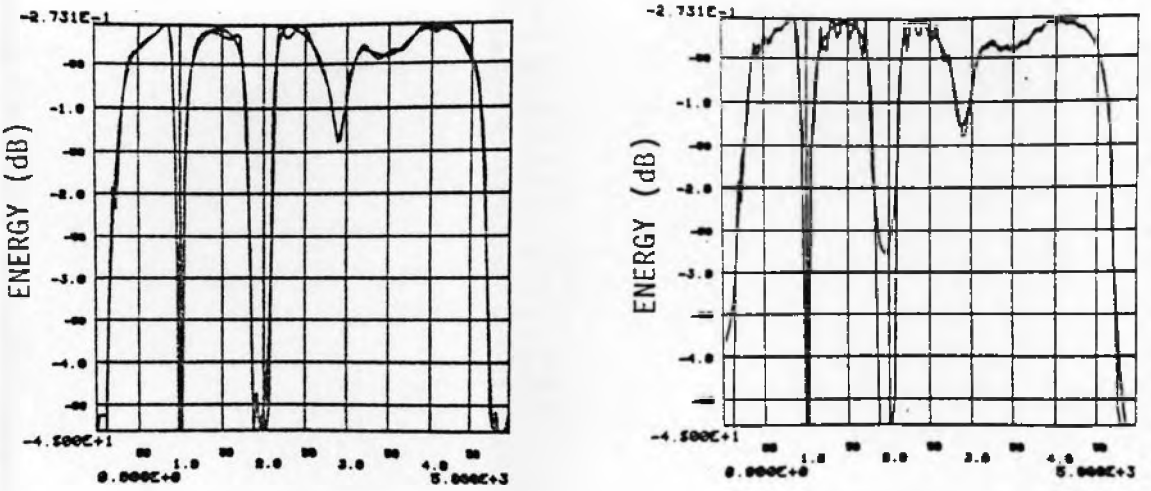


Figure 4-3. Wiener filter smoothed spectrum for short interval.



(a) (b)

Figure 4-4. (a) Wiener smoothed spectrum superimposed with 30-Pole and 30-Zero model spectrum (Autocorrelation Prediction).

(b) Wiener smoothed spectrum superimposed with 60-Pole model spectrum (Linear Prediction).

$1/H(kT,\omega)$ is realized as the cascade of two lattice filters described in Section 3.6. To improve the filtering process, both sets of the parcor parameters are interpolated for each sample of the input to the filter.

In theory, the value of $W(kT,\omega)$ given by (4.1) is non-negative for all frequencies. In practice, however, $W(kT,\omega)$ occasionally becomes negative at some frequencies. This is so because of the following: $\hat{\Phi}_{S+N}(kT,\omega)$ is an estimate for the speech smoothed spectrum plus the noise smoothed spectrum. On the other hand, the interval T (30 msec.) during which the speech is stationary is <u>not</u> long enough to give a close estimate of the noise smoothed spectrum. This makes the $\hat{\Phi}_{S+N}(kT,\omega)$ a <u>rough</u> estimate and consequently $W(kT,\omega)$, given by equation (4-1), becomes negative at some frequencies. To correct for the above problem, using the function depicted in Figure 4-5, $W(kT,\omega)$ is tailored to positive values at those frequencies where it is negative.

The above Dynamic Pole-Zero Filtering process was applied to speech degraded with stationary additive colored noise. Spectrograms of corresponding portions of the degraded speech, filtered speech and the original clean speech are shown in Figure 4-6 for comparison. The process was also used to suppress helicopter noise in a recording of pilot's speech degraded by helicopter noise. Spectrograms of corresponding portions of degraded pilot's speech and the filtered speech are shown in Figure 4-7. Figure 4-7a shows the spectrogram of the helicopter noise from a portion of the recording where the pilot was silent.

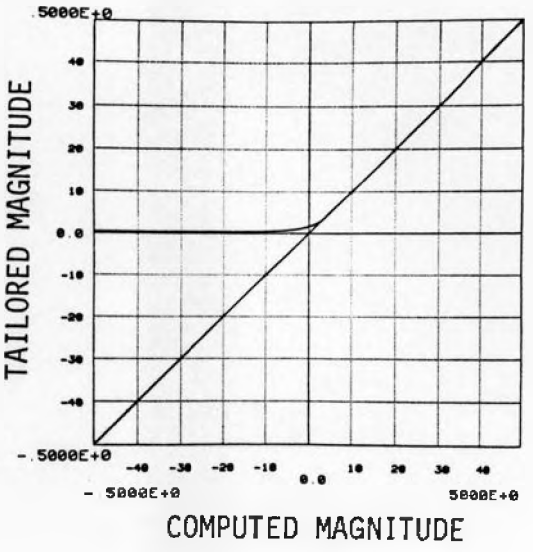
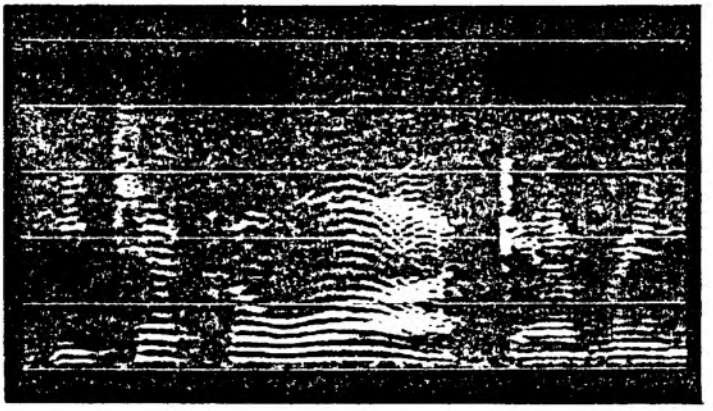


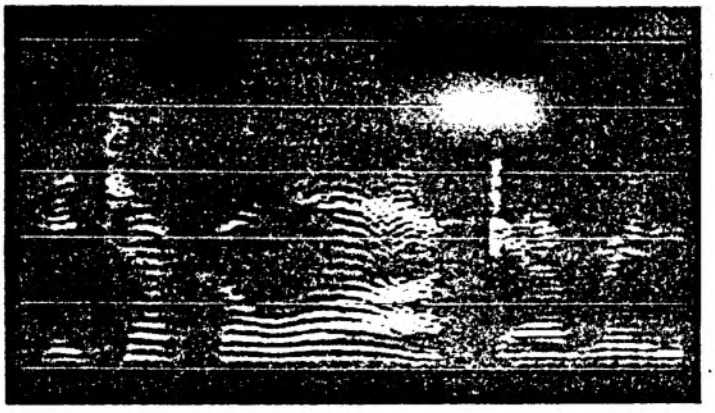
Figure 4-5. Tailoring function (after T. Petersen).

The Linear Predition (LP) spectrum of the helicopter noise is also shown in Figure 4-8. Note the harmonic structure in the helicopter noise depicted in both Figure 4-7a and Figure 4-8. Finally, corresponding portions of time domain pilot's speech with background helicopter noise and the filtered speech are shown in Figure 4-9.

(a)



(b)



(c)

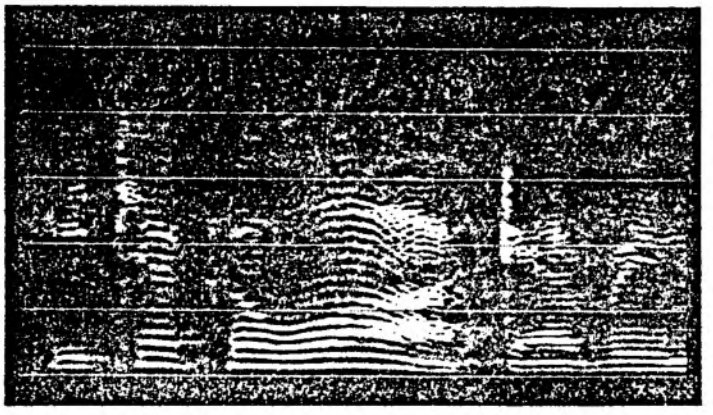


Figure 4-6. Corresponding spectrograms of :

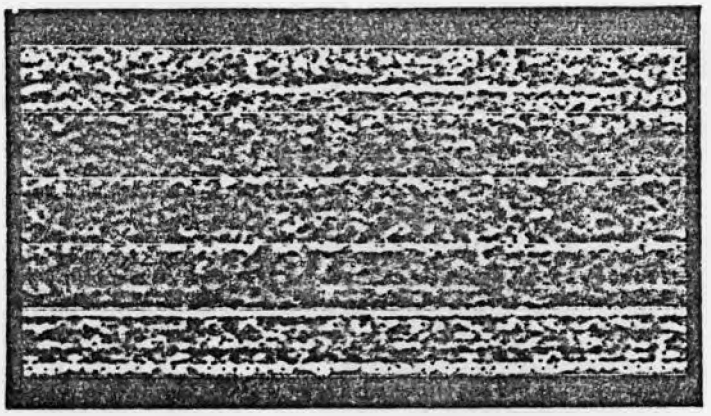
(a) Input degraded speech; clean speech plus stationary colored noise.

(b) Filtered speech; output of the Dynamic Pole-Zero Filtering process.

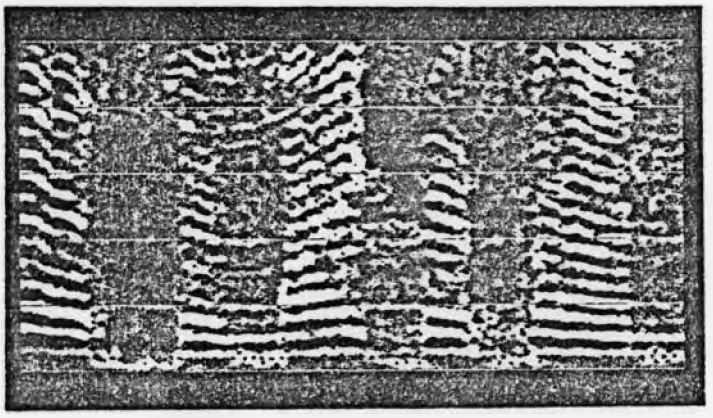
(c) Clean speech.

All three spectrograms have been scaled 6 dB/oct above 400 Hz.

(a)



(b)



(c)

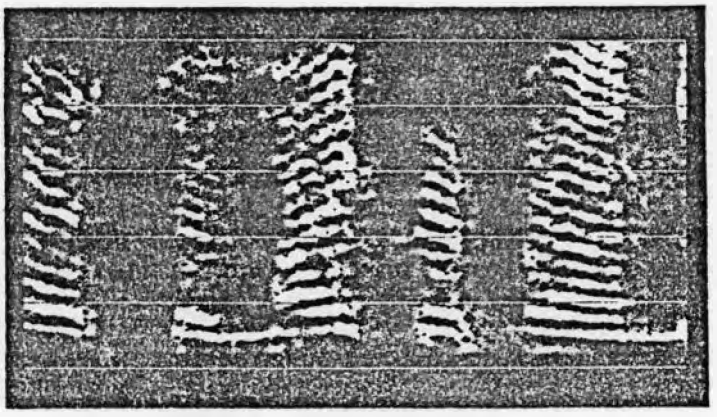


Figure 4-7. Spectrograms of :

(a) Helicopter noise.

(b) Pilot's speech degraded with background helicopter noise.

(c) Filtered pilot's speech; output of the Dynamic Pole-Zero Filtering process.

All three spectrograms have been scaled 6 dB/oct above 400 Hz.

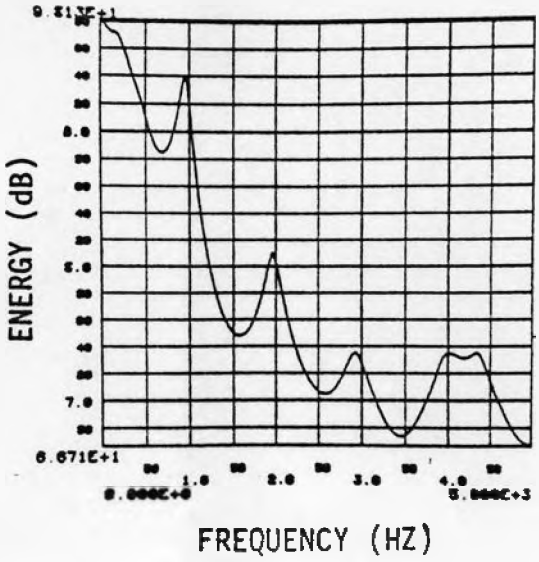
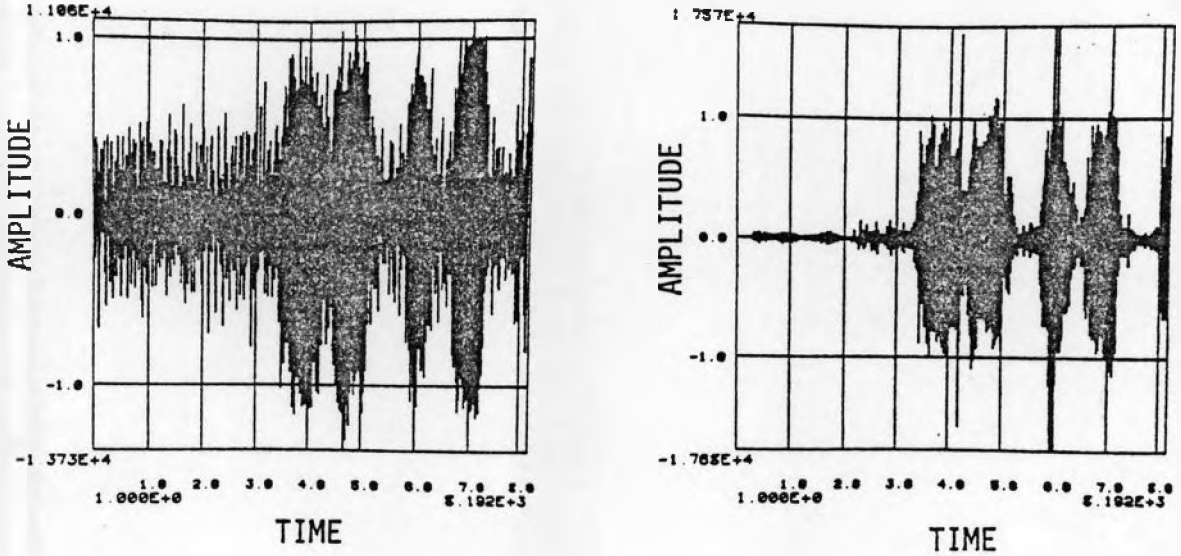


Figure 4-8. Linear Prediction (LP) spectrum of the helicopter noise.



(a) (b)

Figure 4-9. Corresponding portions of :

(a) Input pilot's speech degraded with background helicopter noise.

(b) Filtered pilot's speech; output of the Dynamic Pole-Zero Filtering process.

CHAPTER 5

POLE-ZERO VOCODER (PZV)

5.1 Introduction

The digital speech production model capitalizing on apriori information on the structure of the speech mechanism and the speech waveform is described. The speech production model is further approximated with a limited number of parameters, including zero as well as pole parameters. Using this parametric representation of the speech production model, a so-called Pole-Zero Vocoder (PZV) is devised for analysis and synthesis of clean speech. To code speech degraded with stationary additive colored noise, the PZV is further modified to account for the noise. The PZV, simulated on a computer, was used successfully in analysis and synthesis of clean speech Finally, some pilot experiments revealing the potential of the modified PZV in coding speech degraded with stationary additive colored noise is presented.

5.2 Speech Production Model

For voiced sounds, the time varying filter model in Figure 5-1 represents the effect of the glottis waveform, the vocal tract, the acoustic coupling of the nasal tract and the radiation. The excitation function is a train of unit samples with the same frequency as the pitch. In contrast, for unvoiced sounds the time varying filter model represents the effect of the vocal tract and the

radiation, and the excitation function becomes zero mean white noise with the variance $\sigma^2=1$.

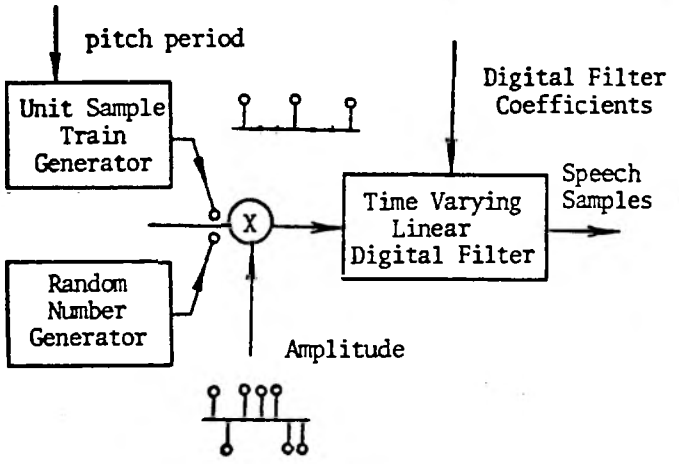


Figure 5-1. Digital speech production model.

The time-varying filter model is considered to be a time-invariant filter during short periods of time (10-30 msec.). This is a reasonable assumption, because the dynamics of the articulatary configuration are slow due to the inertia of muscle controlled jaw, tongue, and lip movements. The finite-time glottis waveform, the acoustic coupling of the nasal tract and the radiation cause the transfer function of the filter model to have zeros as well the usual poles, [14].

The above arguments imply that representing the filter model transfer function by a finite dimension pole-zero model [6] rather than a finite dimension all-pole model [5] may improve the quality of the synthesized speech and/or reduce the representation parameters.

5.3 Autocorrelation Function of the Filter Model

We show how the short-time autocorrelation function of the speech gives an estimate for the autocorrelation function of the filter model in Figure 5-1. The short-time autocorrelation function of the speech is defined as the autocorrelation function of windowed speech using a smoothed window of proper length (20-30 msec.)

Recalling the speech production model in Figure 5-1, we denote the excitation function, the train of pulses or the white noise, by p(n) and the scaled impulse response of the filter model by v(n). The speech s(n) is, therefore, expressed as:

$$s(n) = p(n) \circledast v(n). \tag{5-1}$$

To obtain a stationary segment of speech, the speech s(n) is weighted by some smoothed window w(n) of proper length (20-30 msec.), i.e.

$$s_w(n) \triangleq s(n)\, w(n) = [p(n) \circledast v(n)]\, w(n) \tag{5-2}$$

Assuming the window w(n) is smooth during the effective duration of the impulse response v(n), the equation (5-2) can be approximated as:

$$s_w(n) \simeq [p(n)\, w(n)] \circledast v(n), \tag{5-3a}$$

or

$$s_w(n) \simeq p_w(n) \circledast v(n), \tag{5-3b}$$

where $P_w(n)$ is the weighted excitation function defined as:

$$p_w(n) \triangleq p(n)\, w(n). \tag{5-4}$$

Short-time Autocorrelation Function of Speech Using (5-3b), the speech short-time spectrum $S_w(z)S_w(\frac{1}{z})$ is approximated by:

$$S_w(z)S_w(1/z) \simeq V(z)V(1/z)\; P_w(z)P_w(1/z). \tag{5-5}$$

The equation (5-5) shows that the speech short-time spectrum approximates the filter model spectrum multiplied by the spectrum of the windowed excitation function. From (5-5), the speech short-time autocorrelation function R(k) is approximately equal to

$$R(k) \simeq R_v(k) \circledast R_{P_w}(k), \tag{5-6}$$

where $R_v(k)$ and $R_{P_w}(k)$ are autocorrelation functions of the filter model and the windowed excitation function, respectively.

Now we find the $R_{P_w}(k)$ for both voiced and unvoiced speech. In the case of voiced speech, the excitation function p(n) is a train of pulses with the same period T as the pitch period. Thus using (5-4), the weighted excitation function $p_w(n)$ takes the form of a weighted train of pulses, i.e.

$$p_w(n) = w(n) \sum_{\ell=-\infty}^{+\infty} \delta(n-\ell T) = \begin{cases} w(n) & n = 0, \pm T, \ldots \\ 0 & \text{otherwise.} \end{cases} \tag{5-7}$$

The autocorrelation function of $P_w(n)$, therefore, becomes

$$R_{P_w}(k) = \begin{cases} \sum_{\ell=-\infty}^{+\infty} w(\ell T)w(\ell T+k) & k = 0, \pm T, \ldots \\ 0 & \text{otherwise.} \end{cases} \qquad (5\text{-}8)$$

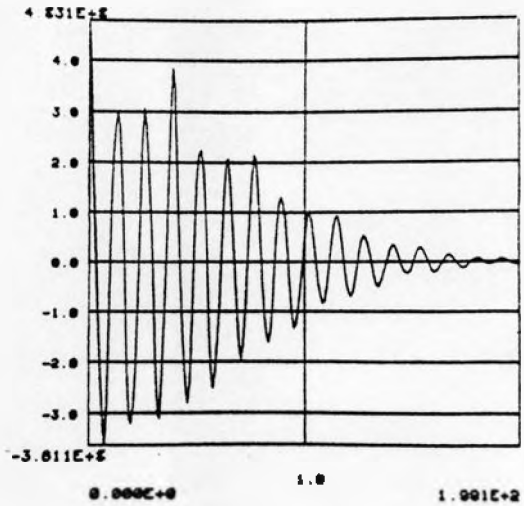
Thus for voiced speech, the short-time autocorrelation function R(k) is approximately equal to the filter model autocorrelation function $R_v(k)$ convolved with the symmetric decaying train of pulses $R_{P_w}(k)$, having the same period T as the pitch period, see Figure 5-2a. Hence, the R(k) is an estimate of $R_v(k)$ for short time-lags. For unvoiced speech, the excitation function p(n) is white noise. Assuming the smoothed window w(n) is long enough, then the autocorrelation function of $P_w(n)$ approximates the autocorrelation function of white noise namely, $\delta(n)$. Thus we have

$$R_{P_w}(k) = \delta(k). \qquad (5\text{-}9)$$

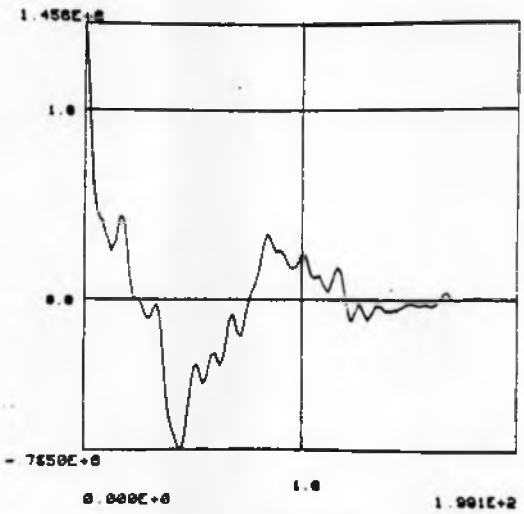
As a result, using (5-6) the short-time autocorrelation function for unvoiced speech is an estimate for the filter model autocorrelation function $R_v(k)$, see Figure 5-2b.

5.4 Pole-Zero Analysis-Synthesis of Speech

Each short segment of speech can be represented by a Voiced/Unvoiced decision, the pitch period, if the decision is voiced, and the filter model transfer function. To be able to represent the short segments of speech with a limited number of parameters, the filter model in Figure 5-1 is represented by a



(a)



(b)

Figure 5-2. (a) Short-time autocorrelation function of voiced sound [O].
(b) Short-time autocorrelation function of unvoiced sound [P].

parametric model. Recalling the discussion in section 5.2, the parametric model is chosen to be the Pole-Zero model H(z) given by equation (3-3). Furthermore, since the numbers of the spectral poles and the spectral zeros in the filter model spectrum are different for different segments of speech, then full advantage of the Pole-Zero modeling is gained by allowing the orders M and L to be dynamic.

The overall block diagram of the so-called Pole-Zero Vocoder (PZV) is shown in Figure 5-3. The Voiced/Unvoiced decision and the pitch period are extracted using any of a variety of existing methods [28]. The parameters of the Pole-Zero model H(z) are obtained by applying Autocorrelation Prediction, Figure 3-2, to the short-time autocorrelation function of the speech. Since the speech short-time autocorrelation function is an estimate for the filter model autocorrelation function, then the spectrum of the resulting Pole-Zero model approximates that of the filter model. In applying Autocorrelation Prediction to the speech short-time autocorrelation function, the parsimoneous (most economical) orders M_p and L_p for each segment may be obtained using the methods described in Section 2-5. Finally, to synthesize speech using the PZV depicted in Figure 5-3, we update the Voiced/Unvoiced decision, the pitch period T and the Pole-Zero model parameters for each segment of speech.

5.5 Pole-Zero Analysis-Synthesis of Speech In Presence of Additive Stationary Noise

The idea is to find the optimal estimate of the speech spectrum for each short interval. The autocorrelation function of this optimal estimate is then matched with that of a Pole-Zero model using

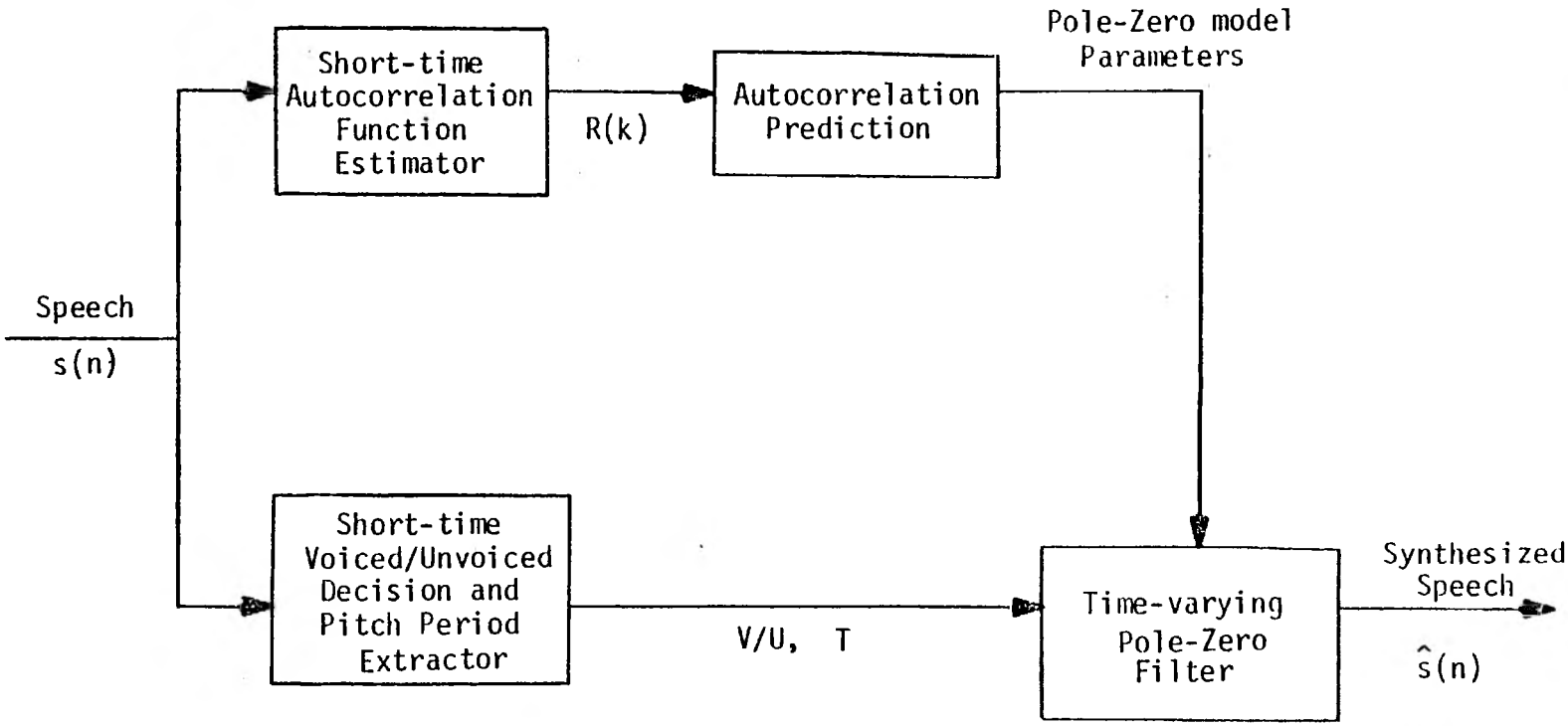


Figure 5-3. Pole-Zero Vocoder (PZV) overall block diagram.

Autocorrelation Prediction described in section 3.4. We make the assumption that the background noise is stationary during a long period of interest, while the speech is stationary only in short intervals (30 msec).

The optimum linear filter, namely the Wiener filter, is designed for each short interval of noisy speech, Figure 5-4.

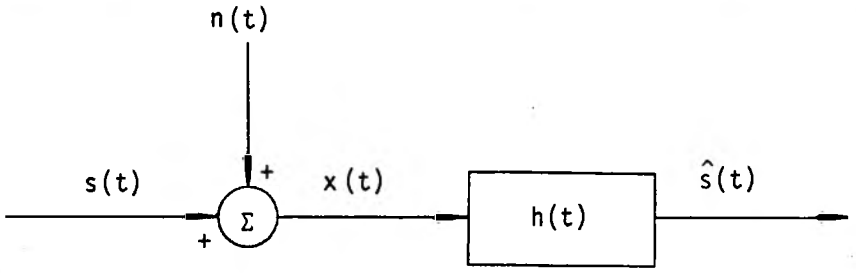


Figure 5-4. Wiener filtering

The transfer function of the Wiener filter is

$$H(\omega) = \frac{\Phi_{SX}(\omega)}{\Phi_{XX}(\omega)} . \qquad (5\text{-}10)$$

Assuming the noise and speech are uncorrelated, then (5-10) takes the form

$$H(\omega) = \frac{\Phi_{SS}(\omega)}{\Phi_{SS}(\omega) + \Phi_{NN}(\omega)} = \frac{1}{1 + \dfrac{\Phi_{NN}(\omega)}{\Phi_{SS}(\omega)}}, \qquad (5\text{-}11)$$

or

$$H(\omega) = \frac{\Phi_{XX}(\omega) - \Phi_{NN}(\omega)}{\Phi_{XX}(\omega)} = 1 - \frac{\Phi_{NN}(\omega)}{\Phi_{XX}(\omega)} . \qquad (5\text{-}12)$$

The noise spectrum $\Phi_{NN}(\omega)$ can be estimated using speechless portions from the noisy speech record. From Figure 5-4, the optimal estimate of the short interval spectrum is

$$\Phi_{\hat{S}\hat{S}}(\omega) = \Phi_{XX}(\omega)\left|H(\omega)\right|^2. \qquad (5\text{-}13)$$

Using (5-11) or (5-12) in (5-13) gives

$$\Phi_{\hat{S}\hat{S}}(\omega) = \frac{\Phi_{SS}(\omega)}{1 + \frac{\Phi_{NN}(\omega)}{\Phi_{SS}(\omega)}} , \qquad (5\text{-}14)$$

or

$$\Phi_{\hat{S}\hat{S}}(\omega) = \Phi_{XX}(\omega)\left[1 - \frac{\Phi_{NN}(\omega)}{\Phi_{XX}(\omega)}\right]^2. \qquad (5\text{-}15)$$

Equation (5-14) shows how the ratio of the optimal estimate $\Phi_{\hat{S}\hat{S}}(\omega)$ to the speech spectrum $\Phi_{SS}(\omega)$ is related to the power ratio $\Phi_{NN}(\omega)/\Phi_{SS}(\omega)$ of the background noise to the speech at different

frequencies. The optimal estimate, however, is computed from (5-15), where $\Phi_{XX}(\omega)$ is the noisy speech spectrum for the short interval (30 msec) and $\Phi_{NN}(\omega)$ is the noise spectrum for the period of interest.

An estimate for $\Phi_{XX}(\omega)$ is obtained by averaging spectra or LP spectra for short overlapping segments of noisy speech. An estimate for $\Phi_{NN}(\omega)$ is computed by averaging the short-time spectra or LP spectra during speechless segments of noisy speech. Finally, the Inverse Fourier Transform of the optimal estimate $\Phi_{\hat{S}\hat{S}}(\omega)$ gives the autocorrelation function (k), i.e.

$$\hat{R}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{\hat{S}\hat{S}}(\omega) e^{-jk\omega} d\omega. \qquad (5\text{-}16)$$

The overall block diagram for the analysis and synthesis of speech in the presence of additive stationary noise is depicted in Figure 5-5. Similar to clean speech analysis, applying Autocorrelation Prediction to the R(k) gives the parameters of the Pole-Zero model, the autocorrelation function of which approximates the R(k). The Voiced/ Unvoiced decision and the pitch period extraction for the short segments of speech, becomes more complicated in the presence of noise [44]. The synthesis remains the same as the one in PZV.

5.6 Implementation and Results

The Pole-Zero Vocoder (PZV) based on Autocorrelation Prediction was simulated on a PDP-10 computer. The parameters of a fixed order

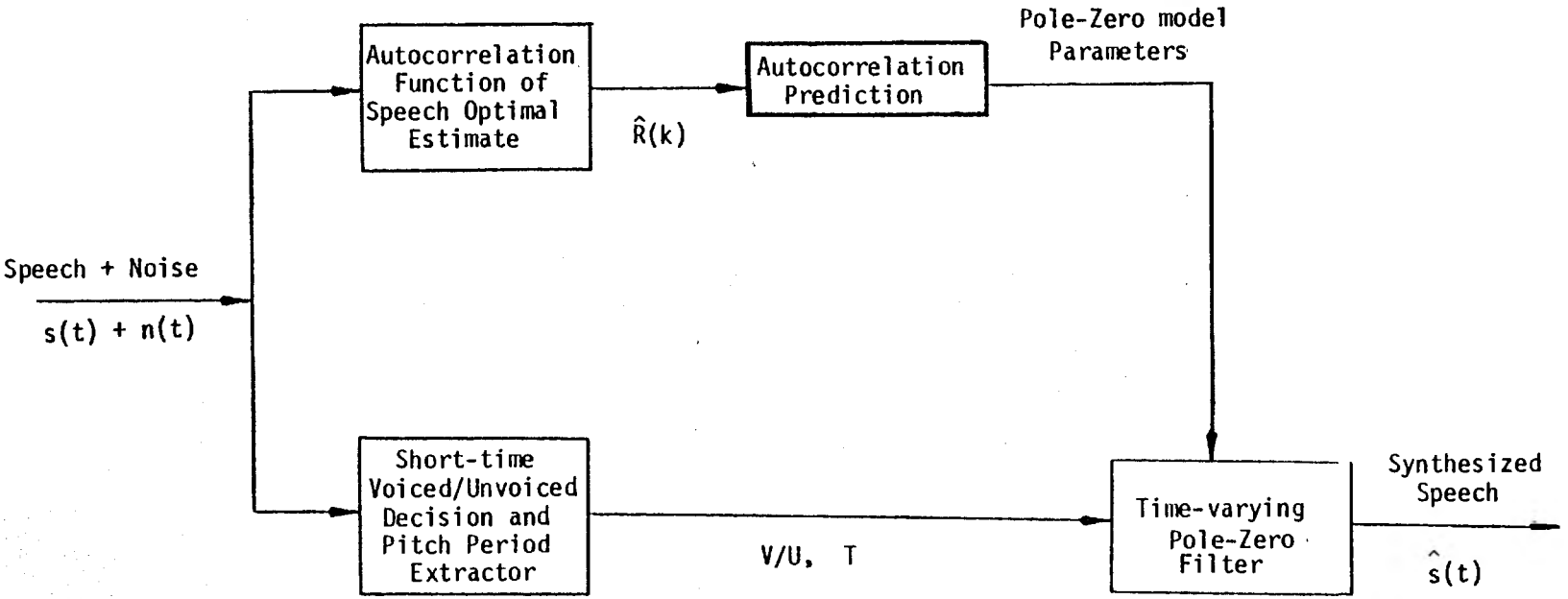


Figure 5-5. Pole-Zero analysis-synthesis of noisy speech overall block diagram.

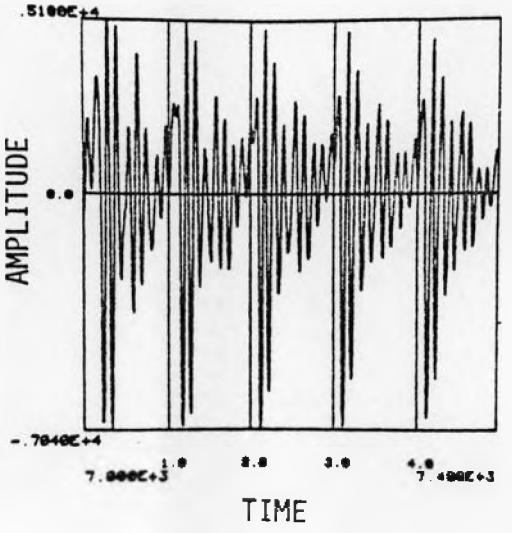
Pole-Zero model are estimated by applying Autocorrelation Prediction to the speech short-time autocorrelation function. The Pole-Zero model is realized by a lattice filter augmented with tap parameters [17]. The gain G for the lattice filter is computed such that the corresponding synthesized speech and the natural speech have equal energy [5,8].

The Voiced/Unvoiced decision and the pitch period is extracted using an implementation of the SIFT algorithm [28] by Boll [9]. The analysis-synthesis is quasi-pitch synchronous [9]. The augmented lattice filter parameters are linearly interpolated between two successive analysis segments.

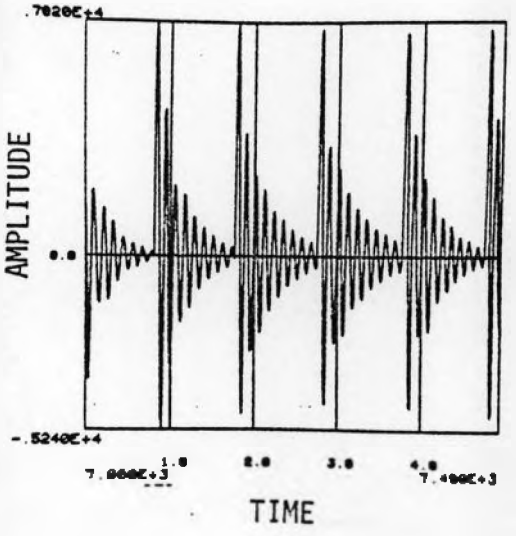
This vocoder was used for analysis and synthesis of passages of natural speech. For proper selection of the fixed orders L and M, informal hearing tests show some improvement of the synthesized speech generated by the Pole-Zero Vocoder over that generated by the All-Pole vocoder having comparable number of parameters. The improvement is more noticeable when the corresponding natural speech has more nasalized sounds. The Pole-Zero synthesized speech has less "ringing" quality than the corresponding All-Pole synthesized speech. Figure 5-6 shows corresponding segments of natural Pole-Zero synthesized and All-Pole synthesized speech.

A pilot experiment to test the performance of the Pole-Zero analysis-synthesis of speech in presence of additive colored noise was implemented as follows: Using the dynamic pole-zero filtering process described in Chapeter 4, the noisy speech was first filtered and then the Pole-Zero Vocoder (PZV) was applied to the resulting

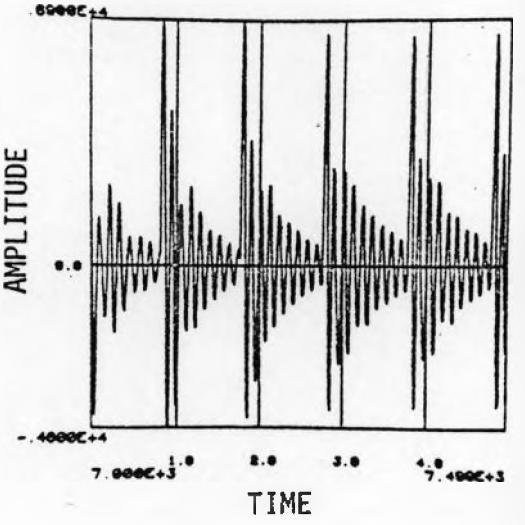
filtered speech. Figure 5-7 shows the spectrograms of corresponding portions of vocoded noisy speech, vocoded filtered speech and vocoded clean speech. Informal hearing tests reveal an improvement in the vocoded filtered speech over the vocoded noisy speech. The Pole-Zero coding of speech in the presence of noise, described in section 5.5, which combines the Pole-Zero filtering of noisy speech and the Pole-Zero coding of clean speech, seems to have the potential to generate improved coded speech.



(a)



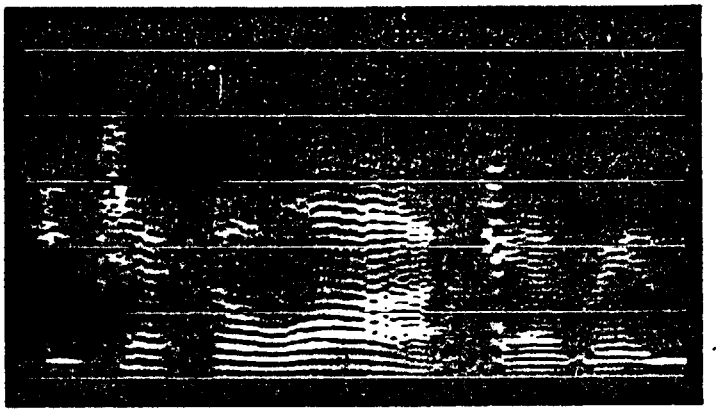
(b)



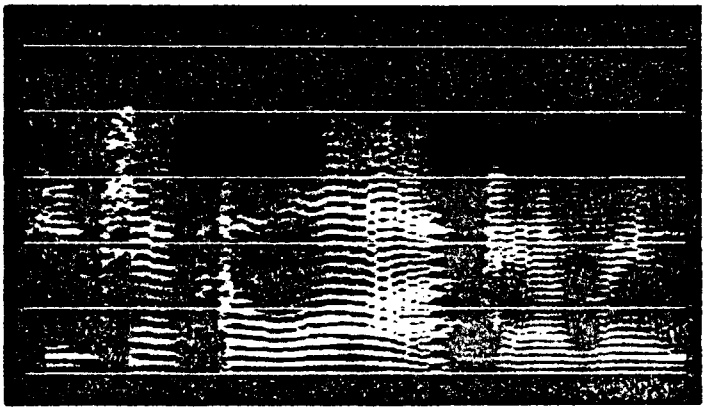
(c)

Figure 5-6. Corresponding segments of :
(a) Voiced natural speech.
(b) Pole-Zero synthesized speech (12P + 4Z).
(c) All-Pole synthesized speech (16P).

(a)



(b)



(c)

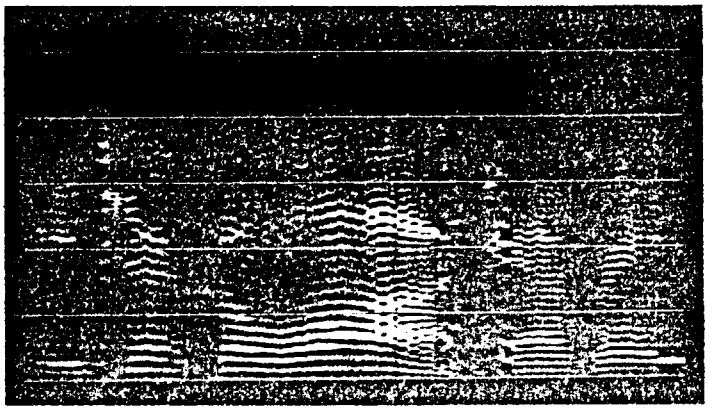


Figure 5-7. Spectrograms of corresponding portions of :
(a) Vocoded filtered speech.
(b) Vocoded clean speech.
(c) Vocoded noisy speech.
All three spectrograms have been scaled 6 dB/oct above 400 Hz.

CONCLUSIONS

6.1 Review

The goal of this research has been to develop new Pole-Zero modeling techniques and apply them to speech processing. This goal has been accomplished. Two Pole-Zero modeling techniques, Autocorrelation Partial Realization (APR) and Autocorrelation Prediction (AP), have been developed and their theories were established. APR, using partially linear operations, identifies the Pole-Zero model whose short time-lag autocorrelations exactly match those of a given spectrum. in contrast AP, using only linear operations, identifies the Pole-Zero model whose short time-lag autocorrelations closely approximate those of a given spectrum. Neither of them uses Fourier Transformation, but fast recursive and/or iterative algorithms to estimate the model parameters. APR and AP have been compared and the properties of the Pole-Zero models identified by them were discussed. It has been shown that the Pole-Zero model, identified by Autocorrelation Prediction, has advantages over the All-Pole model, identified by Linear Prediction (LP), when the the envelope of the given spectrum has deep valleys. A cascade of two lattice filters has been proposed as a realization of the Pole-Zero models identified by AP or APR.

A dynamic filtering process, based on Wiener filtering and Autocorrelation Prediction has been developed and implemented to suppress background noise from degraded speech. Using AP, rather

than LP, to model the estimated Wiener filter spectrum, has improved the performance of this dynamic filtering process. Moreover, a Pole-Zero Vocoder (PZV) based on AP has been developed and implemented. PZV becomes an All-Pole Vocoder (Linear Predictive Vocoder) on one extreme and an All-Zero Vocoder (different from the Homomorphic Vocoder but similar in quality of the synthesized speech) on the other extreme. For proper selection of the Pole-Zero model orders (M is roughly three times L), informal tests has shown that PZV generates more "natural sounding" synthesized speech than the other two extremes.

6.2 Future Research

To use the Pole-Zero model obtained from APR in applications other than spectral matching, like filtering, the stability of the model should be resolved. This area needs futher investigation.

To obtain a low bit rate high quality vocoder, the orders of the Pole-Zero model in PZV can be made dynamic; preferably equal to the parsimonious orders M and L for each frame of anlysis.

Modifying the PZV to account explicitly for the background noise of degraded speech seems promising when the noise is stationary and only the degraded speech is available. The restriction of the stationarity of the background noise may be relaxed if a correlated version of the background noise, recorded by another microphone away from the speaker, is also available.

Finally, the applications of Autocorrelation Prediction, as a Pole-Zero modeling technique, in areas other than speech processing are open for further research.

APPENDIX A

DURBIN AND PARCOR RECURSIVE ALGORITHMS

A system of linear equations can be solved recursively by the Bordering method [13]. When the coefficient matrix of the system of linear equations is a Symmetric Toeplitz matrix, then the Bordering method is simplified to Levinson's algorithm [42]. If the constant vector in the right hand side of the system of linear equations is also of the following form, equation A1, then Levinson's algorithm is further simplified to Durbin's algorithm [12, 27].

$$\begin{bmatrix} R_0 & R_1 & R_2 & \cdots & R_{M-1} \\ R_1 & R_0 & R_1 & \cdots & R_{M-2} \\ R_2 & R_1 & R_0 & \cdots & R_{M-3} \\ \vdots & & & \ddots & \vdots \\ R_{M-1} & R_{M-2} & R_{M-3} & \cdots & R_o \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_M \end{bmatrix} = - \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ \vdots \\ R_M \end{bmatrix} \qquad \text{(A1)}$$

Durbin's Recursive Algorithm

The system of linear equations (A1) can be solved recursively by Durbin's algorithm:

1) $E_0 = R_0$, $j = 0$,

2) $j = j + 1$,

3) $K_j = -\left[R_j + \sum_{i=1}^{j-1} a_i^{(j-1)} R_{j-i}\right] \Big/ E_{j-1}$,

4) $a_j^{(j)} = K_j$,

5) $a_j^{(j)} = a_i^{(j-1)} + K_j a_{j-i}^{(j-1)}$, $\quad 1 \leq i \leq j-1$,

6) $E_j = (1 - K_j^2)E_{j-1}$,

7) If $j = M$ stop; otherwise go to 2 (A2)

When the algorithm stops, the solution to the system of equation (A1) is:

$$a_i = a_i^{(M)}, \qquad 1 \leq i \leq M \ . \tag{A3}$$

The corresponding by-product parameters K_i, for $1 \leq i \leq M$ are referred to as the reflection coefficients, partial correlations or parcor parameters.

If the function R_k represents an autocorrelation function, then the coefficient matrix in (A1) is positive definite [34]. In this case, it can be shown that the parcor parameters K_i have the property

$$\left|K_i\right| < 1, \qquad 1 \leq i \leq M, \tag{A4}$$

and the polynomial $A(z) = 1 + \sum_{i=1}^{M} a_i z^{-i}$ is minimum phase. That is, all the roots of A(z) are inside the unit circle [18].

There is a unique set of parcor parameters K_i for the set of predictor coefficients $\{a_i\}$ and vice versa. To compute directly the corresponding parcor parameters $\{k_i\}$ from the predictor coefficients $\{a_i\}$, a recursive algorithm is derived. Substituting j-1 for i in step 5 of Durbin's algorithm and rearranging the terms gives

$$a_{j-i}^{(j-1)} = a_{j-i}^{(j)} - K_j \, a_i^{(j-1)}, \qquad 1 \leq i \leq j-1. \tag{A5}$$

Again substituting for $a_{j-i}^{(j-1)}$ in step 5 of Durbin's algorithm from (A5) and some simplifications results in

$$a_i^{(j-1)} = \left[a_i^{(j)} - K_j \, a_{j-i}^{(j)} \right] \Big/ (1-K_j^2), \qquad 1 \leq i \leq j-1, \tag{A6}$$

Thus from step 4 in Durbin's algorithm, (A3) and (A6) we obtain the sought for parcor recursive algorithm.

1) $j = M, \; a_i^{(M)} = a_i, \qquad 1 \leq i \leq M$,

2) $K_j = a_j^{(j)}$,

3) If j = 1 stop

4) $a_i^{(j-1)} = \left[a_i^{(j)} - K_j \, a_{j-i}^{(j)} \right] \Big/ (1-K_j^2), \qquad 1 \leq i \leq j-1$,

5) $j = j - 1$,

6) Go to 2 . (A7)

APPENDIX B

PADÉ APPROXIMATION

A function represented by a one-sided power series is approximated with a rational function using the Padé approximation.

Let $X(z)$ be a function represented by the following one-sided power series

$$X(z) = \sum_{k=0}^{\infty} x_k z^{-k}, \qquad \text{for} \quad |z| > r_1 . \tag{B1}$$

Also, consider the rational function

$$\hat{X}(z) = \frac{C(z)}{D(z)} = \frac{\sum_{i=0}^{L} c_i z^{-i}}{\sum_{i=0}^{M} d_i z^{-i}} \qquad \begin{array}{l} d_0 = 1, \\ d_M \neq 0, \\ L \leq M, \end{array} \tag{B2-a}$$

whose power series representation is

$$\hat{X}(z) = \sum_{k=0}^{\infty} \hat{x}_k z^{-k}, \qquad \text{for} \quad |z| > r_2 . \tag{B2-b}$$

From where (B2-a) and (B2-b), the coefficient $\hat{x}_k$ is equal to

$$\hat{x}_k = \sum_{i=0}^{L} c_i \delta(k-1) - \sum_{i=1}^{M} d_i \hat{x}_{k-i} \qquad k \geq 0 . \tag{B3}$$

To approximate the $X(z)$ by $\hat{X}(z)$, the

$$N = L + M + 1, \tag{B4}$$

unknown coefficients $\{d_i\}$ and $\{c_i\}$ are computed such that the first N terms of the $X(z)$ and $\hat{X}(z)$ power series representations are equal, i.e.

$$\hat{x}_k = x_k, \qquad 0 \leq k \leq N-1. \tag{B5}$$

Using (B5) in (B3) results in the following system of linear equations

$$x_k = \sum_{i=0}^{L} c_i \delta(k-i) - \sum_{i=1}^{M} d_i \, x_{k-i}, \qquad 0 \leq k \leq N-1 . \tag{B6}$$

In matrix form, equations (B6) become

$$\begin{bmatrix} x_L & x_{L-1} & \cdot\cdot & x_0 & 0 & \cdot\cdot & 0 & 0 \\ x_{L+1} & x_L & \cdot\cdot & x_1 & x_0 & \cdot\cdot & 0 & 0 \\ \cdot & \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot & \cdot \\ x_{M-1} & x_{M-2} & \cdot\cdot & x_L & x_{L-1} & \cdot\cdot & x_1 & x_0 \\ \cdot & \cdot & \cdot\cdot & \cdot & \cdot & \cdot\cdot & \cdot & \cdot \\ x_{L+M-2} & x_{L+M-3} & \cdot\cdot & x_{M-2} & x_{M-3} & \cdot\cdot & x_L & x_{L-1} \\ x_{L+M-1} & x_{L+M-2} & \cdot\cdot & x_{M-1} & x_{M-2} & \cdot\cdot & x_{L+1} & x_L \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ d_{M-L} \\ \cdot \\ d_{M-1} \\ \cdot \\ d_M \end{bmatrix} = - \begin{bmatrix} x_{L+1} \\ x_{L+2} \\ \cdot \\ x_M \\ \cdot \\ x_{L+M-1} \\ x_{M+L} \end{bmatrix}, \tag{B7-a}$$

$$c_k = \sum_{i=0}^{M} d_i \, x_{k-i} \,, \qquad d_o = 1 \,, \qquad 0 \leq k \leq L \,, \tag{B7-b}$$

where the matrix $(x_{L+i-j})_1^M$ (B7-a) is non-symmetric toeplitz. The Trench recursive algorithm [39, Appendix C] can be used solve the system of linear equations (B7-a). After computing the $\{d_i\}$ the coefficients $\{c_i\}$ are obtained simply by performing the summation in (B7-b).

Padé Approximation Error

Using (B1), (B2-b) and (B5), the Padé approximation error, Figure B-1, is defined by the following power series

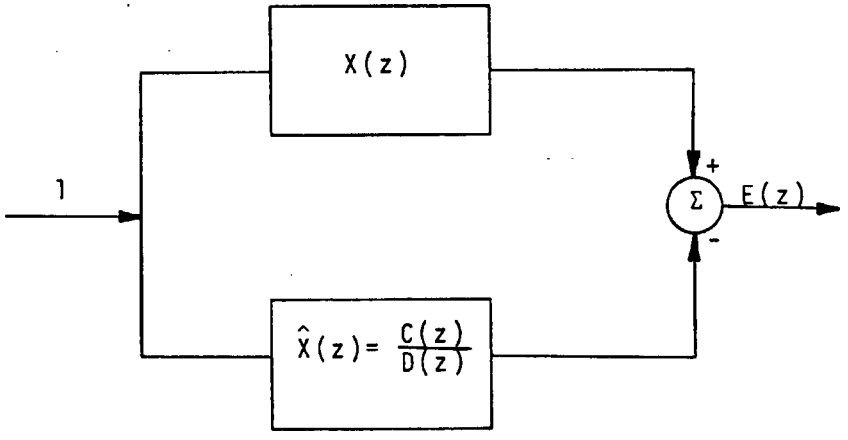


Figure B-1. Padé approximation error

$$E(z) = \sum_{k=0}^{\infty} e_k z^{-k} \triangleq x(z) - \hat{x}(z) = \sum_{k=0}^{\infty} (x_k - \hat{x}_k) z^{-k}$$

$$= \sum_{k=N}^{\infty} (x_k - \hat{x}_k) z^{-k}$$

$$|z| > \text{Max}(r_1, r_2). \qquad \text{(B8-a)}$$

Using (B5) and substituting for $\hat{x}_k$ from (B3) in (B8-a) gives another expression for E(z), i.e.

$$E(z) = \sum_{k=0}^{\infty} e_k z^{-k} = \sum_{k=N}^{\infty} \left[\sum_{i=0}^{M} d_i x_{k-i} \right] z^{-k} . \qquad \text{(B8-b)}$$

Relation (B8-a) shows that the first discrepancy between the corresponding power series coefficients of the X(z) and $\hat{X}(z)$ occurs at k = N. The relation (B8-b) shows that different errors E(z) are obtained, depending on the selection of the orders L and M for a given N. All of these different errors E(z), nevertheless, have the following property:

$$e_k = 0 , \qquad 0 \leq k \leq N-1 . \qquad \text{(B9)}$$

Special Case

The Padé approximation $\frac{C(z)}{D(z)}$ for the special case L=M, and consequently N=2M+1, is considered here. Also, the value of the first discrepancy e_N is expressed in terms of only the given power series coefficients x_k.

The matrix equation (B7-a) for L=M takes the form

$$\begin{bmatrix} x_M & x_{M-1} & \cdots & x_2 & x_1 \\ x_{M+1} & x_M & \cdots & x_3 & x_2 \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ \cdot & \cdot & \cdots & \cdot & \cdot \\ x_{2M-2} & x_{2M-3} & \cdots & x_M & x_{M-1} \\ x_{2M-1} & x_{2M-2} & \cdots & x_{M+1} & x_M \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ \cdot \\ \cdot \\ d_{M-1} \\ d_M \end{bmatrix} = - \begin{bmatrix} x_{M+1} \\ x_{M+2} \\ \cdot \\ \cdot \\ x_{2M-1} \\ x_{2M} \end{bmatrix} . \qquad \text{(B10)}$$

The non-symmetric toeplitz matrix $(x_{M+i-j})_1^M$ in (B10) is symmetric around its second diagonal. Therefore, by rearranging the unknown coefficients $\{d_i\}$ of (B10) in reverse order, the matrix equation (B10) takes the following form

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_{M-1} & x_M \\ x_2 & x_3 & \cdots & x_M & x_{M+1} \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ x_{M-1} & x_M & \cdots & x_{2M-3} & x_{2M-2} \\ x_M & x_{M+1} & \cdots & x_{2M-2} & x_{2M-1} \end{bmatrix} \begin{bmatrix} d_M \\ d_{M-1} \\ \cdot \\ \cdot \\ d_2 \\ d_1 \end{bmatrix} = - \begin{bmatrix} x_{M+1} \\ x_{M+2} \\ \cdot \\ \cdot \\ x_{2M-1} \\ x_{2M} \end{bmatrix} . \qquad \text{(B11)}$$

The symmetric matrix $(x_{i+j-1})_1^M$ in (B11) is a Hankel matrix. The Berlekamp-Massey algorithm (Appendix C) can be used to solve the system of linear equations (B11). From (B8-b) and (B5), the first

discrepancy e_N, for the case $L=M$ is given by

$$e_N = e_{2M+1} = \sum_{i=0}^{M} d_i \, x_{2M+1-i} \, . \tag{B12}$$

To compute e_N directly in terms of only the given power series coefficients x_k, the system of equations (B11) is augmented by the equation (B12). e_N is considered as an unknown. This leads to

$$e_{2M+1} = \frac{\left| x_{i+j-1} \right|_1^{M+1}}{\left| x_{i+j-1} \right|_1^{M}} \, , \tag{B13}$$

where $\left| x_{i+j-1} \right|_1^{M}$ is the determinant of the Hankel matrix $(x_{i+j-1})_1^M$ given in (B11) and the $\left| x_{i+j-1} \right|_1^{M+1}$ is the determinant of the following Hankel matrix:

$$\left[\begin{array}{cccc|c} x_1 & x_2 & \cdots & x_M & x_{M+1} \\ x_2 & x_3 & \cdots & x_{M+1} & x_{M+2} \\ \cdot & & & \cdot & \cdot \\ \cdot & & & \cdot & \cdot \\ \cdot & & & \cdot & \cdot \\ x_M & x_{M+1} & \cdots & x_{2M-1} & x_{2M} \\ \hline x_{M+1} & x_{M+2} & \cdots & x_{2M} & x_{2M+1} \end{array}\right] . \tag{B14}$$

APPENDIX C

TRENCH AND BERLEKAMP-MASSEY RECURSIVE ALGORITHMS

In using the Padé approximation, one encounters the system of linear equations of the form

$$\begin{bmatrix} \phi_0 & \phi_{-1} & \phi_{-2} & \cdots & \phi_{-M+1} \\ \phi_1 & \phi_0 & \phi_{-1} & \cdots & \phi_{-M+2} \\ \phi_2 & \phi_1 & \phi_0 & \cdots & \phi_{-M+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{M-1} & \phi_{M-2} & \phi_{M-3} & \cdots & \phi_0 \end{bmatrix} \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \vdots \\ \psi_M \end{bmatrix} = - \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_M \end{bmatrix}, \tag{C1}$$

where the coefficient matrix $(\phi_{i-j})_1^M$ in (C1) is a non-symmetric Toeplitz matrix. Using the Bordering method [13], Trench [39], and Whittle [41] derive a fast recursive algorithm to solve simultaneously (C1) and the following system of equations

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \cdots & \phi_{M-1} \\ \phi_{-1} & \phi_0 & \phi_1 & \cdots & \phi_{M-2} \\ \phi_{-2} & \phi_{-1} & \phi_0 & \cdots & \phi_{M-3} \\ \vdots & & & & \vdots \\ \phi_{-M+1} & \phi_{-M+2} & \phi_{-M+3} & \cdots & \phi_0 \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \eta_3 \\ \vdots \\ \eta_M \end{bmatrix} = - \begin{bmatrix} \phi_{-1} \\ \phi_{-2} \\ \phi_{-3} \\ \vdots \\ \phi_{-M} \end{bmatrix}. \tag{C2}$$

Trench's Recursive Algorithm:

1) $\Delta_0 = \phi_0, \quad j = 0,$

2) $j = j + 1,$

3) $\psi_j^{(j)} = -\left[\phi_j + \sum_{i=1}^{j-1} \psi_i^{(j-1)} \phi_{j-i}\right] \Big/ \Delta_{j-1},$

4) $\eta_j^{(j)} = -\left[\phi_{-j} + \sum_{i=1}^{j-1} \eta_i^{(j-1)} \phi_{-(j-i)}\right] \Big/ \Delta_{j-1},$

5) $\psi_i^{(j)} = \psi_i^{(j-1)} + \psi_j^{(j)} \eta_{j-i}^{(j-1)}, \qquad 1 \le i \le j-1,$

6) $\eta_i^{(j)} = \eta_i^{(j-1)} + \eta_j^{(j)} \psi_{j-i}^{(j-1)}, \qquad 1 \le i \le j-1,$

7) $\Delta_j = (1 - \psi_j^{(j)} \eta_j^{(j)}) \Delta_{j-1},$

8) If $j = M$ stop; otherwise go to 2.

The system of equations (C1) becomes identical to the system of equations (B7-a) by replacing ϕ_k by

$$\phi_k = \begin{cases} x_{L+k} & -L \le k \le M-1, \\ 0 & \text{Otherwise}, \end{cases} \tag{C3}$$

and ψ_k by d_k for $1 \le k \le M$. Thus Trench's recursive algorithms can be used to solve the system of equations (B7-a) encountered in the Padé approximation. Note also, the system of equations (C1) becomes identical to (C2), except for the unknown names, when the coefficient matrix $(\phi_{i-j})_1^M$ is symmetric, i.e., $\phi_k = \phi_{-k}$. In this case, the Trench

algorithm reduces to Durbin's algorithm.

Berlekamp-Massey Algorithm:

1) $1 \rightarrow D(z) \quad 1 \rightarrow P(z) \quad 1 \rightarrow \beta$

$0 \rightarrow M \quad 1 \rightarrow p \quad 0 \rightarrow k.$

2) If $k = N$, stop; otherwise compute

$$e = x_k + \sum_{i=1}^{M} d_i \, x_{k-i} .$$

3) If $e = 0$, then $\beta + 1 \rightarrow \beta$ and go to 6).

4) If $e \neq 0$ and $2M > k$, then

$D(z) - ep^{-1} z^{-\beta} P(z) \rightarrow D(z)$

$\beta + 1 \rightarrow \beta$

and go to 6).

5) If $e \neq 0$ and $2M \leq k$, then

$D(z) \rightarrow T(z)$ (temporary storage of $D(z)$)

$D(z) - ep^{-1} z^{-\beta} P(z) \rightarrow D(z)$

$k + 1 - M \rightarrow M$

$T(z) \rightarrow P(z)$

$e \rightarrow p$

$1 \rightarrow \beta.$

6) $k + 1 \rightarrow k$ and return to 2).

where

D(z): current denominator

M: order of current denominator

p(z): last denominator of lower order

e: next discrepancy

p: the discrepancy corresponding to the last denomination of lower order

k: current number of matched terms

β: difference between current number of matched terms and that corresponding to last denominator of lower order.

The Berlekamp-Massey algorithm [7, 30] recursively finds the weights $\{d_i\}$ of the shortest Linear Feedback Shift Register (LFSR) that can generate exactly the given sequence $\{x_k\}_{i=0}^{N-1}$, Figure C-1. The LFSR given in Figure C-1 represents the rational function

$$\frac{C'(z)}{D(z)} = \frac{\sum_{i=0}^{M-1} c'_i z^{-1}}{\sum_{i=0}^{M} d_i z^{-1}} = \sum_{i=0}^{\infty} x_i z^{-1}, \qquad d_0 = 1. \tag{C4}$$

The coefficients $\{c'_i\}$ are obtained from the first M values of the given sequence $\{x_k\}_0^{N-1}$ stored in the shift register as initial values, and the weights $\{d_i\}$ using the following relation:

$$c'_i = \sum_{k=0}^{M} d_k x_{i-k}, \qquad 0 \le i \le M-1 .$$

Hence one can use the Berlekamp-Massey algorithm to obtain the denominator coefficients of the least dimension rational function of the form

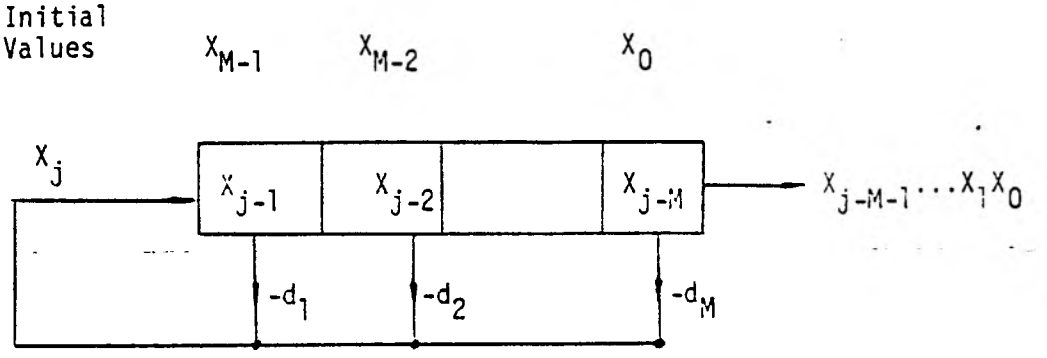


Figure C-1. Linear Feedback Shift Register (LFSR) representing $\frac{C'(z)}{D(z)}$.

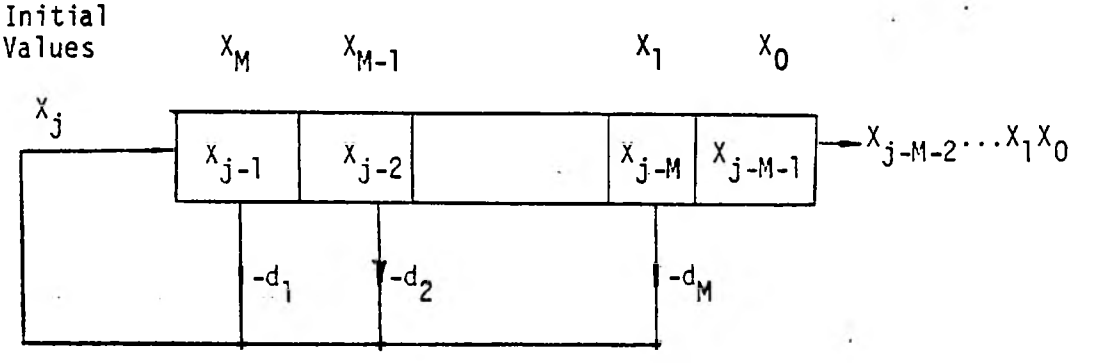


Figure C-2. Linear Feedback Shift Register (LFSR) representing $\frac{C(z)}{D(z)}$.

(C4) for which the first N terms of its power series representations are given.

We are interested, however, in the rational function of the form

$$\frac{C(z)}{D(z)} = \frac{\sum_{i=0}^{M} c_i z^{-1}}{\sum_{i=0}^{M} d_i z^{-1}} = c_0 + z^{-1} \frac{\sum_{i=0}^{M-1} c_i' z^{-1}}{\sum_{i=0}^{M} d_i z^{-1}} = x_0 + \sum_{i=1}^{\infty} x_i z^{-1}, \quad d_0 = 1, \tag{C5}$$

rather than the form (C4). From (C5) and (C4) one concludes that the denominator coefficients $\{d_i\}$ in (C5) are the feedback weights of the shortest LFSR which generates the sequence $\{x_i\}_{i=1}^{N-1}$ rather than the sequence $\{x_i\}_{i=0}^{N-1}$.

Therefore, applying the Berlekamp-Massey algorithm to the sequence $\{x_i\}_{i=1}^{N-1}$ gives the desired denominator coefficients $\{d_i\}$. Using (C5), the numerator coefficients are obtained from the following formula

$$c_i = \sum_{k=0}^{i} d_k x_{i-k}, \qquad 0 \leq i \leq M. \tag{C6}$$

Figure C-2 represents the rational function $\frac{C(z)}{D(z)}$ defined in (C5).

If the given sequence $\{x_i\}_{i=0}^{N-1}$ was indeed generated by some unknown rational function of the form (C5) and the condition $N \geq 2M+1$ holds, then the Berlekamp-Massey algorithm detects the order M and gives the corresponding denominator coefficients $\{d_i\}_{i=1}^{M}$. When the sequence $\{x_i\}_{i=0}^{N-1}$ corresponds to some real data, however, the Berlekamp-Massey algorithm applied to the sequence $\{x_i\}_{i=1}^{N-1}$, recursively computes the solution to the following system of equations, (C7), of order

$M = [N/2]$. The $[N/2]$ means the largest integer less than or equal to the real number $N/2$.

$$\begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_M \\ x_2 & x_3 & x_4 & \cdots & x_{M+1} \\ x_3 & x_4 & x_5 & \cdots & x_{M+2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ x_M & x_{M+1} & x_{M+2} & \cdots & x_{2M-1} \end{bmatrix} \begin{bmatrix} d_M \\ d_{M-1} \\ d_{M-2} \\ \cdot \\ \cdot \\ \cdot \\ d_1 \end{bmatrix} = - \begin{bmatrix} x_{M+1} \\ x_{M+2} \\ x_{M+3} \\ \cdot \\ \cdot \\ \cdot \\ x_{2M} \end{bmatrix} . \qquad \text{(C7)}$$

APPENDIX D

NEWTON-RAPHSON ITERATIVE ALGORITHM

To decompose the symmetric polynomial

$$P(z) = \sum_{i=-M}^{M} p_{|i|}\, z^{-i}\,, \tag{D1}$$

into the product of the form

$$B(z)B(1/z) = \left(\sum_{i=0}^{M} b_i z^{-i}\right)\left(\sum_{i=0}^{M} b_i z^{i}\right) = \sum_{i=-M}^{M}\left(\sum_{j=0}^{M-|i|} b_j b_{j+|i|}\right) z^{-i}, \tag{D2}$$

the polynomial $P(z)$ should be non-negative on the unit circle in z-plane. Assuming $P(z)$ has this property, an iterative algorithm based on Newton-Raphson method is formulated to compute the polynomial $B(z)$ such that

$$P(z) - B(z)B(1/z) = \sum_{i=-M}^{M}\left(p_{|i|} - \sum_{j=0}^{M-|i|} b_j b_{j+|i|}\right) z^{-i} = 0\,. \tag{D3}$$

Consider the vector function (f_i) whose i-th element is defined as

$$f_i = p_i - \sum_{j=0}^{M-i} b_j b_{j+i}\,, \qquad 0 \le i \le M\,, \tag{D4}$$

Computing the B(z) such that the relation (D3) holds, is equivalent to finding the vector

$$(b_i) = \begin{bmatrix} b_0 \\ b_1 \\ \cdot \\ \cdot \\ b_M \end{bmatrix}, \tag{D5}$$

for which the vector function (f_i) has zero value. Therefore, the Newton-Raphson iterative method can be used to find the vector (b_i) and as a result, the desired polynomial B(z). The iterative Newton-Raphson formula for the vector function (f_i) takes the form

$$(f_i) + [(b_i)^{t+1} - (b_i)^t] \frac{\partial(f_i)}{\partial(b_i)} = 0 , \tag{D6}$$

where $(b_i)^t$ is the estimate for the desired (b_i) at the t-th iteration.

From (D2), the matrix $T \triangleq \frac{\partial(f_i)}{\partial(b_i)}$ is equal to

$$T \triangleq \frac{\partial(f_i)}{\partial(b_i)} = \left(\frac{\partial f_i}{\partial b_j}\right) = \begin{bmatrix} b_0 & b_1 & \cdots & b_M \\ b_1 & b_2 & & b_M \\ \cdot & & \cdot & \\ \cdot & & \cdot & \\ \cdot & \cdot & & \bigcirc \\ b_M & & & \end{bmatrix} + \begin{bmatrix} b_0 & b_1 & \cdots & b_M \\ & b_0 & \cdots & b_{M-1} \\ & & \cdot & \cdot \\ & & \cdot & \cdot \\ & \bigcirc & & \cdot \\ & & & b_0 \end{bmatrix} \tag{D7}$$

Using (D7) in (D6) and rearranging the terms, results in

$$(b_i)^{t+1} = (b_i)^t - T^{-1} (f_i) . \tag{D8}$$

The relations (D4) and (D7) and (D8) provide the iterative algorithm for computing the desired vector (b_i).

Assuming $p_o>0$, the starting vector $(b_i)^0$ can be chosen to be

$$(b_i)^0 = \begin{bmatrix} \sqrt{p_o} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} . \qquad \text{(D9)}$$

The iteration is considered to have converged when for some prescribed value ε the following inequalities hold

$$|f_i| < \varepsilon , \qquad 0 \leq i \leq M . \qquad \text{(D10)}$$

It can be shown that the convergence is of second order and the polynomial $B_t(z)$ corresponding to the vector $(b_i)^t$, having the $(b_i)^0$ as starting vector, is minimum phase [43].

REFERENCES

[1] H. Akaike, "Power spectrum estimation through autoregressive model fitting," Ann. Inst. Statist. Math., vol. ,21,pp. 407-419, 1969.

[2] ---, "Statistical predictor identification," Ann. Inst. Statist. Math., vol. 22, pp. 203-217, 1970.

[3] ---, "Information theory and an extension of the maximum likelihood principle," in Proc. 2nd Int. Symp. Information Theory (Supplement to Problems of Control and Information Theory), 1972.

[4] ---, "A new look at statistical model identification," IEEE Trans. Automat. Contr., vol. AC-19, pp. 716-723, Dec. 1974.

[5] B. S. Atal and S. L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," J. Acoust. Soc. Amer., vol. 50, no. 2, pp. 637-655, 1971.

[6] M. A. Atashroo, "Cepstral prediction analysis of digital waveforms,"IEEE Proc. of region six-western USA-conference, pp. 68-72, 1975.

[7] E. R. Berlekamp, Algebraic Coding Theory, New York: McGraw-Hill Book Corp., 1968, p. 180.

[8] S. F. Boll, "A priori ,digital speech analysis," Computer Science Div., University of Utah, Salt Lake City, UTEC-CSC-73-123, 1973.

[9] ---, " Selected methods for improving synthesis speech quality using linear predictive coding: system description, coefficients smoothing and STREAK," Computer Science, Univ. of Utah, Salt Lake city, UTEC-CS-74-151, Nov. 1974.

[10] G. E. Box and G. M. Jenkins, Time Series Analysis Forecasting and Control. San Francisco, Calif.: Holden-Day, 1970. p. 74.

[11] M. J. Chang, "A comparitive performance study of several pitch detection algorithms," M.S. thesis, M.I.T, June 1975.

[12] J. Durbin, "The fitting of time-series models," Rev. Inst. Int. Statist., vol. 28, no. 3, pp. 233-243, 1960.

[13] D. K. Faddeev and V. N. Faddeeva, Computational Methods of Linear Algebra. San Francisco, Calif.: Freeman, 1963, p. 163.

[14] J. L. Flanagan, Speech Analysis Synthesis and Perception, 2nd Edition. New York: Springer-Verlag, 1972.

[15] W. Gersch and D. R. Sharpe, "Estimation of power spectra with finite-order autoregressive models," IEEE Trans. Automat. Contr., vol. AC-18, pp. 367-369, Aug. 1973.

[16] B. Gold and C. M. Rader, Digital Processing of Signals. New York: McGraw-Hill, 1969, chapter 2.

[17] A. H. Gray, Jr., J. D. Markel, "Digital lattice and ladder filter synthesis," IEEE Trans. on Audio and Electroacoustics, vol. AV-21, no. 6, pp. 491-500, Dec. 1973

[18] U. Grenander and G. Szego, Toeplitz Forms and Their Applications. Berkeley, Calif.: Univ. California Press, 1958.

[19] F. Itakura and S. Saito, "Digital filtering techniques for speech analysis and synthesis," in Conf. Rec., 7th Int. Congr. Acoustics, Paper 25 C 1, 1971.

[20] G. E. Kopec, "Speech analysis by homomorphic prediction," S.M. thesis, Dept. of Elect. Eng. M.I.T, Nov. 1975.

[21] G. E. Kopec, A. V. Oppenheim, J. M. Tribolet,"Speech analysis by homomorphic prediction," submitted to IEEE Trans. on acoustics, speech and signal processing.

[22] R. E. Kromer, "Asymptotic properties of the autoregressive spectral estimator," Ph.D. dissertation. Dept. of Statistics, Stanford Univ., Stanford, Calif., 1969.

[23] N. Levinson, "The Wiener RMS (root mean square) error criterion in filter design and prediction," J. Math. Phys., vol. 25, no. 4, pp. 261-278, 1947. Also Appendix B, in N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series. Cambridge Mass., M.I.T. Press, 1949.

[24] J. Makhoul and J. J. Wolf, "Linear prediction and the spectral analysis of speech," Bolt Beranek and Newman Inc., Cambridge, Mass., NTIS AD-749066, Rep. 2304, Aug. 1972.

[25] J. Makhoul, "Spectral analysis of speech by linear prediction," IEEE Trans. Audio Electroacoust., vol. AU-21, pp. 140-148, June 1973.

[26] ---, "Speech compression research at BBN," BBN report no. 2976, December 1974.

[27] ---, "Linear prediction: a tutorial review," Proc. IEEE, special issue on digital signal processing, vol. 63, no. 4, April 1975.

[28] J. D. Markel, "The SIFT algorithm for fundamental frequency extimation," IEEE Trans. on audio and electroacoustics. vol. AM-20, no. 5, December 1972.

[29] J. D. Markel, A. H. Gray, Jr., Linear prediction of speech, New York: Springer-Verlag, 1976.

[30] S. L. Massey, "Shift-register synthesis and BCH decoding," IEEE Trans. on Inf. Theory, vol IT-15, no. 1, January 1969.

[31] A. V. Oppenheim, R. W. Schafer and T. G. Stockham, "Nonlinear filtering of multiplied and convolved signals," Proc. IEEE, vol. 56, pp. 1264-1291, Aug. 1968.

[32] A. V. Oppenheim and R. W. Schafer, Digital Signal Processing, Englewood cliffs., New Jersey: Prentice-Hall, p. 66, 1975.

[33] A. V. Oppenheim, G. Kopec, J. Tribolet, "Signal analysis by homomorphic prediction," to be published in IEEE Trans. on acoustics, speech and signal processing.

[34] A. Papoulis, Probability, Random Variables, and Stochastic Processes, ps. 349 and 431, New York: McGraw-Hill, 1965.

[35] E. Parzen, "Some recent advances in time series modeling, IEEE Trans. on Auto. Control, vol. AC-19, no. 6, December 1974.

[36] L. R. Rabiner, B. Gold, Theory and Application of Digital Signal Processing, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1975.

[37] E. A. Robinson, "Predictive decomposition of time series with application to seismic exploration," Geophysics, vol. 32, no. 3, pp. 418-484, June 1967.

[38] R. W. Schafer and L. R. Rabiner, "Digital representations of speech signals," IEEE Proc., special issue on digital signal processing, vol. 63, no. 4, pp. 662-677, April 1975

[39] W. F. Trench, "An algorithm for the inversion of finite toeplitz matrics," L. Soc. Indust. Appl. Math, vol. 12, no. 3 pp. 515-522, printed in USA, September 1964.

[40] J. M. Tribolet, "Identification of linear discrete systems with applications to speech processing," M.S. thesis, Dept. Elect. Eng., M.I.T., Cambridge, Mass., Jan. 1974.

[41] P. Whittle, "On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix," Biometrika, vol. 50, nos. 1 and 2, pp. 129-134, 1963.

[42] N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications, Appendix B, Cambridge, Mass., M.I.T. Press, 1949.

[43] G. Wilson, "Factorization of the covariance generating function of a pure moving average process," SIAM J. Numer. Anal. vol. 6, no. 1, pp. 1-7, March 1969.

[44] J. D. Wise, J. R. Caprio, T. W. Parks, "Maximum likelihood pitch estimation," Rice Univ., Technical Report 7516, Oct. 17, 1975.

ACKNOWLEDGEMENTS

The successful completion of this dissertation required valuable assistance and support from many fine people.

I am indebted to Dr. Steven F. Boll and Dr. Thomas G. Stockham, Jr. who co-supervised this work, and whose support, insight and encouragement made this work an enjoyable experience for me. I would also like to thank Dr. LaMar K. Timothy and Dr. Alan V. Oppenheim for their initial input and continued enthusiasm.

I express my appreciation to Steven Boll, Mike Callahan, Tracy Petersen, James Kajiya, Brent Baxter and Robert Armantrout for stimulating discussions and software contributions. Particularly, I am grateful to Bill Done for his careful reading of the dissertation draft.

Also, special thanks must go to Carol Brown for her beautiful typing of the formuli and Mike Milochik for his professional photographic reproduction.

At the end, I would like to express a special love and appreciation to my mother, Tahereh, who implanted within me the desire for scholastic achievements.