SPECTRAL ESTIMATION OF GASTROINTESTINAL SIGNALS

by

© Yassin Saad El-Cherif, B.Sc.E.E., M.Sc.E.E.

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Engineering

McMaster University

August, 1981

MASTER OF ENGINEERING (1981) McMASTER UNIVERSITY
Electrical and Computer Engineering Hamilton, Ontario

TITLE: Spectral Estimation of Gastrointestinal Signals

AUTHOR: Yassin Saad El-Cherif
B.Sc.E.E. (Cairo University)
M.Sc.E.E. (Cairo University)

SUPERVISOR: Dr. S.K. Sarna

NUMBER OF PAGES: vii, 105

ABSTRACT

The problem of estimating the power spectral density of Gastrointestinal (GI) signals is studied. Classical nonparametric methods based on the discrete Fourier Transform are presented along with modern parametric methods which are generally based on autoregressive-moving average (ARMA) time series models. Some algorithms of ARMA parameter estimation are presented under the classification of optimal and suboptimal methods. Two recently proposed ARMA methods are particularly considered and proven to be equivalent. Additive interference between autoregressive signals is shown to produce an ARMA process with equal orders of the MA and the AR parts. This has been made use of in modelling and spectral estimation of the Electrical Control Activity (ECA) in the small intestine.

Four methods for spectral estimation of GI signals have been implemented in a general minicomputer based program.

The performance of the different methods is demonstrated by examples taken from small intestinal ECA. The spatial distribution and the temporal variations have been investigated for the ECA spectra in the small intestine.

ACKNOWLEDGEMENTS

The author would like to thank Dr. S. K. Sarna for his guidance and supervision throughout the course of this work.

The stimulating discussions of Dr. S. N. Reddy are greatly appreciated.

Thanks are also due to the colleagues R. Bulat, V. Carbajal, F. Hung, P. Northcott and J. Yu.

The excellent typing of Miss Linda Hunter is very much acknowledged by the author.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Gastrointestinal Electrical Activity

The signals recorded from the smooth muscle layers of the digestive tract have gained increasing attention in the last decade. There have been remarkable efforts in the analysis and modelling of these signals and in interpreting their patterns to explain the motility function of the different parts of the digestive tract [e.g. 1-6]. The presently accepted classification of the control mechanisms controlling gastrointestinal motility is as follows:

1.1.1 Myogenic Control

This refers to the control mechanism governed by the electrical activity of the smooth muscle layers. The recorded signals are thought to originate in the longitudinal muscle layer and are thought to be a myogenic phenomenon (i.e. autonomous muscular activity) without the intervention of any neural activity. This implies that there is an oscillatory behaviour in the ionic flow across the membrane of the smooth muscle cell. This oscillatory behaviour is common in many other involuntary biological phenomena. These signals are of the rhythmic type which means that they exhibit some periodicity. In fact, this periodicity is sometimes very obvious and the number of

fundamental periods can be counted by visual inspection. However, in many other cases, especially in the signals recorded from the colon and the distal small intestine, the periodicity cannot be determined by visualization due to the existence of large background noise. This rhythmic activity is continuously present in normal subjects, whether contractions occur or not. The collection of such signals recorded from different positions in a certain organ has been termed "Electrical Control Activity" or simply ECA [6]. It is sometimes also called "Slow wave activity". The reason for these terminologies will be clear when we talk about other control mechanisms.

1.1.2 Neurogenic Control

There is another kind of electrical activity which appears in the form of a high frequency signal superimposed on the ECA and associated with a mechanical activity.

This phenomenon is thought to be due to neural excitation which results in both the mechanical contraction and the high frequency electrical signals. Henceforth, the latter has been termed "Electrical Response Activity (ERA)" to indicate that its presence comes in response to a neural input. This also clarifies the terminology given to the relatively slow wave activity as being a control signal. In effect, the ECA at a certain position is responsible for determining the maximum rate of contractions at that position as it appears that there could be no more than one contraction during a control wave cycle. Also, in the same sense, the totality of the ECA at different positions

in a certain organ is responsible for organizing the contraction pattern in this organ.

1.1.3 Hormonal Control

The neural control takes place upon the release of a chemical mediator (Acetylcholine). The release of Acetylcholine may take place in response to some hormones such as pentogastrin that stimulates the nerves or it may be spontaneous. However, in either case, the presence of the relevant hormones may cause contractions and ERA.

1.2 Scope and Objectives

This thesis is concerned only with the analysis of the ECA, the goal being to determine the frequency contents in the signals recorded from the small intestine.

It has been necessary to study methods of spectral analysis, which are either well known or recently proposed. Namely, these methods are: The periodogram and the standard autocorrelation methods and the autoregressive and the autoregressive-moving average methods. For this purpose, a minicomputer based general program has been developed which is capable of providing quantitative information as well as convenient graphing of the spectra of the analyzed signals.

1.3.1 Basic Definitions and Assumptions

Throughout this thesis, assumptions of ergodicity and weak (second order) stationarity are always implied [7]. By weak station-

arity, we mean that the first and the second order statistics of the process do not change with time. Ergodicity means that these statistics can be inferred from the time history of the signal. In making these assumptions, we are overwhelmed by the fact that the signal is only a single member (sample function) of a stochastic process. Therefore, time averaging is used instead of ensemble averaging. The assumption of stationarity is also necessary if we are to follow Wiener definition of the power spectral density.

1.3.1 Wiener Definition of the Power Spectral Density [S(f)]

The spectral density function of an ergodic second order stationary process x(t) is given by:

$$S(f) = \int_{-\infty}^{\infty} r(\tau)\, e^{-j2\pi f\tau}\, d\tau \tag{1.1}$$

where $r(\tau)$ is the autocorrelation function and is given by:

$$r(\tau) = \lim_{P\to\infty} \frac{1}{2P} \int_{-P}^{P} x(t)\, x(t+\tau)\, dt \tag{1.2}$$

This means that S(f) and $r(\tau)$ are a Fourier transform pair and each of them can be obtained from the other by the appropriate transformation. It should be noted that the total power in the signal x(t) is the same as r(0) which can be obtained from the inverse Fourier transform of S(f) with $\tau = 0$. This gives:

$$r(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(f)\, df \tag{1.3}$$

This makes it obvious why S(f) is termed power spectral density, as it gives the power distribution over the frequency scale. For rhythmic signals, S(f) is expected to peak around the repetition frequencies.

1.3.2 Power Spectral Density of a Discrete Time Signal

As a result of sampling, all the signals that we are dealing with are actually functions in discrete time, i.e. they are defined only at the sampling instants. Accordingly, the autocorrelation function becomes a sequence of autocorrelation coefficients defined for integral multiples of the sampling period. This is equivalent to multiplying the autocorrelation function by an impulse train, and the spectrum of the discrete process will be given by:

$$S(f) = \int_{\infty}^{\infty} \sum_{k=-\infty}^{\infty} \delta(t-kT)\, r(t)\, e^{-j2\pi fkT}\, dt \qquad (1.4)$$

$$S(f) = \sum_{k=-\infty}^{\infty} r(kT)\, e^{-j2\pi fkT} \qquad (1.5)$$

where T is the sampling period. The above discrete Fourier transform is an infinite Fourier series expansion which is an even periodic function in frequency with a period of repetition equal to 1/T. This means that the sampling rate should be taken at least equal to twice the maximum frequency in the signal to avoid overlapping of the spectra centered around the zero frequency and 1/T [8]. If such an overlap occurs, the resulting phenomenon is known as "aliasing". We will make an assumption that the signal is band limited to a frequency less than

half the sampling rate. This can always be validated by analog filtering prior to sampling.

1.3.3 Power Spectral Density of a Truncated Time Series

In addition to the periodicity of the spectrum resulting from the sampling process, the number of correlation lags used in estimating the spectrum has another unavoidable effect. For example, if a maximum lag (L) is considered, this will effectively result in multiplying the original autocorrelations by the "box car" function as shown in fig. 1.1. This gives the following estimate:

$$\hat{S}(f) = \sum_{-\infty}^{\infty} w(kT)\, r(kT)\, e^{-j2\pi fkT} \qquad (1.6)$$

This multiplication has the effect of a convolution in the frequency domain, which means that the original spectrum becomes convolved with the Fourier transform of the box car function as shown in fig. 1.c. The result of the convolution may be negative at some frequencies.

1.4 Organization of the Thesis

The rest of the thesis is organized as follows:

Chapters 2 and 3 are devoted to discussing the methods of spectral estimation. Chapter 2 deals briefly with the periodogram and the autocorrelation methods which are based on calculating the discrete Fourier Transform. These methods have been extensively treated in literature (e.g. 6, 7, 8). Then the autoregressive/maxi-

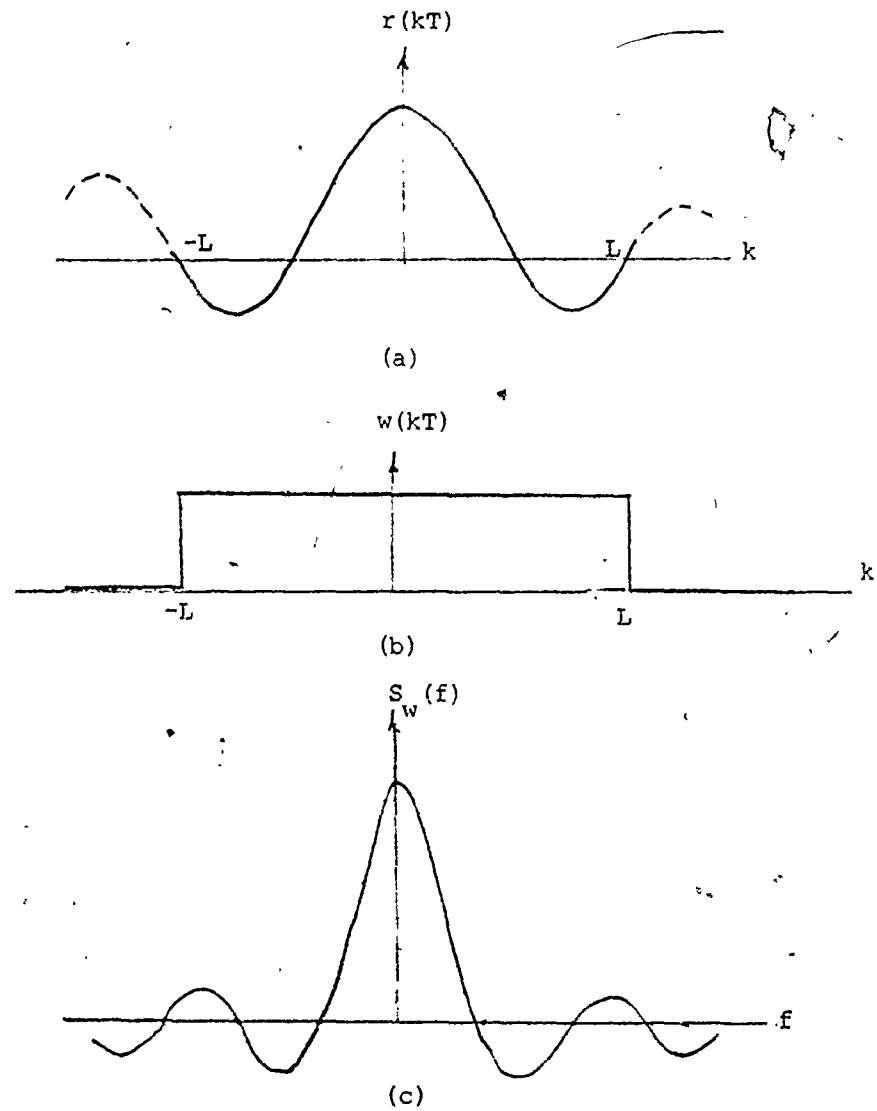


Fig 1.1 a) The truncated autocorrelation function

b) The box car function

c) The Fourier Transform of the box car function

mum entropy method (AR/MEM) is presented with some estimation algorithms and methods for checking the adequacy of the spectral estimate. The estimation of the spectrum is based upon linear estimates of some intermediate parameters.

The autoregressive-moving average (ARMA) method, which is also called "rational method", is discussed in Chapter 3. The optimal methods which result in a nonlinear estimation problem are first presented and a class of suboptimal methods which are recently proposed are then discussed.

Chapter 4 describes the implementation of the above methods in an interactive minicomputer-based program.

Chapter 5 discusses the results of estimating the spectra of signals recorded from the small intestine and gives an evaluation of the methods investigated here.

Conclusions of our findings and suggestions for future research are presented in Chapter 6.

CHAPTER 2

FFT and AR/MEM Spectral Estimation

2.1 Standard Autocorrelation Method

The procedure described by Blackman and Tukey in [9] for estimating the spectrum is based upon numerical evaluation of the original Wiener definition, and the compensation for the effect of truncation by using a specially designed autocorrelation window to replace the box car function. This windowing approach is an empirical technique to improve the performance of the estimation method. The performance is usually evaluated on the basis of two main factors [10]:

i) Leakage

Considering a sine wave of frequency (f_o), the true spectrum and autocorrelation are shown in fig. 2.1 and given by:

$$S(f) = A^2/4\ [\delta(f-f_o) + \delta(f+f_o)] \qquad (2.1)$$

and

$$r(nT) = A/2\ \cos\ (4\pi f_o nT) \qquad (2.2)$$

If the autocorrelation is multiplied by the rectangular window in fig. 1.b, and the DFT is evaluated for the product, the resulting spectrum S'(f) will be equal to the convolution of the rectangular window spec-

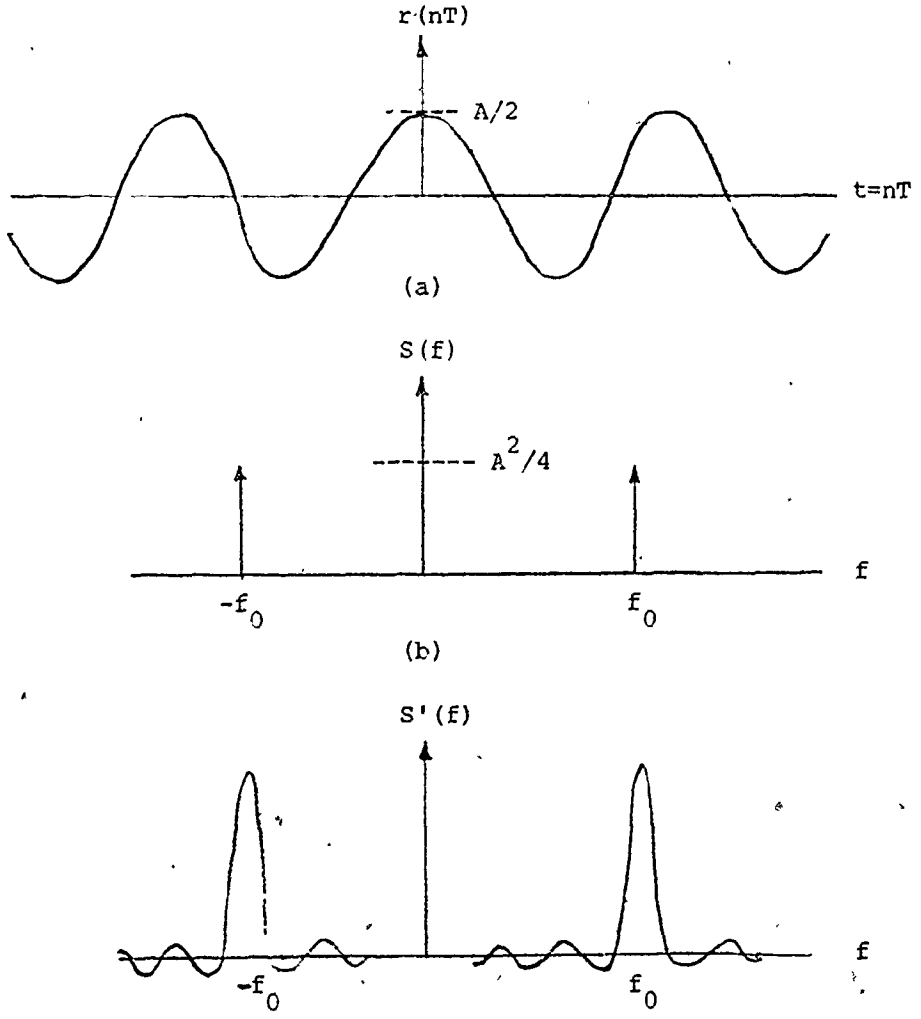


Fig 2.1 a) The autocorrelation function.

b) The true spectrum.

c) The Fourier Transform of the truncated autocorrelation function.

trum W(f) and the two impulses at f_o and $-f_o$. This is shown in fig. 2.1(c) where the impulse is converted to the (sin x)/x function and the amplitude is reduced according to the data length. The rest of the power seems to leak into the side lobes. The decrease in the main lobe amplitude and hence the amount of leaking power will be less if the length of the rectangular window is increased. If the length of the data approaches ∞ the original impulses will be restored.

ii) Resolution

Another important feature is the width of the main lobe. This is called the "resolution band width" or simply resolution. If there was another sine wave with frequency $(f_o+\Delta f)$ added to the original sine wave, then the spectrum of the summation will be the composition of two (sin x)/x functions with possible overlap between the two main lobes if Δf is less than 1/LT which may result in only one big lobe. Hence, estimating the spectrum by direct DFT of the autocorrelations up to lag L does not resolve frequencies less than 1/LT apart. In case of signals composed of several sinusoidal components, a spectrum with good resolution should show sharp peaks at the sinusoidal frequencies. A quantitative measure for the resolution (ξ) has been suggested by Markel and Gray [11] and is given by:

$$\frac{\exp\left[1/2\pi \int_{-\pi}^{\pi} \ln S(w)\, dw\right]}{1/2\,\pi \int_{-\pi}^{\pi} S(w)\, dw} \qquad (2.4)$$

This is, in fact, equal to the ratio between the geometric and the

arithmetic means of the spectrum. Its magnitude is between 0 and 1, and the spectrum peaks when it approaches zero.

In statistical terms, the estimate is judged according to the following statistical properties [12]:

1. Bias: The estimate $\hat{S}(w)$ is unbiased if

$$E[\hat{S}(w)] = S(w) \tag{2.5}$$

2. Variance: $\mathrm{Var}[\hat{S}(w)] = E[\{\hat{S}(w) - S(w)\}^2]$ (2.6)

3. Consistency: The estimate is consistent if its variance approaches zero as the number of points approaches infinity.

A theoretical bound to the variance of unbiased estimates is given by:

$$\sigma_s^2 \geq \frac{1}{NE\left\{\left[\frac{\partial \ln f(x,t)}{\partial t}\right]^2\right\}} \tag{2.7}$$

This is known as Cramer-Rao inequality [12], where N is the number of samples and f(x,t) is the first order statistic.

2.2 The Periodogram

The possibility of estimating the power spectral density, without calculating the autocorrelation function is discussed in (Papoulis 1965) for the continuous time case. The estimate $S_p(f)$ is obtained from a signal of length 2p as follows:

$$S_p(f) = \frac{1}{2p} \left| \int_{-p}^{p} x(t)\, e^{-j2\pi ft}\, dt \right|^2 \qquad (2.8)$$

The expected value of this estimate is shown to approach the true spectrum $S(f)$ as $p \to \infty$ iff:

$$\int_{-\infty}^{\infty} \left| \tau\, r(\tau) \right| d\tau < \infty \qquad (2.9)$$

This gives the necessary condition for obtaining an assymptotically unbiased estimate of the spectrum. In practical estimation procedure, however, the signal length is always finite; hence, the estimate is always biased.

The average squared magnitude of the N-point discrete Fourier transform of the series $x(kT)$ is called the periodogram $P(f)$. This is given by:

$$P(f) = \frac{1}{N} \left| \sum_{k=0}^{N} x(kT)\, e^{-j2\pi fkT} \right|^2 \qquad (2.10)$$

The statistical properties of the periodogram are analyzed in [13] under the assumption of Gaussianity. It is shown that the variance of the periodogram is given by:

$$\mathrm{Var}[P(f)] = S^2(f) \left\{ 1 + \left[\frac{\sin(wN)}{N \sin w} \right]^2 \right\} \qquad (2.11)$$

This means that the variance of the periodogram is proportional to and larger than the square of the true spectrum. Moreover, as N is increased to ∞, the variance does not become zero which means that the periodogram is not a consistent estimator of the power spectral density.

2.3 Tapering and Windowing

The effect of truncating the original series before transformation is to incorporate sharp rising and falling edges at the beginning and end of the data. This introduces high frequency components in the signal spectrum which were not already there. The seemingly reasonable way for treating this is to round off these sharp edges by a suitable data window. This is known as data tapering. This approach, however, is not welcomed by statisticans [14]. The statistical reasoning behind windowing the autocorrelations is to multiply the autocorrelation at a certain lag by a weight which is proportional to its variance at this lag. The increase of the variance at higher lags is evident, since the number of data points that contribute to the autocorrelation estimate becomes small. This is not the case with the data since the signal is assumed stationary and hence all data samples, presumably, have the same variance. Although the windowing technique is said to be an empirical one, yet there are some factors which govern the design of the window. The object is always to improve the statistical properties, for example, eliminate the bias or reduce the variance. For details of window design, see Ref. [9, 10].

2.4 Fast Estimation of the Autocorrelation Function

In discrete time, the autocorrelation is estimated by the following mean lag products:

$$r(k) = \frac{1}{N-K} \sum_{n=1}^{N-K} x(n)\ x(n+k) \qquad (2.12)$$

However, it can also be obtained by the inverse Fourier transform of the periodogram, thus making use of the FFT. To avoid a time domain aliasing (also called wrap around) at high lags, we have to satisfy the condition:

$$\frac{1}{\text{frequency spacing}} > 2 \text{ maximum lag} \qquad (2.13)$$

which can be considered as a frequency sampling requirement dual to Nyquist rate in the time domain. The procedure for obtaining N-lag autocorrelations from N-sample time series x(n) is as follows [13]:

1. Augment the sequence x(n) with N zeros to obtain a new sequence x'(n) with 2N samples.
2. Compute the 2N-point FFT X(k), k = 0, 1, 2, ... 2N-1
3. Compute the raw periodogram $P(k) = 1/N \mid X(k) \mid^2$, k = 0, 1, ... 2N-1
4. Compute the inverse FFT of P(k) and multiply the first N values by N/(N-k) to obtain the correct divisor. The autocorrelation will then be given by:

$$r(k) = \frac{N}{N-k} F^{-1} [P(k)] , \quad k = 0, 1, \ldots, N-1 \qquad (2.14)$$

2.5 Maximum Entropy Spectral Estimation

In the previous methods, the signal and the autocorrelations were assumed to be zero beyond the window edges, which is incorrect. The length of data is often governed by the computational feasibility.

The concept of maximum entropy has been introduced by Burg in 1967 [15]. Burg derived an expression for the spectrum which corresponds to a certain number of lags L and assumes the most unpredictable values to the unknown lags. This is outlined below. (See [16] for details).

In terms of the spectral density, the entropy H of a Gaussian process is expressed as:

$$H = \frac{1}{4f_{max}} \int_{-f_{max}}^{f_{max}} \log S(f)\, df \tag{2.15}$$

In terms of the autocorrelations, this becomes:

$$H = \frac{1}{4f_{max}} \int_{-2f_{max}}^{2f_{max}} \log \left[\sum_{k=-\infty}^{\infty} r(k)\, e^{-j2\pi fkT} \right] df \tag{2.16}$$

The object now is to maximize H w.r.t. the unknown lags (i.e. for $|k| > L$) under the constraint that S(f) is consistent with the known values for $|k| < L$). Hence,

$$\frac{\partial H}{\partial r(k)} = 0 \qquad |k| > L \tag{2.17}$$

This gives:

$$\int_{-f_{max}}^{f_{max}} \frac{e^{-j2\pi fkT}}{\hat{S}(f)}\, df = 0, \quad |k| > L \tag{2.18}$$

$1/S(f)$ can then be expanded in an L-term series, as follows:

$$1/S(f) = \sum_{k=-L}^{L} c_k\, e^{-j2\pi fkT} \tag{2.19}$$

where $c_k = c_{-k}$ to ensure that S(f) is real. The coefficients $\{c_k\}$ are determined such that the known lags of r(k) are consistent with the spectral estimate, i.e.

$$r(k) = \int_{-f_{max}}^{f_{max}} \hat{S}(f)\, e^{j2\pi fkT}\, df \quad |k| < L \tag{2.20}$$

The solution for this gives the following estimate:

$$\hat{S}(f) = \frac{P_L}{f_{max} \left| 1 + \sum_{k=1}^{L} a_k\, e^{-j2\pi fkT} \right|^2} \tag{2.21}$$

where P_L is a constant and $\{a_k\}$ can be estimated from the autocorrelations by the following set of equations:

$$\begin{bmatrix} r(0) & r(1) & -- & r(L) \\ r(1) & r(0) & -- & -- \\ -- & -- & -- & -- \\ -- & -- & -- & -- \\ r(L) & -- & -- & r(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ -- \\ -- \\ a_L \end{bmatrix} = \begin{bmatrix} P_L \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{2.22}$$

This will be shown next to be the same estimate as that obtained from an autoregressive representation. Burg also suggested another method for estimating the coefficients directly from the data without calculating the autocorrelation function [17].

2.6 Autoregressive Spectral Estimation

A stationary time series is said to be autoregressive if it

can be modelled by a difference equation (regression model) in which x(n) is regressed on its own past values [18] plus a white noise term. This is written as:

$$x(n) = a_1 x(n-1) + a_2 x(n-2) + \ldots + a_L x(n-L) + w(n) \qquad (2.23)$$

where w(n) is a white noise sequence, the coefficients $\{a_i\}$ are the regression parameters and L is the model order. The autoregressive modelling was shown by Akaike [19] to be useful in estimating the power spectrum of x(n). Equation (2.23) can be rewritten as:

$$x(n) - a_1 x(n-1) \ldots - a_L x(n-L) = w(n) \qquad (2.24)$$

Taking the z transform of the above equation, we get:

$$A(z)\ X(z) = W(z) \qquad (2.25)$$

where $A(z) = (1 - a_1 z - a_2 z^2 \ldots a_L z^L)$.

The power spectrum density can then be obtained as:

$$S(z) = X(z)\ X^*(z) = S_w(z) \,/\, A(z)\ A(z^{-1}) \qquad (2.26)$$

where * denotes complex conjugation and $S_w(z)$ is the white noise spectrum which is constant with frequency and equal to the noise variance (σ_w^2).

The relation between the maximum entropy and the autoregressive spectral estimates was found by Van den Bos [20]. In fact, the AR spectrum is equivalent to the ME spectrum without entropy consideration. A difference may arise from the way the AR parameters are estimated.

2.6.1 Yule-Walker Solution for the AR Parameters

The classical method for estimating the autoregressive parameters is known as Yule-Walker solution [18] and this can be obtained by multiplying equation (2.23) by x(n-k) and then taking the expectation of both sides. We get:

$$E[x(n)\ x(n-k)] = a_1 E[x(n-1)\ x(n-k)] + \ldots + a_L E[x(n-L)x\ (n-k)] + E[w(n)\ x(n-k)] \qquad (2.27)$$

Hence,

$$r(k) = a_1 r(k-1) + \ldots + a_L r(k-L) \qquad (2.28)$$

where we made use of the fact that x(n-k) is uncorrelated with the future noise sample w(n). If equation (2.28) is written in a matrix form for k = 1, 2, ... L, we obtain:

$$\begin{bmatrix} r(1) \\ r(2) \\ -- \\ -- \\ r(L) \end{bmatrix} = \begin{bmatrix} r(0) & r(1) & -- & r(L-1) \\ r(1) & r(0) & -- & r(L-2) \\ -- & -- & & \\ -- & -- & & \\ r(L-1) & -- & -- & r(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ -- \\ -- \\ a_L \end{bmatrix} \qquad (2.29)$$

and it should be noted that in finding the entries of the autocorrelation matrix on the right-hand side, we made use of the symmetry of the autocorrelation function of stationary processes, [i.e. r(k) = r(-k)].

The solution for the parameters a_i can be obtained in vector notation as follows:

$$\underline{a} = R^{-1} \underline{r} \tag{2.30}$$

where $\underline{a}^T = [a_1\ a_2\ \ldots\ a_L]$, $\underline{r}^T = [r_1\ r_2\ \ldots\ r_L]$, R is the autocorrelation matrix and T denotes transposition. The autocorrelation entries in this matrix equation can be estimated by evaluating the mean lag products in equation (2.12) or using the fast method. The problems that are involved in obtaining the solution are:

1. Finding the order (L).
2. Inverting the matrix R.
3. Checking for the adequacy of the model.

2.6.2 Least Squares Estimation of AR Parameters

The estimation may be based on a least squares solution to equation (2.23). The reasoning in this is to try to make w(n) as close as possible to its expected value which is zero by minimizing $\Sigma\ w^2(n)$ for all such equations that can be formed from the available data. (This is also called linear least square prediction [21]). In vector notation:

$$\begin{bmatrix} x_{L+1} \\ x_{L+2} \\ -- \\ -- \\ x_N \end{bmatrix} = \begin{bmatrix} x_L & x_{L-1} & -- & x_1 \\ x_{L+1} & x_L & -- & x_2 \\ -- & -- & & \\ -- & -- & & \\ x_{N-1} & x_{N-2} & & x_{N-L} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ -- \\ a_L \end{bmatrix} + \begin{bmatrix} w_1 \\ w_2 \\ -- \\ -- \\ w_N \end{bmatrix} \tag{2.31}$$

or

$$\underline{x}_{N,1} = X_{N,L}\,\underline{a}_{L,1} + \underline{w}_{N,1} \tag{2.32}$$

The quantity to be minimized is $\underline{w}\,\underline{w}^T$ which is given by:

$$\underline{w}^T\underline{w} = (\underline{x}_{N,1} - X_{N,L}\,\underline{a}_{L,1})^T (\underline{x}_{N,1} - X_{N,L}\,\underline{a}_{L,1}) \tag{2.33}$$

Taking the derivative w.r.t. the parameters $\underline{a}$, then equating this to zero we obtain the famous pseudo inverse formula for the solution that follows [22]:

$$\underline{a}_{L,1} = (X_{N,L}^T\,X_{N,L})^{-1}\,X_{N,L}^T\,\underline{x}_{N,L} \tag{2.34}$$

A close look at this shows that it is exactly the same as Yule-Walker solution if the autocorrelations are estimated by the mean lag products. We assume that the inverse of the matrix exists in both cases. This means that the matrix is of full rank, which should be true if the order is correct.

2.6.3 Recursive Estimation of AR Parameters

The problem with the above solutions is the need for matrix inversion which is computationally very time consuming when the order is high. Durbin [23] obtained a solution which is recursive in order, i.e. the parameters of a certain model are estimated from those of a lower order model. Hence, the solution for any order can be obtained by starting from the first order which does not need matrix inversion.

This simplification is due to the "Toeplitz" property of R, i.e. its symmetry with respect to both diagonals. The solution can be obtained as follows: Consider the simple case of obtaining the parameters of an AR(3) model from those of an AR(2) model. Hence,

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} r_0 & r_1 \\ r_1 & r_0 \end{bmatrix} \begin{bmatrix} a_1(2) \\ a_2(2) \end{bmatrix} = R_2 \underline{a}(2) \qquad (2.35)$$

and

$$\begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} = \begin{bmatrix} r_0 & r_1 & r_2 \\ r_1 & r_0 & r_1 \\ r_2 & r_1 & r_0 \end{bmatrix} \begin{bmatrix} a_1(3) \\ a_2(3) \\ a_3(3) \end{bmatrix} = R_3 \underline{a}(3) \qquad (2.36)$$

From the first two rows in equation (2.36), $a_1(3)$ and $a_2(3)$ can be obtained as:

$$\begin{bmatrix} a_1(3) \\ a_2(3) \end{bmatrix} = R_2^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} - R_2^{-1} \begin{bmatrix} r_2 \\ r_1 \end{bmatrix} a_3(3) \qquad (2.37)$$

Hence

$$\begin{bmatrix} a_1(3) \\ a_2(3) \end{bmatrix} = \begin{bmatrix} a_1(2) \\ a_2(2) \end{bmatrix} - a_3(3) \begin{bmatrix} a_2(2) \\ a_1(2) \end{bmatrix} \qquad (2.38)$$

Hence, $a_1(3) = a_1(2) - a_2(2)\, a_3(3)$ (2.39)

and $a_2(3) = a_2(2) - a_1(2)\, a_3(3)$ (2.40)

$a_3(3)$ can be obtained from the third row of equation (2.36) and equations (2.39) and (2.40) as follows:

$$a_3(3) = \frac{r_3 - a_1(2)\, r_2 - a_2(2)\, r_1}{r_0 - a_1(2)\, r_1 - a_2(2)\, r_2} \tag{2.41}$$

Equations (2.39), (2.40) and (2.41) completely identify the parameters of the AR(3) model from those of the AR(2). The general recursive formulas for orders p and p+1 are:

$$a_j(p+1) = a_j(p) - a_{p+1}(p+1)\, a_{p-j+1}(p) \tag{2.42}$$

and

$$a_{p+1}(p+1) = \frac{r_{p+1} - \sum_{j=1}^{p} a_j(p) r_{p+1-j}}{r_o - \sum_{j=1}^{p} a_j(p)\, r_j} \tag{2.43}$$

2.6.4 Hypothesis Testing as a Diagnostic Check

The autoregressive modelling is built around a hypothesis which is the whiteness of the sequence w(n). Since this has been exploited in deriving the solution, then it should be checked if the whiteness feature has been achieved or not. The deviation from whiteness is taken as a measure for the lack of fit. The most peculiar thing about a white noise sequence is that it possesses an impulsive autocorrelation which is zero everywhere except at zero lag. It can be shown that under the assumption of Gaussian distribution of these autocorrelations $\{r_w(k)\}$, the quantity $Q = N \sum_{k=1}^{M} r_w^2(k)$ has a chi-square distribution with (M-L) degrees of freedom [7]. Therefore, if the sequence is white noise, then this quantity should be as close as possible to zero. In other words, the probability that Q is larger than a certain value

should be less than a predetermined value. This is called the level of significance of the test. However, determining the level of significance is a problem in itself that needs a subjective judgement. This is usually taken as .05 for system identification purposes [18], but our experience with it showed that it is rather loose and it should be lessened if a better spectral resolution is desired.

2.6.5 Order Choice for AR Models

i) Akaike's Information Criterion (AIC)

Akaike devised a test [24] which is intended to eliminate the subjectivity encountered in model fitting. This is based on an information theoretic criterion which suggests that the best order is the one which minimizes the quantity:

$$AIC = N \ln \sigma_w^2 + 2L \tag{2.44}$$

Since most physical processes arise actually from a very high order then AIC is considered as a compromise between obtaining a small noise variance and the effort spent in fitting high order models. This is widely used in fitting parametric statistical models and its theoretical basis is beyond the scope of this thesis.

ii) Final Prediction Error (FPE)

This also has been suggested by Akaike [24] and it requires the best model to minimize the following quantity:

$$FPE = \frac{N + L + 1}{N - L - 1} \sigma_w^2 \tag{2.45}$$

It also gives a similar compromise since it results from multiplying σ_w^2 which decreases with the order by another quantity which increases. It can be seen that AIC and FPE are asymptotically equivalent, (i.e. as $N \to \infty$).

iii) Autocorrelation Matrix Determinant

From equation (2.29) it can be seen that r(k) can be obtained from the previous L lag values. Therefore, if the order is increased, from L to L+1, then each lag value in the L + 1st row can be obtained by the same linear combination from the corresponding lags in the upper L rows. In other words, the L + 1st row is linearly dependent on the other L rows. Thus, the matrix contains only L linearly independent vectors if the correct order is (L). Then we should expect the matrix to become singular (of zero determinant) if the dimension is increased beyond the true order. However, because the autocorrelation entries in the matrix are not exact, then we do not expect the determinant to be exactly zero but possibly very small. Accordingly, we can consider the ratio of the determinants (DRT) obtained from two successive orders as an indication to how close we are from the true order.

$$DRT(L) = \frac{\text{determinant } (L+1)}{\text{determinant } (L)} \qquad (2.46)$$

The above ratio should be peaky at the correct order.

CHAPTER 3

RATIONAL SPECTRAL ESTIMATION

3.1 Parametric Spectral Representation

It has been shown by Tretter and Steiglitz [25] that the power spectral density can be obtained from rational spectral models, which translates the spectrum estimation task to a system identification problem. The problem can be attacked from the point of view of different subjects. We will consider a linear prediction approach through Wold's decomposition principle [26] which states that a stochastic sequence $x(n)$ can be decomposed into a deterministic component $x_1(n)$ and a random component $w(t)$. The deterministic part of rhythmic signals will be mainly equal to the summation of sine waves having the rhythmic frequencies. However, a general representation to $x_1(n)$ could be put in a difference equation form. If the random part $w(n)$ is white noise, we obtain the same representation as the autoregressive model, that is,

$$x(n) = a_1x(n-1) + a_2x(n-2) + \ldots a_Lx(n-L) + w(n)$$

A more general representation can be obtained by considering $w(n)$ as coloured noise which can be generated by a weighted average (called moving average (MA)) of a white noise sequence $e(n)$. The difference equation model then becomes a mixed autoregressive-moving average

(ARMA) model of the form:

$$x(n) = \sum_{i=1}^{L} a_i \, x(n-i) + \sum_{i=0}^{M} b_i \, e(n-i) \qquad (3.1)$$

The rational spectral model can be obtained from the z-transform (or Fourier transform with $z = e^{j2\pi f}$) of the above equation. This gives:

$$A(z)\, X(z) = B(z)\, E(z) \qquad (3.2)$$

where $A(z) = 1 - a_1 z - a_2 z^2 - \ldots a_L z^L$

and $B(z) = b_0 + b_1 z + b_2 z^2 \ldots b_M z^M$

The power spectral density can then be obtained as follows:

$$S(z) = X(z)\, X^*(z) = \frac{B(z)B^*(z)}{A(z)A^*(z)}\, S_e(z) \qquad (3.3)$$

where $S_e(z)$ is the spectrum of the white noise e(n), which is constant with frequency and equal to the noise variance σ_e^2. In this model, the sequence x(n) is described in terms of a group of parameters $\{a_i\}$ and $\{b_i\}$ which are the coefficients of the difference equation (3.1). This is an alternative way for characterizing a stationary process instead of the autocovariance or the spectral density functions. This implies that if the model is adequate, then all the information that can be extracted from the signal is retained in the parameters, and the spectral density can be restored from the parameters as in equation (3.3).

The problem of the power spectral density estimation now reduces

to that of parameter estimation. The methods of estimating the ARMA parameters can be classified into optimal and suboptimal methods, the optimality being in the maximum likelihood or the least square sense. (for details of optimal estimators, see Ref. [12, 18]). The suboptimal methods have been recently proposed [27 - 29] to reduce the computational effort of the optimal ones.

3.2 Nonlinear Least Square Estimation

The estimation in this case is based upon the minimization of the squared sum of e(n) with respect to the parameters. From equation (3.2) we have (with $a_0=1$),

$$e(n) = a_0 x(n) - \sum_{i=1}^{L} a_i x(n-i) - \sum_{i=1}^{M} b_i e(n-i) \qquad (3.4)$$

From the available N data points, we can write N-L such equations assuming $L \geq M$ (We will show later that a noisy periodic signal can be represented by an ARMA model with L = M).

Hence:

$$\begin{bmatrix} e_{L+1} \\ e_{L+2} \\ -- \\ -- \\ \\ e_N \end{bmatrix} = \begin{bmatrix} x_{L+1} & -x_L & -- & -x_1 & -e_L & -e_{L-M+1} \\ x_{L+2} & -x_{L+1} & -- & -x_2 & -e_{L+1} & -e_{L-M+2} \\ -- & -- & & -- & -- & -- \\ -- & -- & & -- & -- & -- \\ & & & & & \\ x_N & -x_{N-1} & & -x_{N-L} & -e_{N-1} & -e_{N-M} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ -- \\ a_L \\ b_1 \\ -- \\ b_M \end{bmatrix} \qquad (3.5)$$

We will denote the parameter vector by $\underline{p}$, the noise vector by $\underline{e}$ and the data matrix by D. Then equation (3.5) can be rewritten as follows:

$$\underline{e} = D\,\underline{P} \tag{3.6}$$

The sum of squares of e(n) is equal to $\underline{e}^T\underline{e}$ and given by:

$$\underline{e}^T\underline{e} = \underline{P}^T D^T D\,\underline{P} \tag{3.7}$$

It can be seen that $\underline{e}^T\underline{e}$ is a highly nonlinear function in the parameters because of the recursive nature of the model equation. Hence, an attempt to minimize Σe^2 by equating the first derivative to zero does not yield a closed form solution. In this case, numerical minimization has to be performed.

3.2.1 Åström's Procedure

We will derive the procedure described by Åström [30] for solving the above problem, which is based on Newton's method [31] for optimizing the scalar function $u = \Sigma e^2$ with respect to the parameters of the vector $\underline{p}$, with the objective of obtaining the values of the parameters which yield the minimum. The basis of the method is to expand the function in a Taylor series around some chosen point $\underline{p}_o$ in a convex region of the function as follows:

$$u(\underline{p}) = u(\underline{p}_o) + (\underline{p}-\underline{p}_o)^T\,\underline{\nabla}\,u(\underline{p}_o) + \tfrac{1}{2}\,(\underline{p}-\underline{p}_o)^T\,H(\underline{p}_o)\,(\underline{p}-\underline{p}_o) + \ldots$$

where $\underline{\nabla}$ and H denote the gradient vector and the Hessian matrix of u, respectively.

Linearizing the gradient of the function about $\underline{p}_o$ (i.e. considering only the first two terms in the expansion), we obtain the condition of a minimum (or a maximum) as follows:

$$\underline{\nabla}u(\underline{p}) = \underline{\nabla}u(\underline{p}_o) + H(\underline{p}_o)\ (\underline{p}-\underline{p}_o) = \underline{0} \tag{3.8}$$

where $\underline{0}$ is the zero vector. Then:

$$\underline{p} - \underline{p}_o = -H^{-1}(\underline{p}_o)\ \underline{\nabla}u(\underline{p}_o) \tag{3.9}$$

The implication of the above is that if the function is quadratic (i.e. could be exactly represented by the first three terms of a Taylor's series expansion, since all the higher order derivatives are zero), then a change in $\underline{p}$ as given by equation (3.9) should lead us to the minimum. However, because the nonlinearity may be higher than quadratic (according to the model order) the above process has to be repeated in order to reach a minimum.

A necessary condition for convergence to a minimum is the positive definitness of the Hessian matrix, which means the convexity of the function at $\underline{p}_o$. Otherwise, a maximum will be obtained. The gradient can be evaluated from the available data as follows:

$$\underline{\nabla}u(\underline{p}) = \frac{\partial}{\partial \underline{p}} \sum_{n=L+1}^{N} e^2(n)$$

$$= \sum_{n=L+1}^{N} 2e(n)\ \frac{\partial e(n)}{\partial \underline{p}} \tag{3.10}$$

The Hessian matrix can be obtained by differentiating equation (3.10).

This gives:

$$H = \underline{\nabla}\,(\underline{\nabla} u)^T$$

$$= 2 \sum_{n=L+1}^{N} \frac{\partial e(n)}{\partial \underline{p}} \left(\frac{\partial e(n)}{\partial \underline{p}}\right)^T + 2 \sum_{n=L+1}^{N} e(n) \frac{\partial}{\partial \underline{p}} \left(\frac{\partial e(n)}{\partial \underline{p}}\right)^T \qquad (3.11)$$

The derivatives of e(n) with respect to the parameters in $\underline{p}$ can be obtained from equation (3.4) as follows:

$$\frac{\partial e(n)}{\partial a_i} = -x(n-i)$$

$$\frac{\partial e(n)}{\partial b_i} = -e(n-i) - \sum_{i=1}^{M} b_i \frac{\partial e(n-i)}{\partial b_i} \quad n = L+1, \ldots, N \qquad (3.12)$$

$$\frac{\partial e(n)}{\partial a_i \partial b_i} = \frac{\partial^2 e(n)}{\partial a_i^2} = 0$$

and

$$\frac{\partial^2 e(n)}{\partial b_i^2} = \frac{\partial e(n-i)}{\partial b_i} - \sum_{i=1}^{M} b_i \frac{\partial^2 e(n-i)}{\partial b_i^2} \quad n = L+1, \ldots, N \qquad (3.13)$$

The procedure can then be summarized in the following steps:

1. Assume an initial parameter vector $\underline{p}_o$
2. Evaluate e(n), n = L+1, ..., N from equation (3.5)

3.a) Evaluate the derivatives from equations (3.12) and (3.13)

b) Evaluate the gradient vector from equation (3.10)

c) Evaluate the Hessian matrix from equation (3.11)

d) Evaluate the new $\underline{p}$ from equation (3.9)

e) Calculate the new noise vector $\underline{e}$ from equation (3.5) and the function $u = \underline{e}^T \underline{e}$.

4. If $\underline{e}^T\underline{e}$ is less than a certain tolerance, stop. Otherwise, go back to step 2.

3.2.2 Steepest Descent and Marquardt's Compromise

The above optimization procedure guarantees convergence to the local minimum. However, it suffers from the inclusion of a tremendous amount of computation which consumes a very long computer time, especially with minicomputers which are relatively slow. The calculation of the Hessian matrix and its inverse constitutes the main effort in the problem. This can be, however, dispensed away with completely and replaced by a unity matrix resulting in the well-known steepest descent algorithm [31]. The change in the parameter vector in this case will be given by:

$$\underline{p} - \underline{p}_o = -I\underline{\nabla}u(\underline{p}_o) \qquad (3.14)$$

A further control to the step change can be introduced by multiplying by a constant (λ) which may be varied from one step to the other.

Another way to avoid calculating the Hessian matrix and obtaining a step change which is more complicated than (3.14) but utilizes only the first derivatives is known as Marquardt's algorithm [18]. This can be obtained by successive linearization of the error $\underline{e}$ around $\underline{p}_o$ as follows:

$$\underline{e} = \underline{e}_o + J\cdot(\underline{p}-\underline{p}_o) \qquad (3.15)$$

where J is the Jacobian matrix of $\underline{e}$ w.r.t. $\underline{p}$ and $\underline{e}_o$ is the noise vector

at $\underline{p}_o$. The step change can be obtained by linear least squares from the above equation. This gives,

$$(\underline{p}-\underline{p}_o) = -(J^TJ)^{-1} J^T \underline{e}_o \qquad (3.16)$$

Comparing the above result with that in equation (3.11), it can be seen that the Hessian matrix has been approximated with the first term only which is equal to J^TJ. A replacement for the effect of the second derivative on the step change can be obtained by adding a constant matrix (λI) to J^TJ to simulate the effect of the second term in equation (3.11) where λ can be changed from one step to the other. Since the second derivative is expected to decrease in magnitude as we get closer to the minimum, then λ should also be made smaller with every new step. This could be achieved by multiplying λ by a constant fraction before each step. The iteration formula can now be written as:

$$\underline{p}_k - \underline{p}_{k-1} = -(J_{k-1}^T J_{k-1} + \mu\lambda_{k-1}I)^{-1} J_{k-1}^T \underline{e}_{k-1} \qquad (3.17)$$

$$k = 1, 2, \ldots$$

where k is the iteration number, $\lambda_k = \mu\lambda_{k-1}$ and μ is a constant fraction.

3.3 Maximum Likelihood Estimation

This is based on Bayes' conditional probability theorem [12, 18]. Given a set of data $\underline{x}$, then by assuming a certain prior distribution of the parameter values $P(\underline{p})$, the posterior distribution after knowledge of the data will be $P(\underline{p}/\underline{x})$. The likelihood function $L(\underline{p}/\underline{x})$ is the

probability that a particular set of parameters have given rise to the available data. From Bayes' theorem, we have:

$$P(\underline{p}/\underline{x}) = P(\underline{p})\, L(\underline{p}/\underline{x}) \tag{3.18}$$

The best estimates of the parameters are those which are most likely to produce the available data which can be obtained by maximizing the likelihood function. If the a priori distribution is taken to be uniform, the likelihood function will be proportional to the posterior distribution. It is shown in [18] that under the assumption of Normality of the data and noise samples, the likelihood function is given by:

$$L(\underline{p}/\underline{x}) = (2\pi\sigma_e^2)^{-N/2} \exp-(\underline{e}^T\underline{e}/2\sigma_e^2) \tag{3.19}$$

This can be seen to be equal to the joint density function of the noise samples assuming they have identical Gaussian distribution and are truly independent by the virtue of whiteness. Maximization of the above function is equivalent to minimizing the value of the negative exponent $\underline{e}^T\underline{e}/2\sigma_e^2$. This leads us back to the minimization of the sum of the squares of the noise samples and the problem can be solved exactly as the least squares estimation.

3.4 Suboptimal Estimation of the Rational Model Parameters

The nonlinearity of the error function to be minimized in the rational models makes use of such spectral models difficult when the number of parameters is large. This is due to the inclusion of the moving average part in the model. However, it has been shown by

Kinkel et al. [27] that the spectral density can be obtained from the ARMA model without prior estimation of the moving average parameters making use of the property of the uniqueness of the autoregressive parameters of a mixed ARMA model [32]. The uniqueness of the autoregressive representation and the autoregressive part of a mixed ARMA model are direct consequences of Wold's decomposition theorem [26] mentioned earlier in this chapter. This can be felt by recalling that the autoregressive part represents the deterministic behaviour of the signal. The same method was adopted independently by Kaveh [28] where he demonstrated its good resolution property. The essence of this method is to consider the ARMA process as arising from an all-pole (autoregressive) filter driven by a moving average process. Another method has been suggested by Cadzow [29] where he exploited the symmetry of the spectral density and the autocorrelation functions of a stationary process. This was referred to in an earlier survey by Morf et al. [33]. The two methods are treated next.

3.4.1 Estimation of the AR Parameters of a Mixed ARMA Process

Consider the zero mean ARMA process x(n) which is modelled by:

$$\sum_{i=0}^{L} a_i x(n-i) = \sum_{i=0}^{M} b_i e(n-i) \qquad (3.20)$$

Multiplying both sides by x(n-k) then taking expectations, we get:

$$\sum_{i=0}^{L} a_i E[x(n-k)\, x(n-i)] = \sum_{i=0}^{M} b_i E[x(n-k)\, e(n-i)] \qquad (3.21)$$

The expectation of the form $E[x(n-k)\ e(n-i)]$ is equal to the cross-correlation $r_{xe}(k-l)$, where

$$\begin{aligned} r_{xe}(k-i) &= 0 \qquad k-i \geq M \\ &\neq 0 \qquad k-i < M \end{aligned} \tag{3.22}$$

This is a result of the structure of the ARMA model where $x(n)$ is produced by recursive filtering of its own L past samples and the noise M past samples. In other words, because of the causality of the ARMA model, then $E[x(n)\ e(m)] = 0$, so long as $m > n$.

In order to solve for the L parameters $\{a_i,\ i=1,\ 2,\ \ldots\ L\}$, we have to have at least (L) independent equations of the form (3.2.1) with zero on the right-hand side. This could be achieved by writing equation (3.21) with $k = M+1,\ M+2,\ \ldots\ M+L$, where we guarantee that all the crosscorrelations within the summation will vanish. Then we obtain:

$$\begin{aligned} &r_{M+1} + a_1 r_M + a_2 r_{M-1} \ \ldots\ + a_L r_{M-L+1} = 0 \\ &r_{M+2} + a_1 r_{M+1} + a_2 r_M \ \ldots\ + a_L r_{M-L+2} = 0 \\ &\quad \ldots \qquad \ldots \qquad \ldots \qquad \ldots \\ &r_{M+L} + a_1 r_{M+L-1} + a_2 r_{M+L-2} \ \ldots\ + a_L r_M = 0 \end{aligned} \tag{3.23}$$

This is similar to the Yule-Walker solution of the pure autoregressive models. However, because of the errors in estimating the autocorrelations, the right-hand side is not exactly zero. A least square solution can be obtained (by minimizing the sum of the squares of the

right-hand side errors) from K equations as follows:

$$\underline{a} = -(R^T R)^{-1} R^T \underline{r} \tag{3.24}$$

where $\underline{r}^T = [r_{M+1} \quad r_{M+2} \quad -- \quad r_{M+K}]$ (3.25)

and

$$R = \begin{bmatrix} r_M & r_{M-1} & -- & r_{M-L+1} \\ r_{M+1} & r_M & -- & r_{M-L+2} \\ -- & -- & -- & -- \\ -- & -- & -- & -- \\ r_{M+K-1} & -- & -- & r_{M+K-L} \end{bmatrix} \tag{3.26}$$

3.4.2 The Cascade Representation Method

The z transform of the time domain model equation gives:

$$X(z) = E(z)\, B(z) \cdot \frac{1}{A(z)} \tag{3.27}$$

This can be regarded as an all pole filter $[1/A(z)]$ driven by the moving average process represented by $E(z)\, B(z)$. Then the spectrum of $X(z)$ is equal to the moving average spectrum $[\sigma_e^2\, B(z)\, B^*(z)]$ multiplied by the all pole filter amplitude characteristic which is equal to $1/[A(z) \cdot A^*(z)]$. The moving average spectrum is given by:

$$S_b(z) = \sum_{k=-M}^{M} c_k\, z^k \tag{3.28}$$

where the c_k's are the autocorrelations of the driving MA process. It should be noted that the above representation is exact since the autocorrelations of a MA process vanish for $k > M$, hence the infinite limits

of the summation are reduced to the value of M. The c_k's are essentially related to the AR parameters and the autocorrelations of the original process and can be obtained by equating the two exact expressions of the spectrum as follows:

$$S_x(z) = \sum_{k=-\infty}^{\infty} r_k z^k = \frac{\sum_{k=-M}^{M} c_k z^k}{A(z)\, A(z^{-1})} \tag{3.29}$$

Hence,

$$(1 + a_1 z + \ldots a_L z^L)\ (1 + a_1 z^{-1} + \ldots + a_L z^{-L})\ (\ldots + r_{-1} z^{-1} + r_o z^o + r_1 z + \ldots) = c_{-M} z^{-M} \ldots + c_{-1} z^{-1} + c_o z^o + c_1 z + \ldots + c_M z^M \tag{3.30}$$

Equating the coefficients of the equal powers of z on both sides, we get,

$$c_k = c_{-k} = \sum_{i=0}^{L} \sum_{j=0}^{L} a_i\, a_j\, r(k-i-j) \qquad k = 0, 1, \ldots, M \tag{3.31}$$

The spectrum can then be estimated by directly evaluating the expression on the right-hand side of equation (3.29).

3.4.3 The One-Sided Spectrum Method

The expression for the power spectral density in terms of the autocorrelation function can be rewritten as follows:

$$S_x(z) = \sum_{k=-\infty}^{\infty} r_k z^k$$

$$= (ro/2 + \sum_{k=-\infty}^{-1} r_k z^k) + (ro/2 + \sum_{k=1}^{\infty} r_k z^k) \qquad (3.32)$$

$$= \bar{S}(z) + S^{+}(z) \qquad (3.33)$$

where $\bar{S}(z)$ and $S^{+}(z)$ are complex conjugate quantities. This has the meaning of decomposing the real process x(n) with the positive real spectrum $S_x(z)$ into two complex processes with the complex spectra $\bar{S}(z)$ and $S^{+}(z)$. $S^{+}(z)$ is the spectrum of the process whose autocorrelation function $r^{+}(k)$ satisfies:

$$\begin{aligned} r^{+}(k) &= r(k) && k > 0 \\ &= r(0)/2 && k = 0 \\ &= 0 && k < 0 \end{aligned} \qquad (3.34)$$

Similarly, the other complex process has an autocorrelation function $\bar{r}(k)$ such that:

$$\begin{aligned} \bar{r}(k) &= 0 && k > 0 \\ &= r(0)/2 && k = 0 \\ &= r(k) && k < 0 \end{aligned} \qquad (3.35)$$

To estimate $S_x(z)$ it is enough to estimate $S^{+}(z)$, then $S_x(z)$ can be obtained as:

$$S_x(z) = 2 \text{ Real } S^{+}(z) \qquad (3.36)$$

Accordingly, if $S_x(z)$ is to be modelled by the rational form in equation (3.3), then $S^+(z)$ is assigned a rational model with the same poles as the roots of $A(z) = 0$. Hence,

$$S^+(z) = C(z)/A(z) \tag{3.37}$$

where $C(z) = c_o' + c_1'z + c_2'z^2 + \ldots + c_L'z^L$. The choice of the numerator order to be equal to (L) will be justfied later. The parameters $\{c_i\}$ can be estimated from the autocorrelations and the autoregressive parameters as follows:

$$S^+(z) = \sum_{k=0}^{\infty} r^+(k)\, z^k = C(z)/A(z) \tag{3.38}$$

Hence,

$$(1 + a_1z + \ldots + a_Lz^L)\ (r^+(0) + r^+(1)z + r^+(2)z^2 + \ldots\ )$$

$$= c_o' + c_1'z + \ldots + c_L'z^L \tag{3.39}$$

Equating the coefficients of equal powers of (z) on both sides, we get:

$$c_k' = r^+(k) + \sum_{i=1}^{L} a_i r^+(k-i) \tag{3.40}$$

3.4.4 Equivalence of the Two Methods

The two methods above prescribe a simple procedure for estimating the rational spectrum which can be summarized in the following three steps:

1. Estimating the autocorrelation function by the mean lag products.

2. Obtaining the autoregressive parameters via solving linear equations.

3. Estimating a set of auxiliary parameters $\{c_k\}$ or $\{c'_k\}$.

In the cascade method, these auxiliary parameters are none but the autocorrelations of the MA stage. With simple analytical manipulation, we will show that the one-sided spectrum method yields the same result as the cascade representation if the sequence $\{c'_k\}$ is made of length 2L.

The cascade representation estimate, written compactly as $\frac{N_1}{D}$, has a moving average spectrum (N_1) which is expressed as:

$$N_I = A(z)\ A(z^{-1}) \cdot S(z)$$

$$= (1 + a_1 z + .. a_L z^L)\ (1 + a_1 z^{-1} + a_L z^{-L}) \sum_{k=-\infty}^{\infty} r_k z^k \qquad (3.41)$$

Since the MA order is taken to be L, the autocorrelations of the original series will be considered up to lag 2L. Hence, in terms of cosine functions:

$$N_1 = |A(e^{j\omega T})|^2 \left[r_o + 2 \sum_{k=1}^{2L} r_k \cos \omega k T\right] \qquad (3.42)$$

The one-sided spectrum gives the estimate:

$$2 \text{ Real } S^+(z) = 2 \text{ Real } \frac{C(z)}{A(z)} = 2 \text{ Real } \frac{C(z)\ A(z^{-1})}{A(z)\ A(z^{-1})} = \frac{N_2}{D} \qquad (3.43)$$

Hence,

$$N_2 = A(z)\, A(z^{-1}) \cdot 2 \text{ Real } S^{+}(z)$$

$$= |A(z)|^2 \quad 2 \text{ Real } [\frac{r_o}{2} + r_1 z + r_2 z^2 + \ldots] \tag{3.44}$$

If the first $2L+1$ terms in the expansion of the one-sided spectrum are considered, then, in terms of cosine functions:

$$N_2 = |A(e^{j\omega T})|^2 \quad [r_o + 2 \sum_{k=1}^{2L} r_k \cos \omega kT] \tag{3.45}$$

which is the same expression obtained from the cascade representation. This requires the polynomial $C(z)$ to be of degree $2L$.

A convenient matrix formulation for calculating the $2L$ parameters $\{c'\}$ through the recurrence relation (3.40) can be deduced in terms of the autocorrelations $\{r(k)\}$ as follows:

$$C(z) = A(z) \quad S^{+}(z)$$

Hence,

$$(c_o + c_1 z + .. \; c_{2L} z^{2L}) = (1 + a_1 z + .. \; a_L z^L) \; (\frac{r_o}{2} + r_1 z + r_2 z^2 + \ldots)$$

$$\begin{bmatrix} c_o \\ c_1 \\ c_2 \\ .. \\ .. \\ c_L \\ c_{L+1} \\ .. \\ c_{2L} \end{bmatrix} = \begin{bmatrix} r_o/2 & 0 & & & & 0 \\ r_1 & r_o/2 & 0 & & & \\ r_2 & r_1 & r_o/2 & 0 & & \\ .. & .. & .. & .. & 0 & \\ .. & .. & .. & .. & .. & 0 \\ r_L & r_{L-1} & .. & & & r_o/2 \\ r_{L+1} & r_L & .. & & & r_1 \\ .. & .. & .. & & & .. \\ r_{2L} & r_{2L-1} & .. & & & r_L \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ .. \\ .. \\ a_L \end{bmatrix} \tag{3.46}$$

3.4.5 Correcting Negative Estimates

The positive definitness of the spectrum cannot be guaranteed by both equations (3.41) and (3.42). The reason for this could be either the mismatch between the spectrum and the rational model due to an inappropriate choice of the order or the error introduced by the uncertainty in the autocorrelations and parameters used in estimation. It has been suggested that the autocorrelation sequence $\{c_k\}$ of the moving average part in the cascade method be windowed by a triangular window to rectify the estimate [34]. However, we do not find any justification for choosing this particular window.

On the other hand, the one-sided spectrum method does not explicitly utilize any autocorrelation sequence except that of the original signal. Cadzow [35] suggested that the autoregressive parameters in his method be obtained by a weighted least square estimate as follows:

$$\underline{a} = -(R^T W R)^{-1} W R^T \underline{r} \qquad (3.47)$$

where W is a weighting diagonal matrix. Again, the choice of the weights in W is a problem in itself, even though Cadzow suggested taking $w(n) = (N-n)^3$ based on empirical results. We did not investigate this problem but we feel that a natural way to compensate for the autocorrelation error is to apply a window on the autocorrelations of the original signal, where the variance of the mean lag products estimate is known to increase as the lag is increased which makes the use of a triangular window rather sensible.

3.4.6 Spectral Factorization

Another approach which needs some more effort beyond the previous two methods is known as spectral factorization [36]. We will present the problem of finding the moving average parameters from the sequence $\{c_k\}$ whereby we can resort to the original positive definite rational spectrum $\sigma_e^2 B(z) B(z^{-1})/A(z) A(z^{-1})$.

Equating the two exact expressions for the moving average spectrum, we obtain,

$$(b_o + b_1 z + \ldots b_M z^M)(b_o + b_1 z^{-1} + \ldots b_M z^{-M}) = \sum_{k=-M}^{M} c_k z^k \qquad (3.48)$$

Equating the coefficients of equal powers of (z), we obtain the following system of nonlinear equations:

$$\begin{aligned}
b_o^2 + b_1^2 + b_2^2 \ldots \ldots + b_M^2 &= c_o \\
b_o b_1 + b_1 b_2 \quad \ldots \ldots + b_{M-1} b_M &= c_1 \\
\ldots \quad \ldots \quad \ldots \ldots \quad \ldots \ldots & \\
b_{M-1} b_M &= c_{M-1} \\
b_M &= c_M
\end{aligned} \qquad (3.49)$$

The solution to these equations is not unique. There are exactly 2^{M-1} solutions, which is equal to the number of possible exchanges of the M roots, of both polynomials B(z) and $B(z^{-1})$, while their product remains unchanged. Uniqueness needs some condition to be imposed on these

roots. An example for that is the minimum phase solution, i.e. all the roots of B(z) must lie inside the unit circle.

3.5 Rational Models of Interfering Autoregressive Signals

The choice of order of the rational model is a more difficult problem than it is in the autoregressive case. However, it can be shown that an autoregressive signal $x_1(n)$ of order M with additive interference from another autoregressive signal $x_2(n)$ of order L can be represented by an ARMA model of order (L+M, L+M). The proof follows. Denote by x(n) the signal resulting from the interference. Hence:

$$x(n) = x_1(n) + x_2(n) + w(n)$$

where w(n) is white noise. In the z-domain, the above equation becomes:

$$X(z) = X_1(z) + X_2(z) + W(z)$$

$$= \left[\frac{1}{A_L(z)} + \frac{1}{B_M(z)} + 1\right] W(z) \qquad (3.50)$$

where $A_L(z)$ and $B_M(z)$ are polynomials of order L and M, respectively. Hence,

$$X(z) = \frac{A_L(z) + B_M(z) + A_L(z)B_M(z)}{A_L(z)\, B_M(z)} W(z) \qquad (3.51)$$

$$X(z) = \frac{C_{L+M}(z)}{D_{L+M}(z)} W(z) \qquad (3.52)$$

The above equation gives an ARMA (L+M, L+M) representation for x(n). It has been shown by Chan et al. [37] that a signal comprised of K sinusoids plus white noise can be uniquely represented by an AR(2K) model. Considering the evident coupling between every two adjacent gastrointestinal oscillators, then it appears that an appropriate model for gastrointestinal signals is an ARMA model of equal orders of the numerator and denominator polynomials. A general case of this including m signals is a multivariate model of dimension (m).

CHAPTER 4

MINICOMPUTER IMPLEMENTATION OF A GENERAL PROGRAM FOR SPECTRAL ESTIMATION

4.1 Main Features of the Program

A general program has been implemented on a NOVA 830 minicomputer for spectral estimation of the (GI) signals. It has been efficiently used in studying the intestinal ECA [38]. The following features have been achieved in the program.

1. Versatility for the user to choose the estimation method which is most appropriate to his application or the possibility to try all methods.
2. Allowance for the analysis of long records of data through time segmentation in order to have complete information of possible time variations in response to some stimulus.
3. Graphing capability for 3 dimensions to provide convenient visualization of the change in the power spectral density with time. A 3D plotting program has been especially developed for this purpose. In addition to this, tables for the spectral peaks are provided on the output for quick inspection of rhythmic activities.

4.2 Storage of Data

Signals are sampled and digitized by the program DATAS which is already available on the system. This program controls an 8-channel A/D converter interfaced to the minicomputer. The output of the program is a multichannel file containing integer data stored in alternate blocks (each block is 256 bytes). A program WRDATA has been written to convert this file into several single channel files for more convenience in further programming and handling data. The names of the output files are chosen by the user. This is illustrated in Figure 4.1.

The storage of data in integer format requires half the space needed than for real data, since real numbers take twice as much space as taken by integers.

4.3 Structure of the General Program

Because of the limited memory space of the minicomputer (28 K available to the user program), it was necessary that the program be built in a modular structure. The program consists of a master program (MASTER) and four subprograms for performing the spectral estimation in the different methods.

The five modules are executed in a chaining mode as shown in Figure 4.2. The execution sequence is determined by the user's choice in the interactive part of the master program.

4.3.1 Master

This program provides an interactive capability through a user-computer dialogue for collecting the information supplied by the user

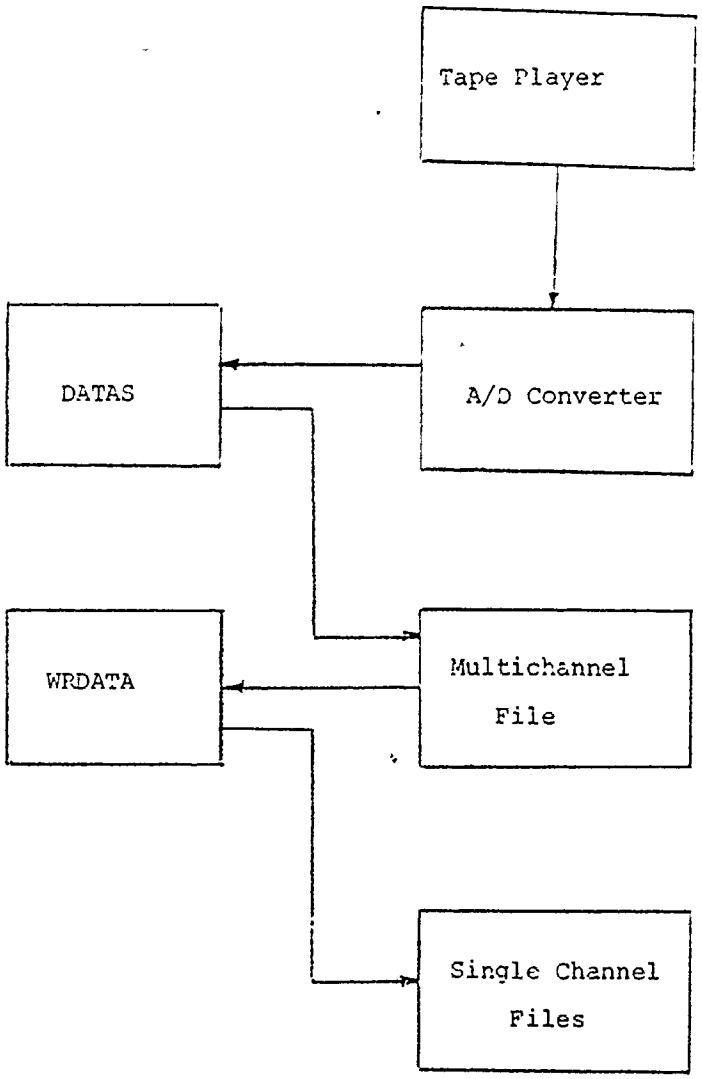


Fig 4.1 Procedure for data storage

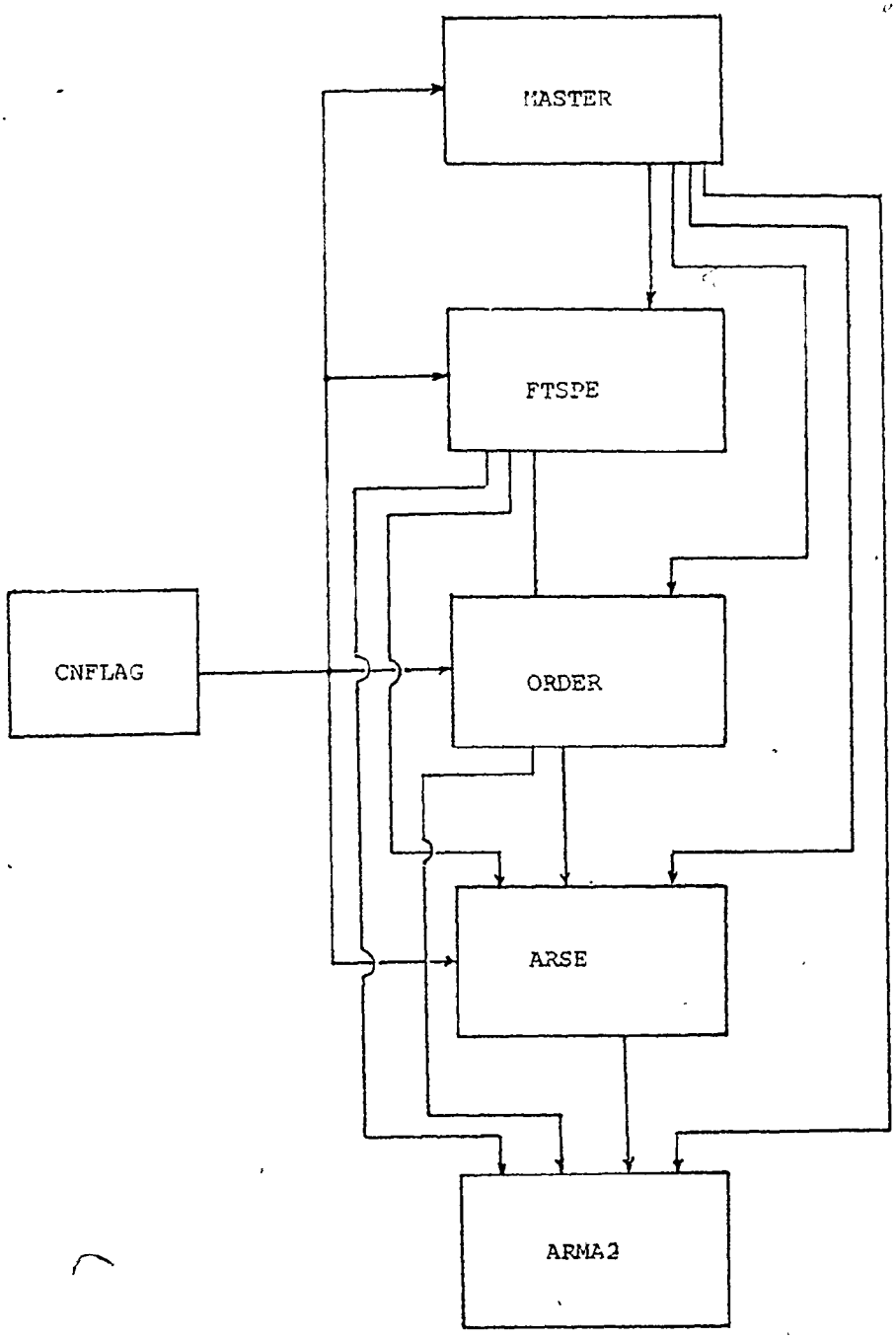


Fig 4.2 Possible chaining sequences during execution

who describes his data and requirements regarding the method to be used in estimation, plotting, etc. In order to supply the information to other modules, the following files are created by MASTER [Figure 4.3].

FDIRECT: A directory for the names of the data files to be analyzed, the number of segments per file and the number of samples per segment.

CNFLAG: A file that contains flags as to which module should be executed next.

CNTRL1, CNTRL2, CNTRL3 and CNTRL4: Files that control the programs FTSPE, ORDER, ARSE and ARMA2, respectively. They also provide file names supplied by the user for saving the spectra obtained by the different methods.

CNTPLT: A file that contains the information needed for plotting.

4.3.2 Execution of the Subprograms

Any of the other four modules may be fetched from the disk, loaded in the core memory and given the control by MASTER which is completely overlayed by the loaded program. This may still happen for any of the three remaining modules if it is needed to satisfy the user requirements. In this case, the next subprogram overlays the previous one in the memory. This chaining sequence is controlled by flags stored in CNFLAG which is accessible by all modules as shown in Figure 4.1 and Figure 4.4. The flow of information during execution is shown in Figure 4.4. Each of the control files CNTRL1, ... CNTRL4 is accessed by the corresponding program for retrieving the parameters governing the

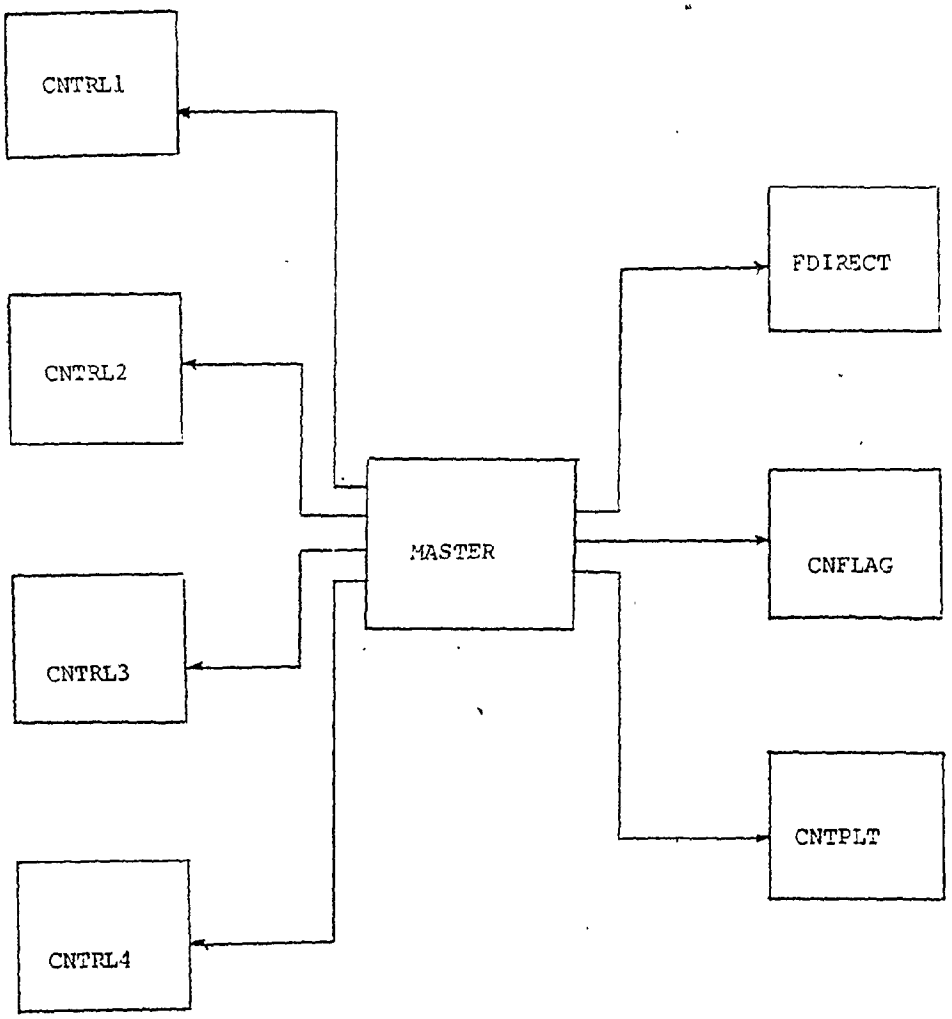


Fig 4.3 Control files generated by the master program

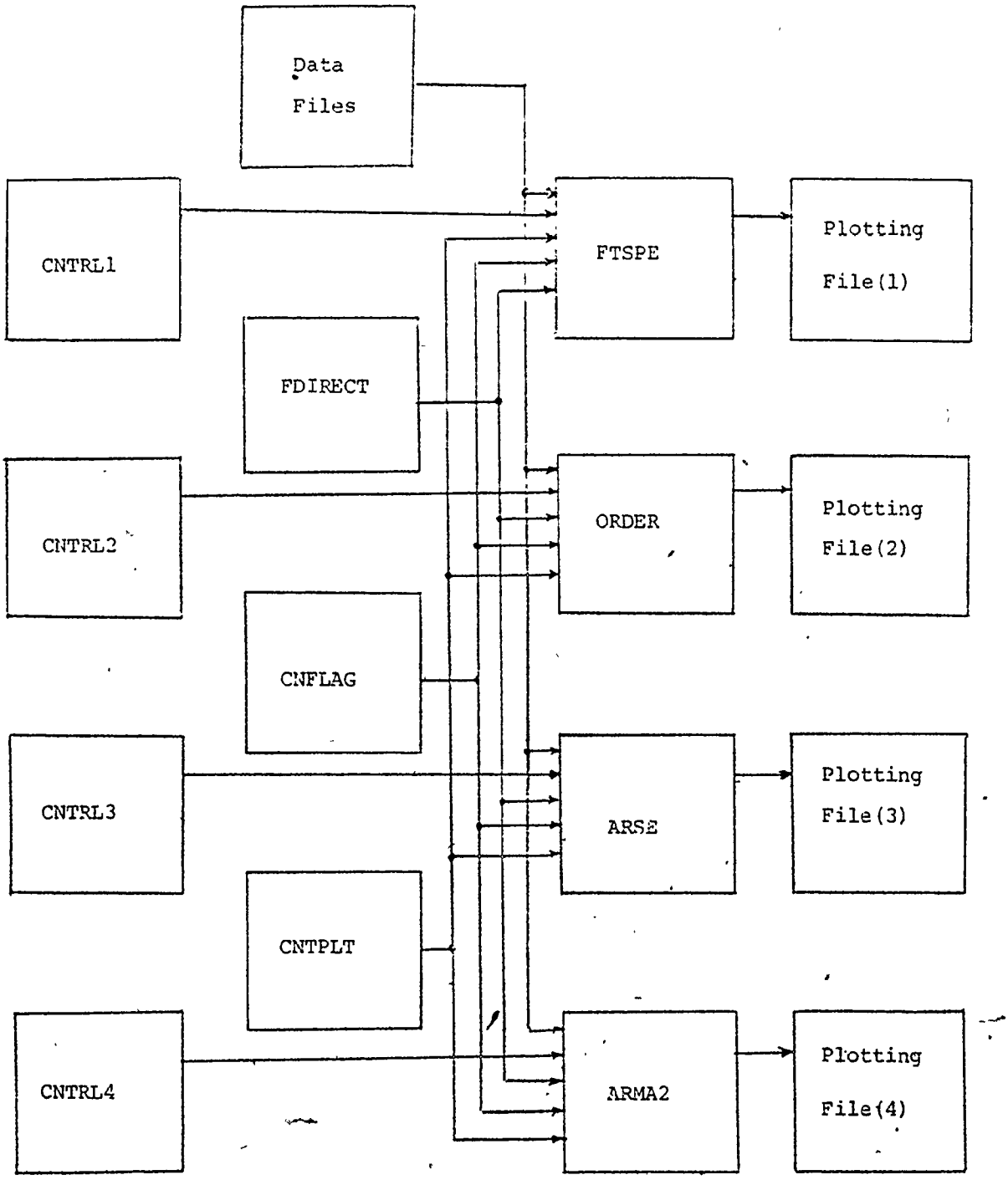


Fig 4.4 Flow of information during execution

execution of the pertinent method. Data to be analyzed is fetched from the disk by the aid of the directory stored in FDIRECT which is also accessible by all four modules. The output of every program is stored in a corresponding file ready for plotting. The dimensions and the form of plots are determined by the information stored in CNTPLT.

4.3.3 Structure of the Subprograms

FTSPE: This program provides either the periodogram of the tapered data or the discrete Fourier Transform of the windowed autocorrelations, or both.

Cosine tapering is first applied to the first and the last tenths of the data according to the following formulae:

$$\begin{aligned} x(n) &= 0.5 - 0.5 \cos\left[\pi(n-1)/(N-1)\right] && n < N/10,\ n > 9N/10 \\ &= x(n) && n/10 \le n \le 9N/10 \\ & && n = 1, 2, \ldots, N-1 \end{aligned} \qquad (4.1)$$

This is called "Hamming window". If the coefficients (0.5, 0.5) are replaced by (0.54, 0.46), the tapering function is called "Hanning". The discrete Fourier Transform is calculated by the Jim Cooley decimation in time algorithm. The autocorrelations are estimated by the Fast Inverse Fourier Transform of the Periodogram. The options provided for the autocorrelation window are:

Barlett window:

$$w(k) = \frac{L - K}{L} \qquad (4.2)$$

Parzen window:

$$w(k) = (1 - 6K/L)^2 \, (1 - K/L) \qquad K < L/2$$

$$= 2(1 - K/L)^3 \qquad K \geq L/2 \tag{4.3}$$

$$K = 0, 1, 2, \ldots, L$$

L zeros are appended to the L autocorrelation lags for improving the computational resolution (i.e. providing finer frequency spacing between spectral samples).

The program consists of the following segments:

FTSPE: main program

CFT: subroutine for tapering the data

FFT: subroutine for calculating the Fast Fourier Transform

PEAKS: subroutine for generating a table for the relative amplitudes of the spectral samples

PLOT3D and PLO3D: subroutines for generating a 3-D plot of the spectra of the different segments

ARSE: This program estimates the autoregressive spectrum with a user-specified order and consists of the following segments:

ARSE: main program

SOLV1: a subroutine for estimating the autoregressive parameters

PEAKS, PLO3D and PLOT3D: (same as in FFT)

ORDER: This program is for calculating the quantities needed in checking the autoregressive order. It utilizes only one subroutine (SOLVE) for finding the autoregressive parameters.

ARMA2: This program provides a suboptimal estimate of the rational spectrum. It has a similar structure to that of ARSE with SOLV1 repalced by SOLV2 for calculating the parameters of the suboptimal one-sided spectrum method.

With the exception of ORDER, these programs have an overlay structure, in the sense that only the main program and the FORTRAN subroutines reside in the memory during execution. The other subroutines are loaded in an overlay area whenever needed, as shown in Figure 4.5. This saves up to 4K of the core memory which allows more flexibility in array dimensioning. Details of the programs and FORTRAN listings are available in [39].

4.4 Three-Dimensional Display of Spectra

The method employed in producing the 3-D graph of the spectra of contiguous segments in a long record, is derived from the first principles of engineering drawing of isometrics. The individual plots are imagined to be arranged inside a parallelpiped, as shown in Figure 4.6, depending on the angle through which the elevation projection is looked at. The plotting is performed on a PRINTRONIX 300 line printer which can be programmed in FORTRAN as an incremental plotter which is commanded on the dot level instead of the dot matrix in the normal printing mode. The programs which interface the plotting routines

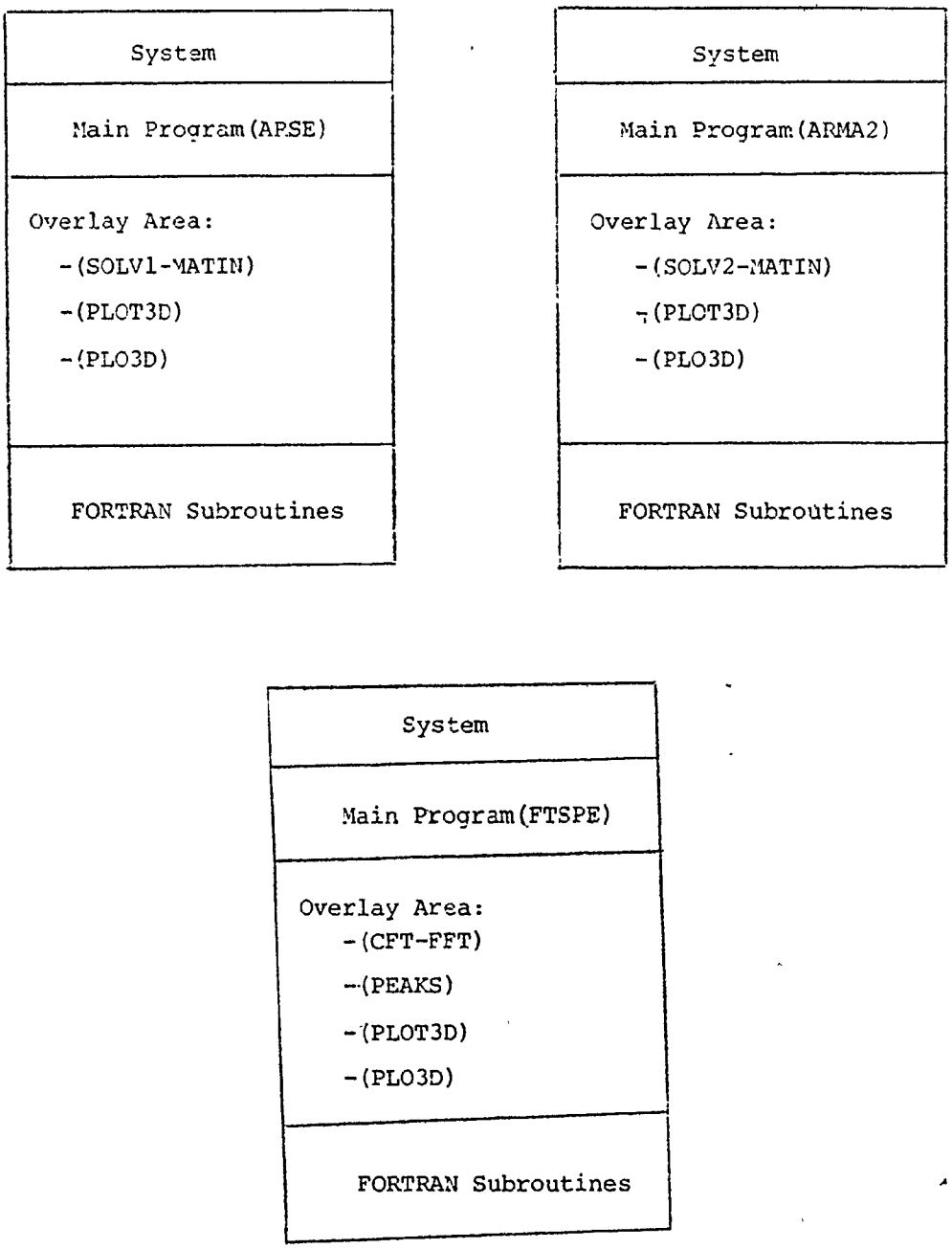


Fig 4.5 Overlay structure of the subprograms

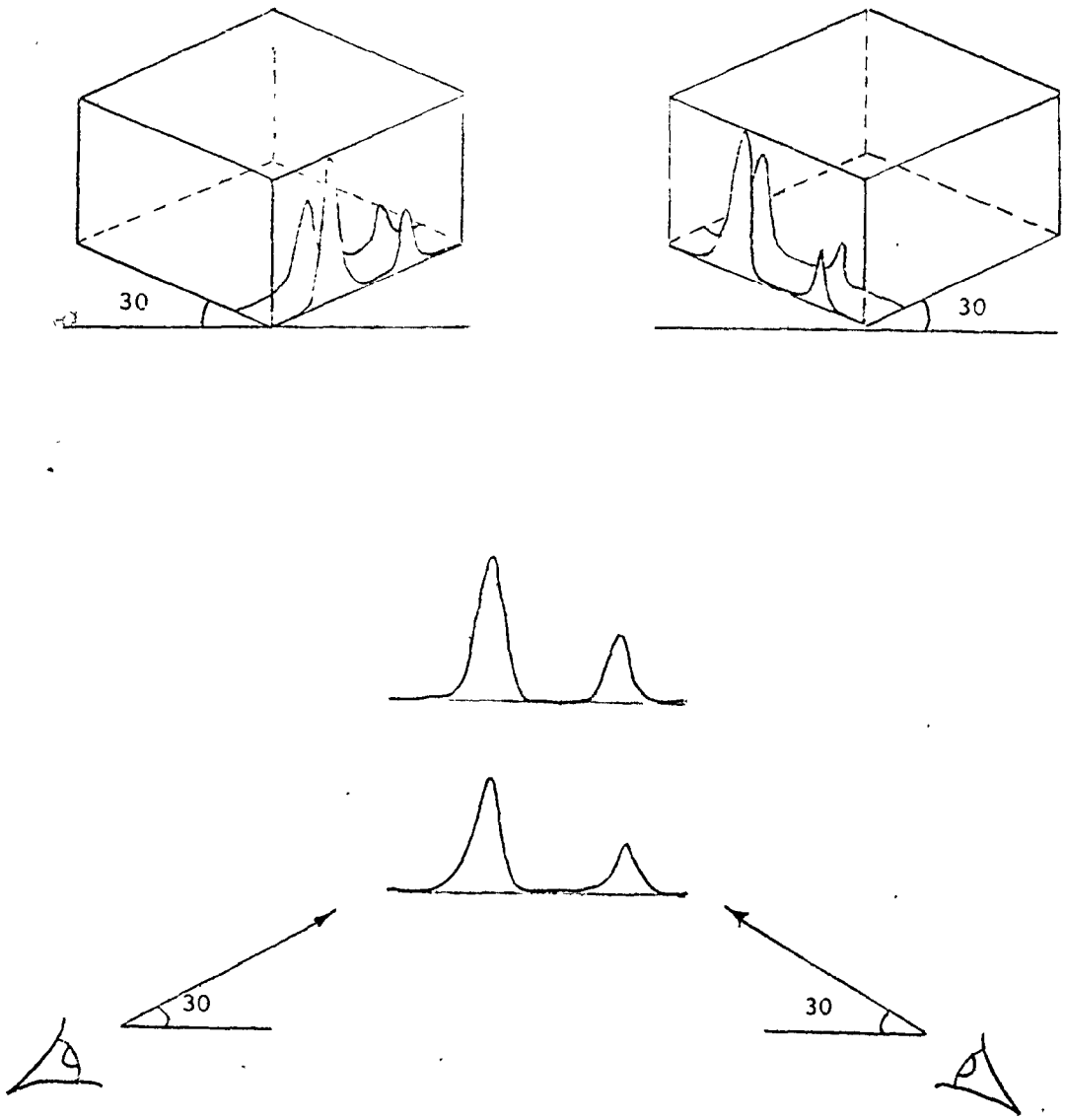


Fig 4.6 Isometric display of spectra in three dimensions

(PLO3D, PLOT3D) are provided by the CCSI library (LPTPLOT.LB). FORTRAN listings for the plotting routines are also available in [39].

4.5 Example

Consider a case when it is desired to investigate the temporal variations of the power spectral density of a single channel stored in a separate file using all the available methods, periodogram, correlation, AR and ARMA. Other specifications are illustrated in the dialogue of Figure 4.7 which is output to the line printer simultaneously while the user is responding to the system console. The program output is shown in Figure 4.8 (a-d).

The time variation of the spectra of a certain channel or the spatial distribution of the spectra of several channels may be obtained by choosing the appropriate mode of operation (spatial or temporal). The access of data for each case is illustrated in Figure 4.9.

```
DATE(MO/DAY/YR)  JN/24/81
ENTER 1 FOR SPATIAL OR 0 FOR TEMPORAL   0
ENTER NO. OF FILES   1
NAME OF FILE 1   DATA5
NUMBER OF SEGMENTS PER FILE  10
NUMBER OF SAMPLES PER SEGMENT   512
SKIP ANY SEGMENTS IN THE BEGINNING OF RECORD? NO
FFT SPECTRAL ESTIMATION? YES
PERIODOGRAM ESTIMATION? YES
ENTER NUMBER OF FFT POINTS  1024
ENTER 1 FOR HAMMING OR 2 FOR HANNING WINDOW   1
USE CORRELATION METHOD? YES
ENTER MAX LAG    256
ENTER 1 FOR BARLETT OR 2 FOR PARZEN WINDOW  1
ENTER NUMBER OF FFT POINTS   1024
FILE NAME FOR SAVING FFT SPECTRA. PLT1
AR SPECTRAL ESTIMATION ? YES
EXAMINE AR ORDER ? NO
ORDER FOR FILE( 1): 36
FILE NAME FOR SAVING AR SPECTRA . PLT2
PRINNT MODEL PARAMETERS ? NO
ARMA SPECTRAL ESTIMATION ? YES
ORDER FOR FILE( 1): 20
FILE NAME FOR ARMA SPECTRA . PLT3 .
PRINT MODEL PARAMETERS ? NO
FOR PLOTTING ENTER
FREQ SPACING FOR RATIONAL METHODS(IN C/MIN)      0.500
SAMPLING FREQUENCY IN HZ     10 0
APPROX MAX FREQ TO PLOTTED(IN C/MIN) :   80 0
PLOT LENGTH IN INCHES :     6.0
PLOT HEIGHT IN INCHES :     2.0
APPROX SEPARATION BETWEEN PLOTS(IN INCHES) .   0 50
   1 TO PLOT LOOKING FROM RIGHT ANGLE
OR 0 TO PLOT LOOKING FROM LEFT ANGLE   0
```

Fig 4.7 Computer user dialogue of the example of section 4.5

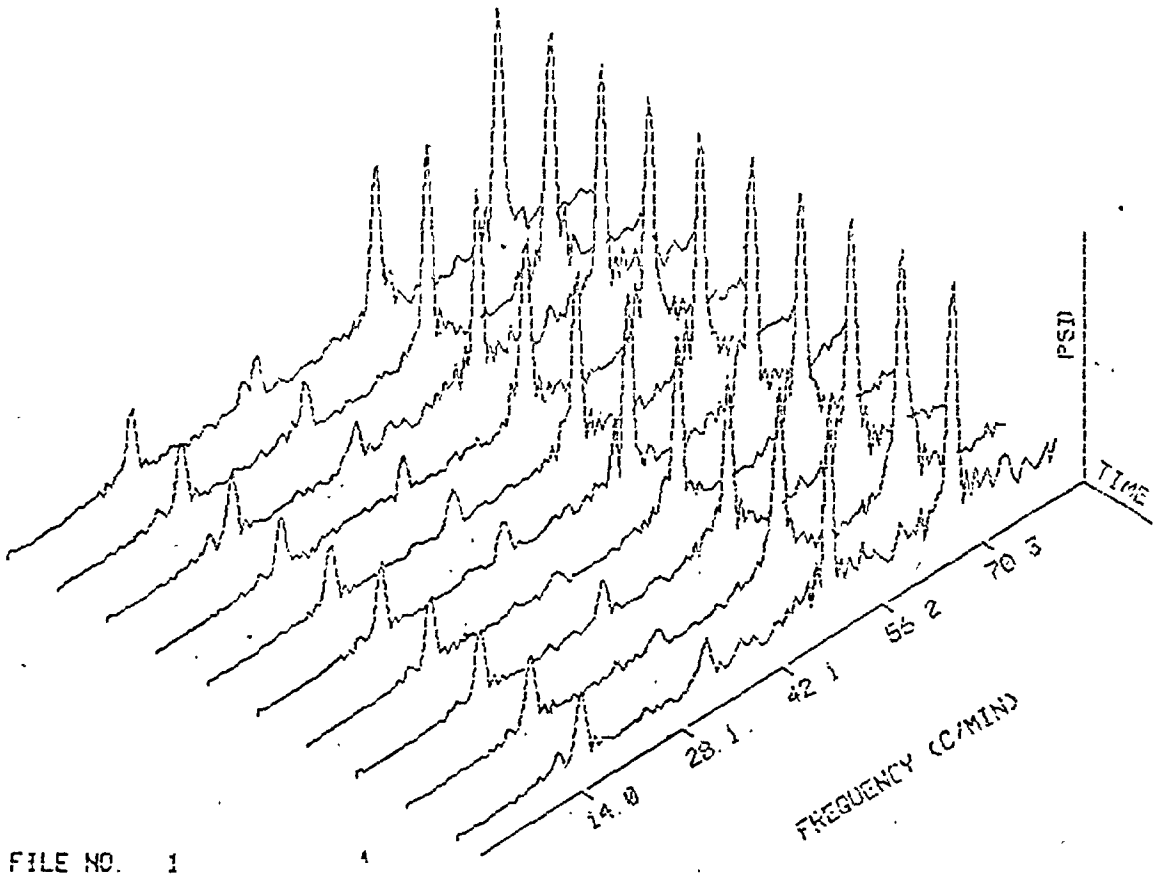
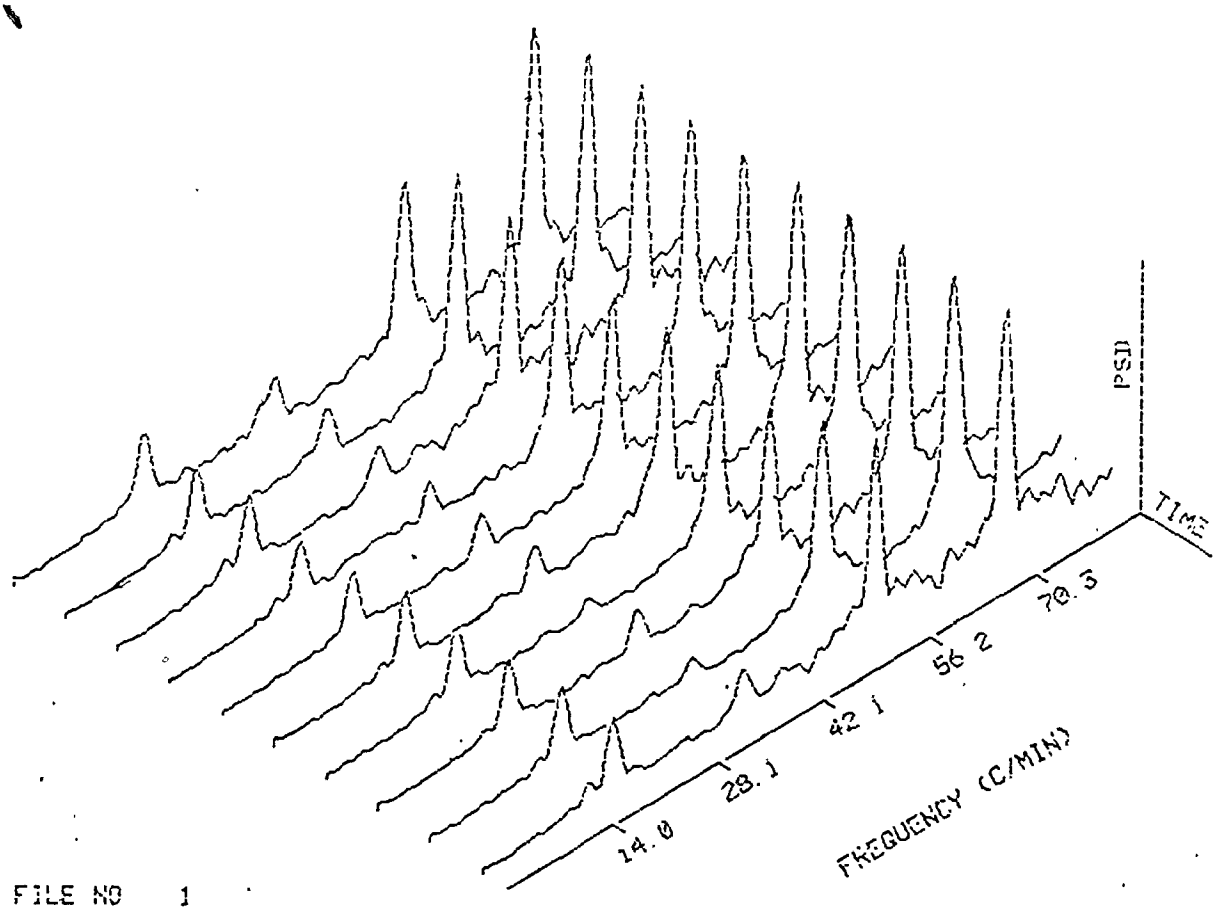


Fig 4.8 (a) Output of the example of section 4.5



FILE NO 1
CORREL METHOD
SEGMENT LENGTH = .8 MIN

Fig 4.8 (b)

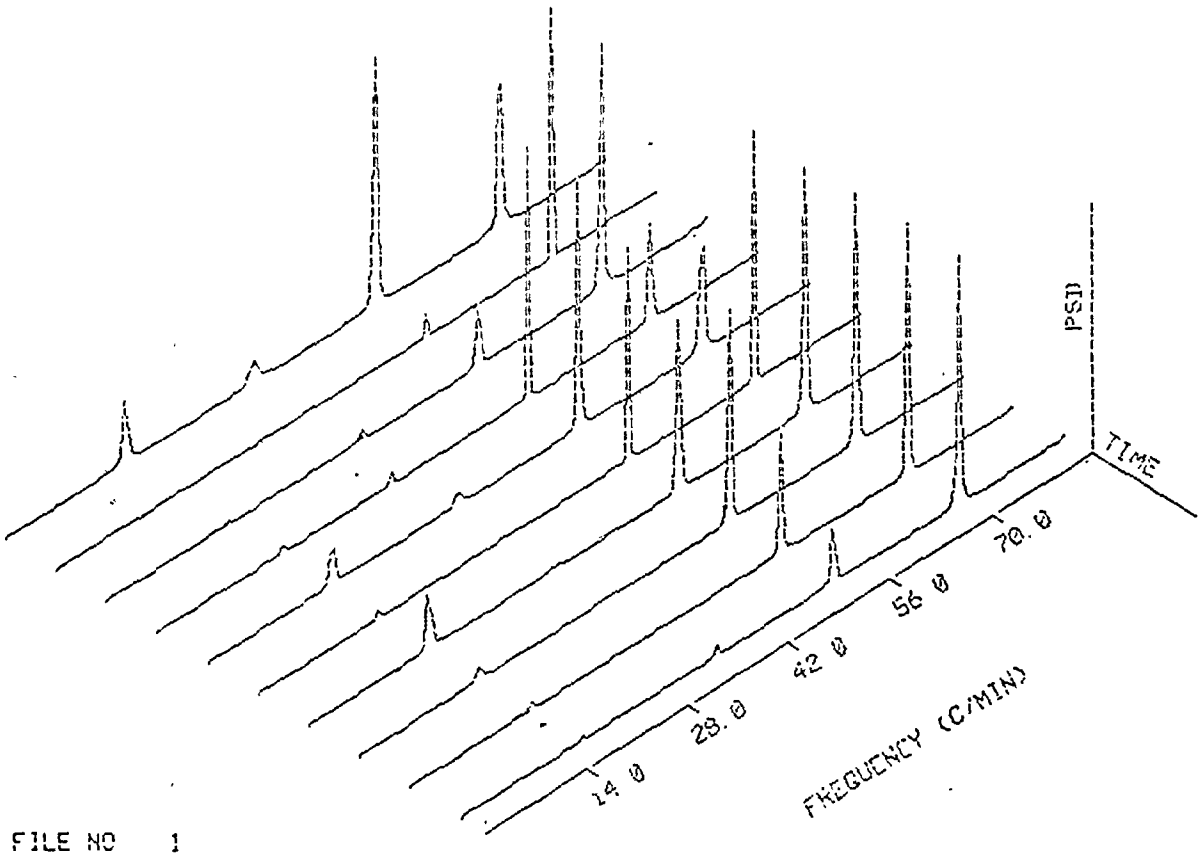
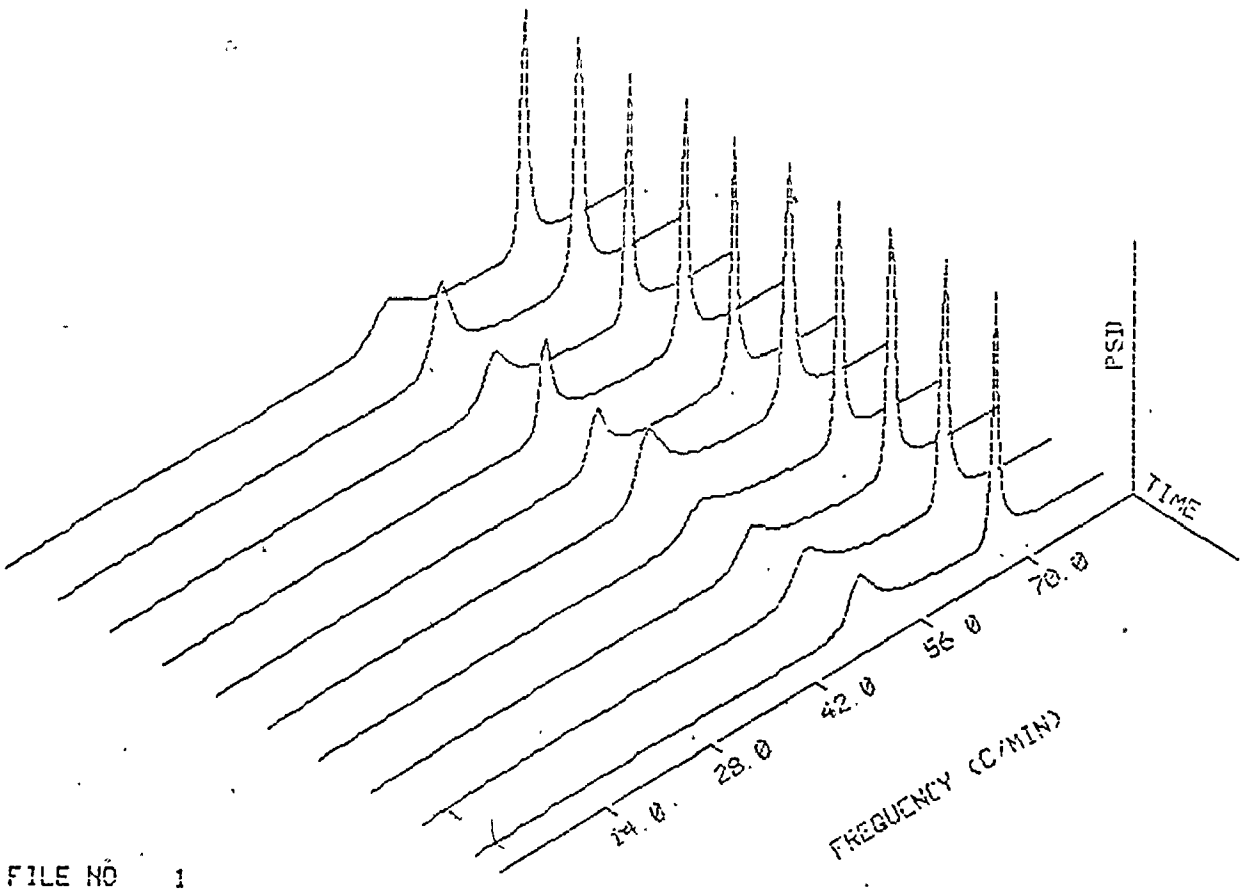

PSD
TIME
14.0
28.0
42.0
56.0
70.0
FREQUENCY (C/MIN)
FILE NO 1
AR 36 SPECTRA
SEGMENT LENGTH = 9 MIN

Fig 4.8 (c)



FILE NO 1
ARMA 20 20SPECTRA
SEGMENT LENGTH= .9 MIN

Fig 4.8 (d)

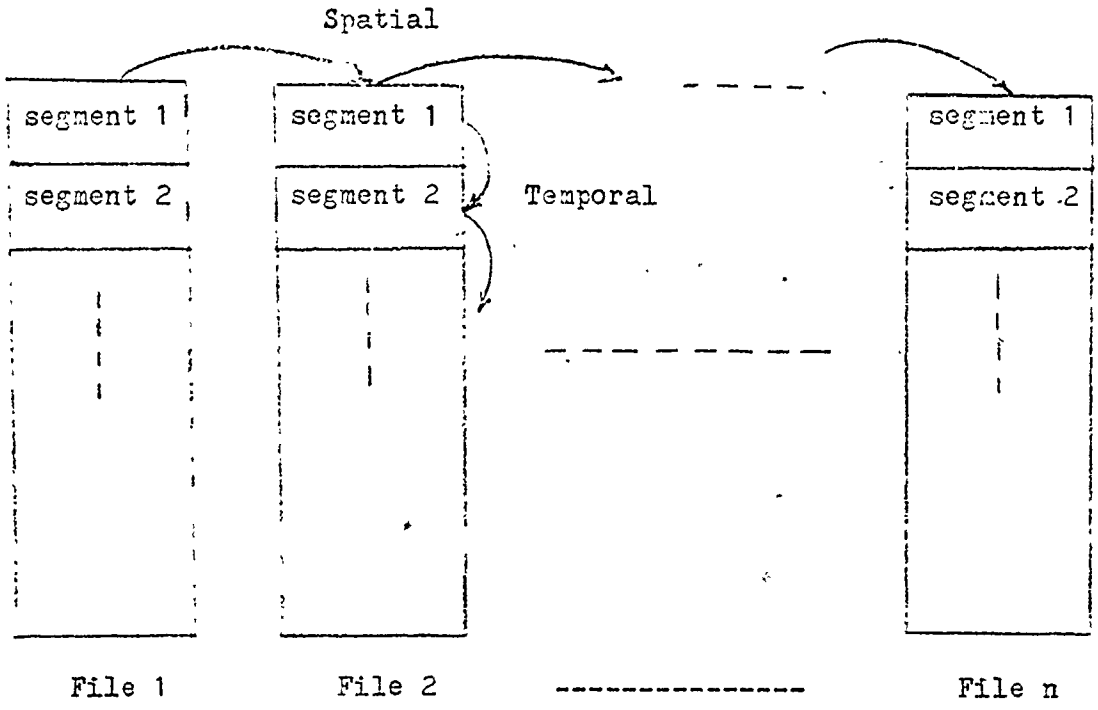


Fig 4.9 Spatial and temporal modes of operation

CHAPTER 5

SPECTRAL ESTIMATION OF THE SMALL INTESTINAL ECA

This chapter presents and discusses the results of applying four of the previously mentioned spectral estimation methods to signals recorded from a dog's small intestine.

The results are arranged to provide a set of comparisons among the four methods used to show some of their relative advantages and drawbacks. The comparisons are of qualitative nature and are based on analyzing many segments in long records to avoid the misconception which may result from a single experiment. This has been made easy by the general program described in Chapter 4.

The temporal variations and the spatial distribution are also investigated for the ECA spectra in the small intestine.

5.1 Experimental Procedure and Data Acquisition

A set of 17 monapolar Trimel wire electrodes was implanted subserosally on the small intestine of a dog. The electrodes were placed equidistantly between the pylorus and the ileocecal junction with the first electrode being at 10 cm from the pylorus, as shown in Figure 5.1. The electrode leads were brought out through a stainless steel cannula embedded in the abdominal wall. The leads were connected to a 31-pin female ampheriol plug in the cannula. The male

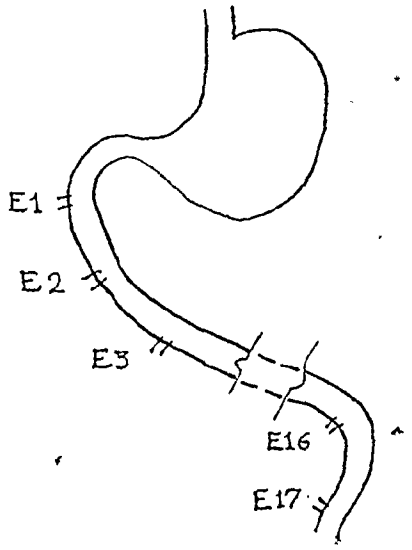


Fig 5.1 Electrode arrangement on the dog small intestine

end was connected to it for recording. All recordings were made on a Beckman recorder with lower and upper cutoff frequencies set at 0.16 and 30 Hz, respectively. Signals were simultaneously recorded on a Hewlett Packard FM tape recorder which was played back for digitization. Signals were first filtered using an analog low pass filter with a cut-off frequency of 4 Hz and the sampling rate was chosen to be 10 Hz.

5.2 Autoregressive Order for the ECA in the Small Intestine

AR models of increasing order were fitted to all 17 signals and the decision criteria were evaluated and plotted against the order. In Figure 5.2 an example is shown for the three criteria as obtained from E3 in the electrode arrangement of Figure 5.1. A typical one-minute segment of the signal is shown in the top of Figure 5.4. The variance of the residuals and the determinant ratio are seen to have an assymptotic behaviour towards continuous improvement as the model order is increased up to 30. The implication of this is that the signal cannot be regarded as a true autoregressive process of a finite order and that the true order may be considerably large or infinite. However, there was a substantial improvement in the model fit as the order was increased from 15 to 20. On the other hand, there was little change in performance as the order was increased from 3 to 12 or from 22 to 30. Therefore, if a "good" autoregressive approximation with a reasonable order is desired, the choice may be limited to those orders which show that drastic change. The AIC in this example has a local minimum at a small order which will be shown to underestimate the order. The global

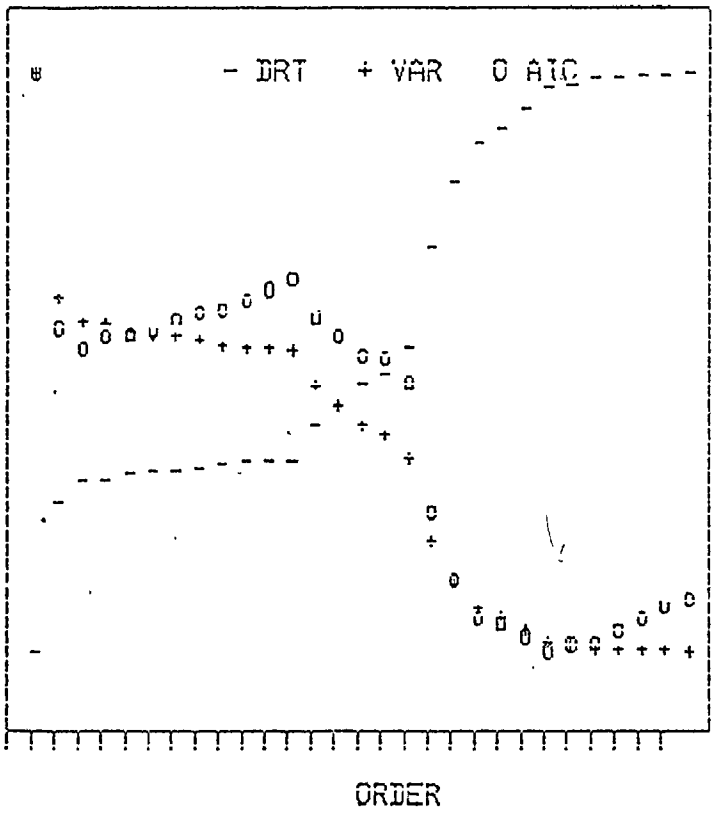


Fig 5.2 Decision criteria for AR order choice in the example of section 5.2

minimum of the AIC within the considered range is consistent with the other two criteria. However, in many other cases, the global minimum does not correspond to the order which yields an improvement in the variance or the determinant ratio.

The dependence of the resolution on the order can be seen by comparing Figures 5.3 and 5.4 where the AR(16) spectra have broad peaks to the extent that makes it difficult to identify the rhythm while the AR(20) spectra obviously have better resolution. This is in agreement with the performance interpretation on the basis of the criteria shown in Figure 5.2. However, the order cannot be chosen arbitrarily large. This depends on the number of samples used in estimation. It has been suggested in [21] that the order should be kept less than half the number of samples. But even with this restriction, the results may be unpleasant in some cases, as will be demonstrated in Section 5.5.

5.3 Applicability of Seasonal Models and the Effect of the Sampling Rate

Seasonal models of time series are based on prior knowledge of the seasonalities of the series, i.e. the frequencies of repetition. Consider a time series having m seasonalities at S_1, S_2, ... S_m samples apart. The seasonal model is built by taking the difference between every sample x(nT) and another sample at a time S_1T samples ahead, i.e., $x(nT + S_1T)$, assuming S_1 to be the largest. This differencing will cancel out the periodicity at the S_1T intervals and produce a new series $x_1(nT)$ which is treated similarly to cancel out the S_2 periodicity. This

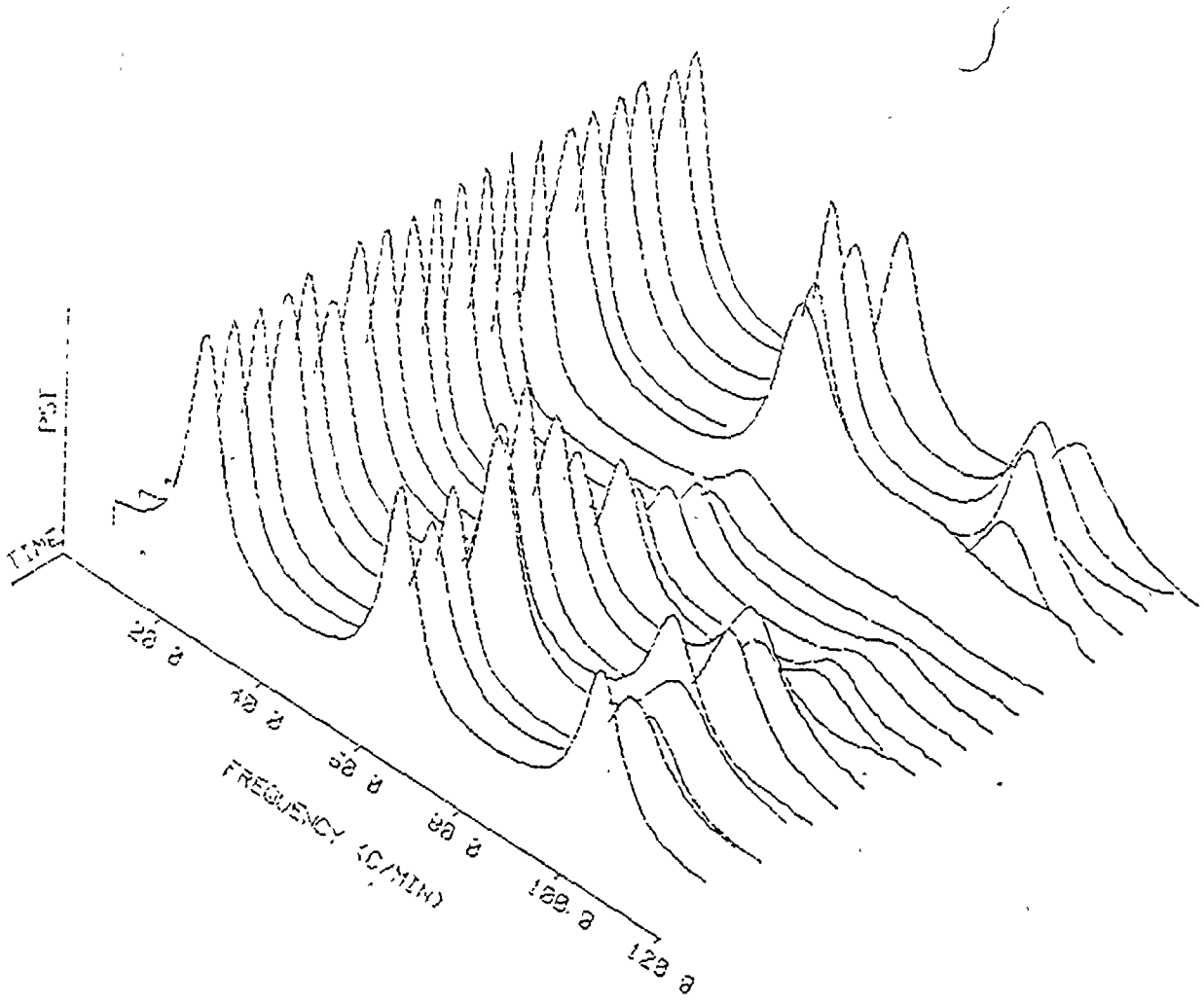


Fig 5.3 AR(16) spectra of the example of section 5.2

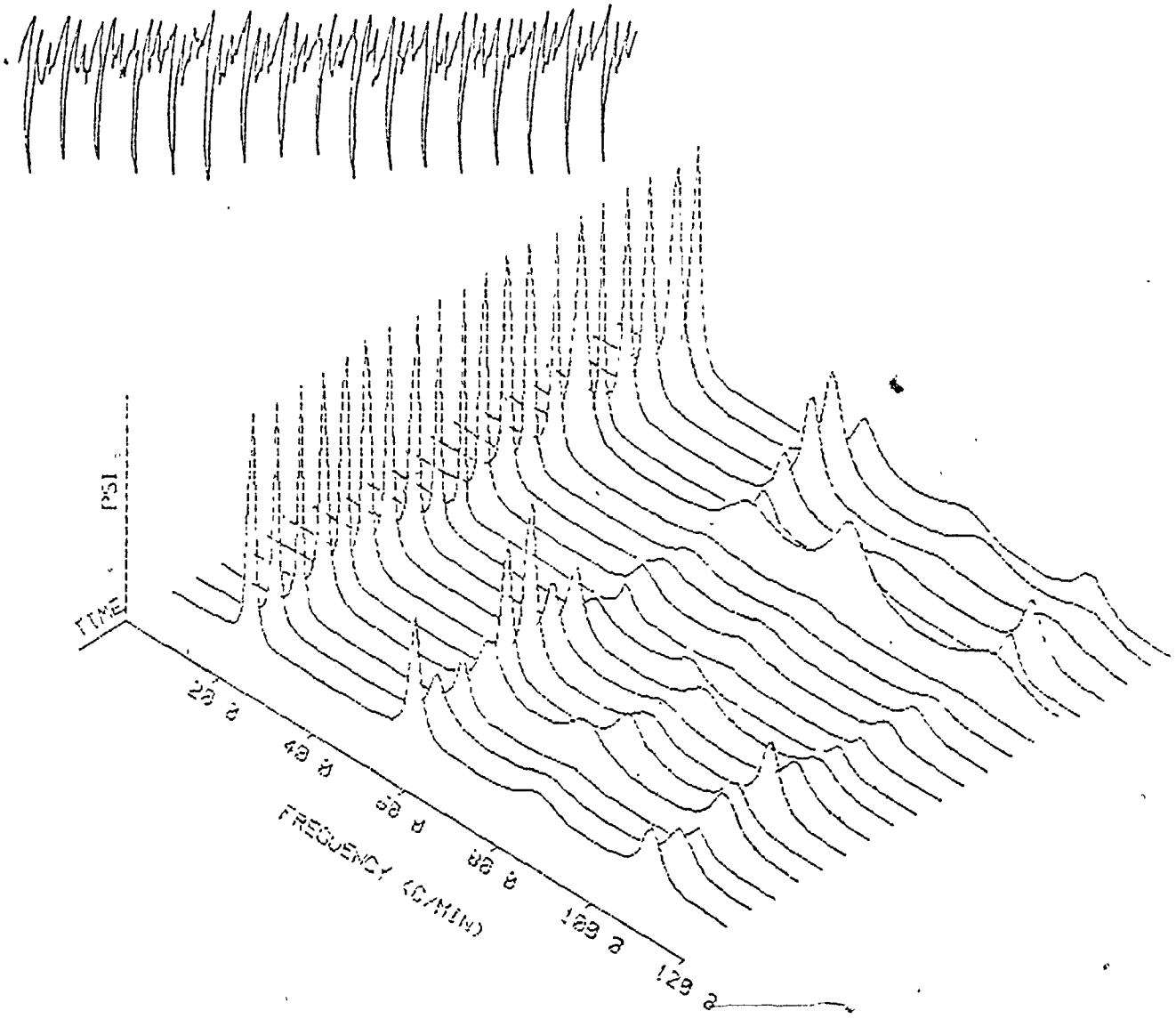


Fig 5.4 AR(20) spectra of the example of section 5.2.
A one minute segment of the signal is shown in the top.

process continues until the S_m periodicity is cancelled and the final series after the m differencing processes will be $x_m(nT)$. If the series is exactly periodic, x_m should be constant, but generally it could be an ARMA process. The above operations can be expressed as follows:

$$x_1(nT) = x(nT) - x(nT + S_1T) \tag{5.1}$$

Hence $$X_1(z) = X(z)\,(1 - z^{S_1}) \tag{5.2}$$

Hence $$X(z) = \frac{1}{(1 - z^{S_1})}\, X_1(z) \tag{5.3}$$

$$X(z) = \frac{1}{(1 - z^{S_1})\,(1 - z^{S_2})\,\ldots\,(1 - z^{S_m})}\, X_m(z) \tag{5.4}$$

This gives a model whose denominator polynomial is factorized as in equation 5.4 and has a degree equal to $(S_1 + S_2 + \ldots + S_m + \text{AR order of } x_m)$ which clearly depends on the sampling rate. The above model is a special case of a general model whose denominator is of the same order but not necessarily factorizable in the same way. This discussion has been meant to show the effect of the sampling rate on the model order, a fact which can be intuitively realized. The seasonal models themselves are inapplicable in case of spectral estimation because they require knowledge of the periodicities which are the quantities sought.

The signal treated in the previous example has been reconsidered at a sampling rate of 5 Hz, which is half the previous rate. The order

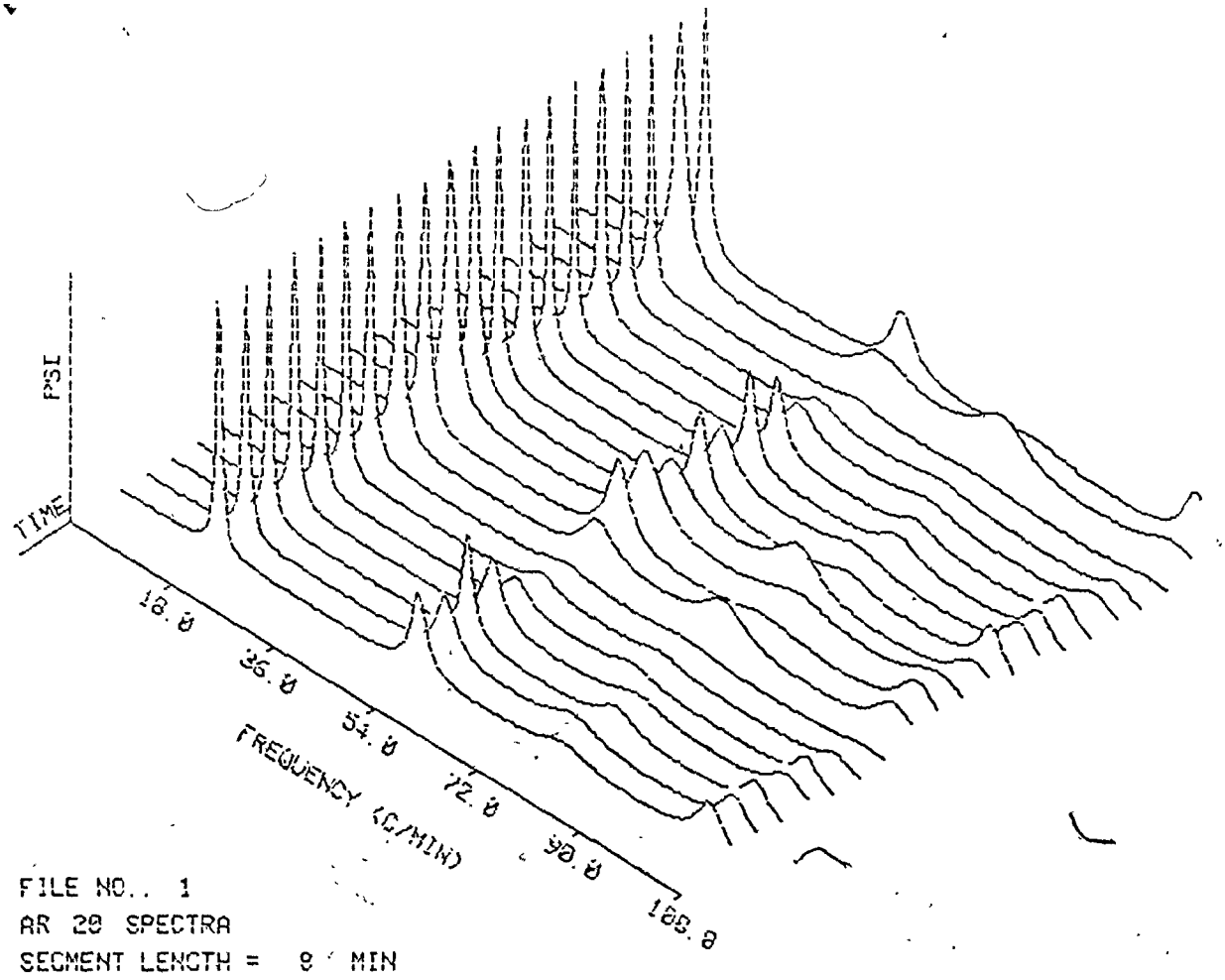


Fig 5.5 Spectra of the example of section 5.3 at a 10 Hz sampling rate.

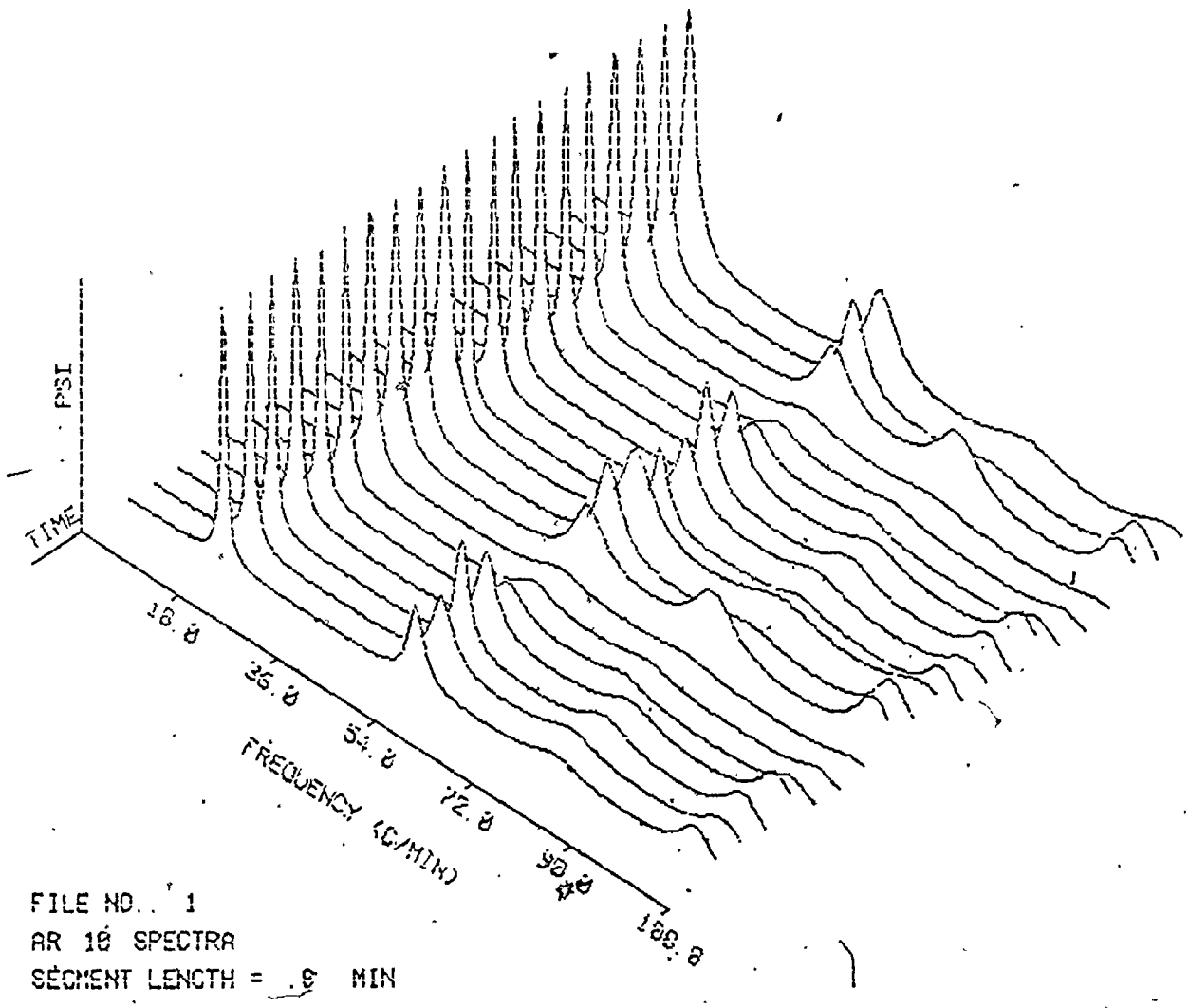


Fig 5.6 Spectra of the example of section 5.3 at a 5 Hz sampling rate.

is accordingly expected to decrease. The spectra of a 16 minute record are shown in Figures 5.5 and 5.6, for 20 equal length segments. The number of samples is 512 and 256 for the 10 Hz and 5 Hz sampling rates, respectively. The AR(10) and the AR(20) models are nearly equivalent in performance as far as the spectrum is concerned. They both resolve the fundamental shythm at about 17 c/min with slight differences in the appearance of harmonics.

5.4 Application of FFT and AR Methods to Noisy ECA

The 8 minute record shown in Figure 5.7 was analyzed in ten 0.8 minute long segments. The first four segments are noisy, as shown in the top four traces. This noise may be caused by the moving abdominal environment. The fourth segment starts to show some regular rhythm which continues steadily for the next 5 minutes. The first look at the first three segments may suggest that the signal is not rhythmic. Moreover, application of the periodogram and the autocorrelation methods results in spectra with large variations and many spurious peaks, as shown in Figures 5.8 and 5.9. However, the autoregressive method gave much smoother spectra with no spurious peaks (Figure 5.10). The shown peaks are relatively broad but they do not lead to the detection of false rhythms, as may be concluded from the FFT methods. The inefficiency of the FFT methods in this case is due to the fact that in order for these methods to be reliable, the data should have periodic extension or should be very long. For example, if the periodogram was calculated for the whole 8 minute record, the result would have certainly been

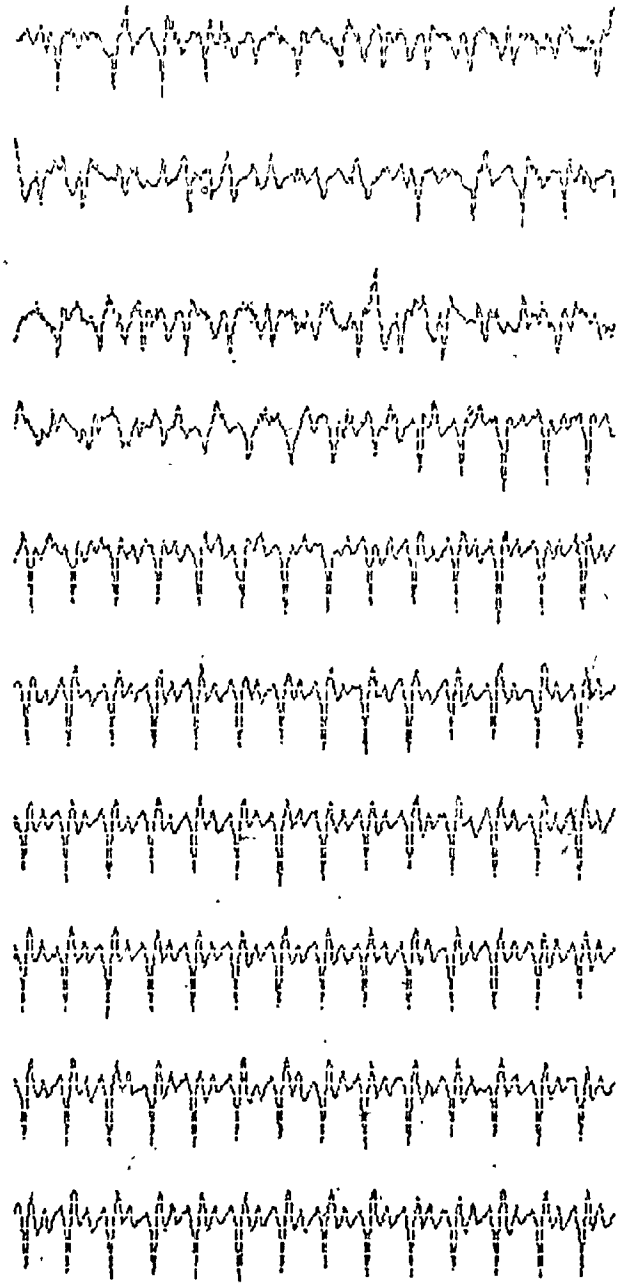


Fig 5.7 The 8 minute record of the first example of section 5.4.

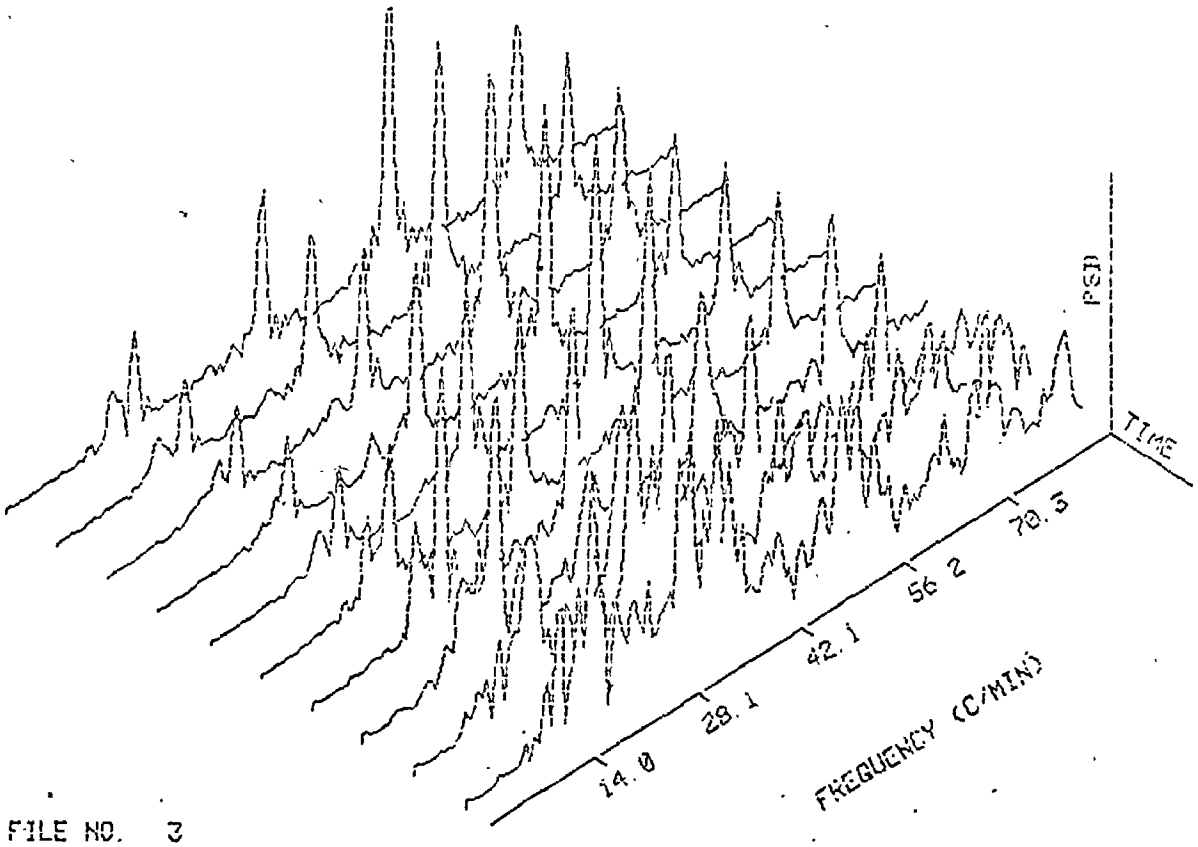
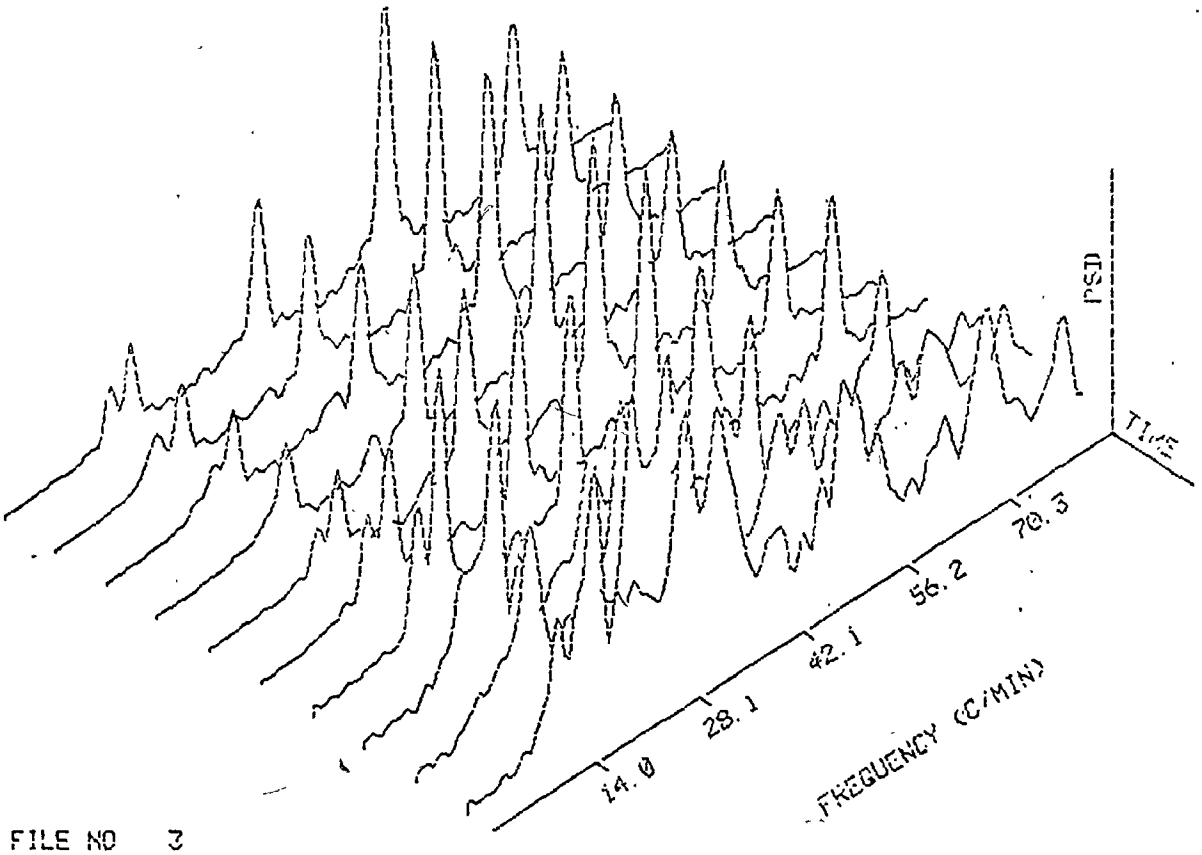
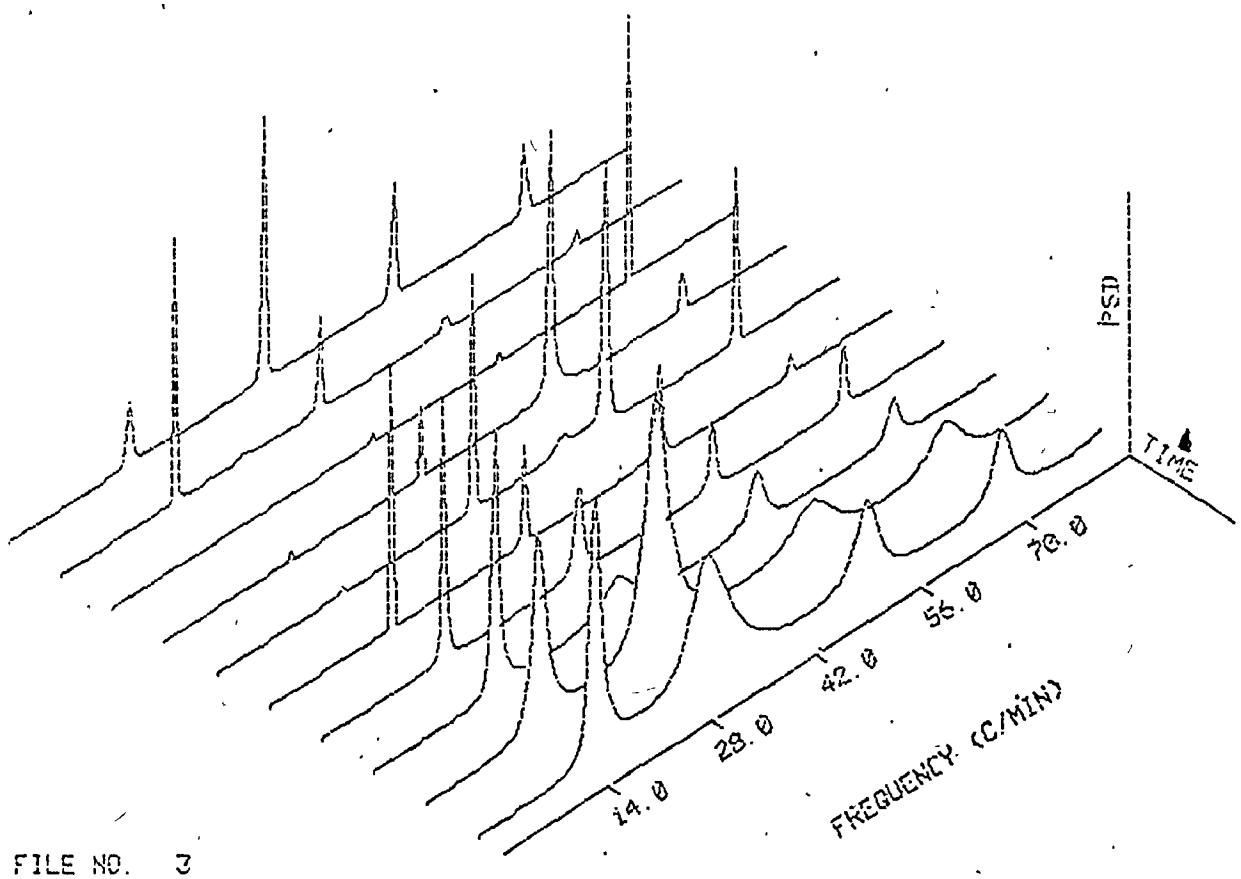


Fig 5.8 Periodogram of the record of fig 5.7 (first example in section 5.4)



FILE NO 3
CORREL METHOD
SEGMENT LENGTH = .8 MIN

Fig 5.9 Spectra of the record of fig 5.7 via the autocorrelation method.



FILE NO. 3
AR 36 SPECTRA
SEGMENT LENGTH = .8 MIN

Fig 5.10 AR(36) spectra of the record of fig 5.7

improved. However, the FFT is ideal for signals which are clearly periodic, as shown in the example of Figures 5.11, 5.12, and 5.13. The signal shown is obtained by E2 from the duodenum.

5.5 Line Splitting in AR Spectral Estimates

In the previous section, it was mentioned that the peaks obtained from the noisy record were broad, which would imply a deterioration in the resolution power of the AR spectrum as the signal-to-noise ratio becomes smaller. This has been analytically dealt with by S. Kay in [40]. Another drawback of the autoregressive spectrum obtained by either Yule-Walker or Burg's parameter estimates is known as the line splitting phenomenon, which sometimes occurs when the order is overestimated. This phenomenon has been dealt with in [41] and [42] for Yule-Walker and Burg's estimates, respectively. Although it is known that the order of any process may be theoretically infinite, the practical estimation procedure requires the order to be a fraction of the total number of samples. To the best of the author's knowledge, there is no definite limit to this fraction to prevent line splitting and it seems that trying different orders, though tedious, may be guarding against this pitfall. Through the author's practice with the intestinal ECA, line splitting was obtained only once for two segments in the record of Figure 5.7. In Figure 5.14, splitting can be seen in the peaks of the fundamental frequency in the 10th segment and the second harmonic of the 6th segment. This happened when the autoregressive order of this example was increased to 40.

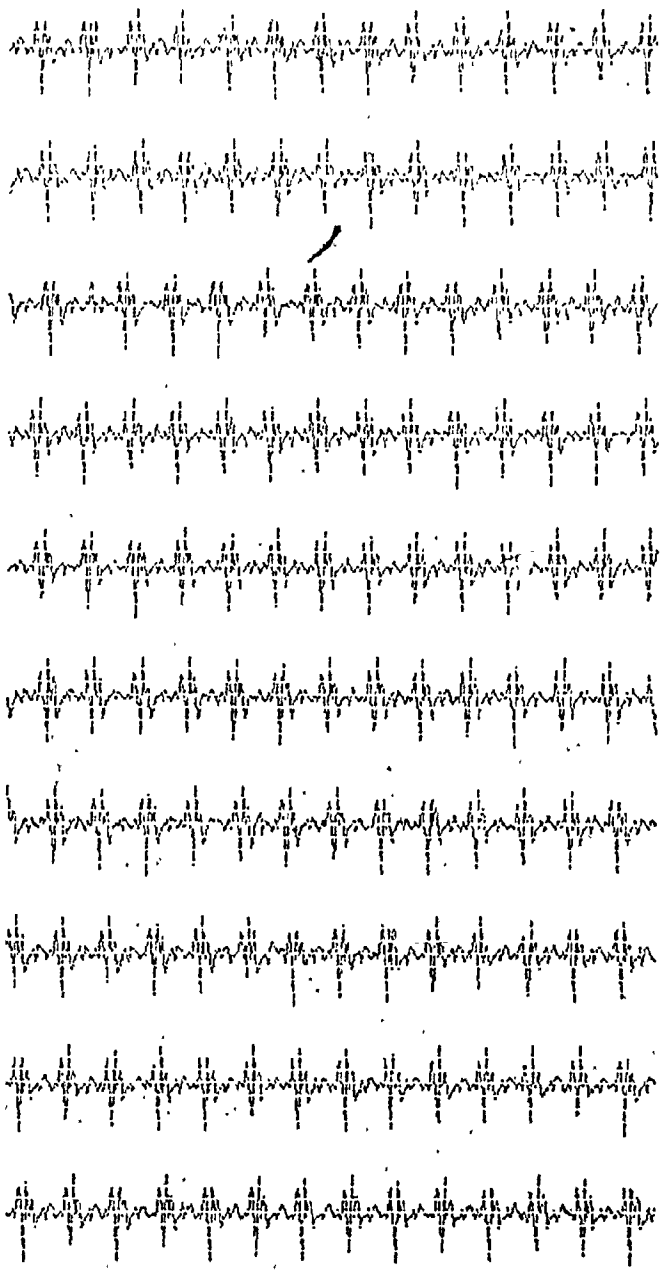
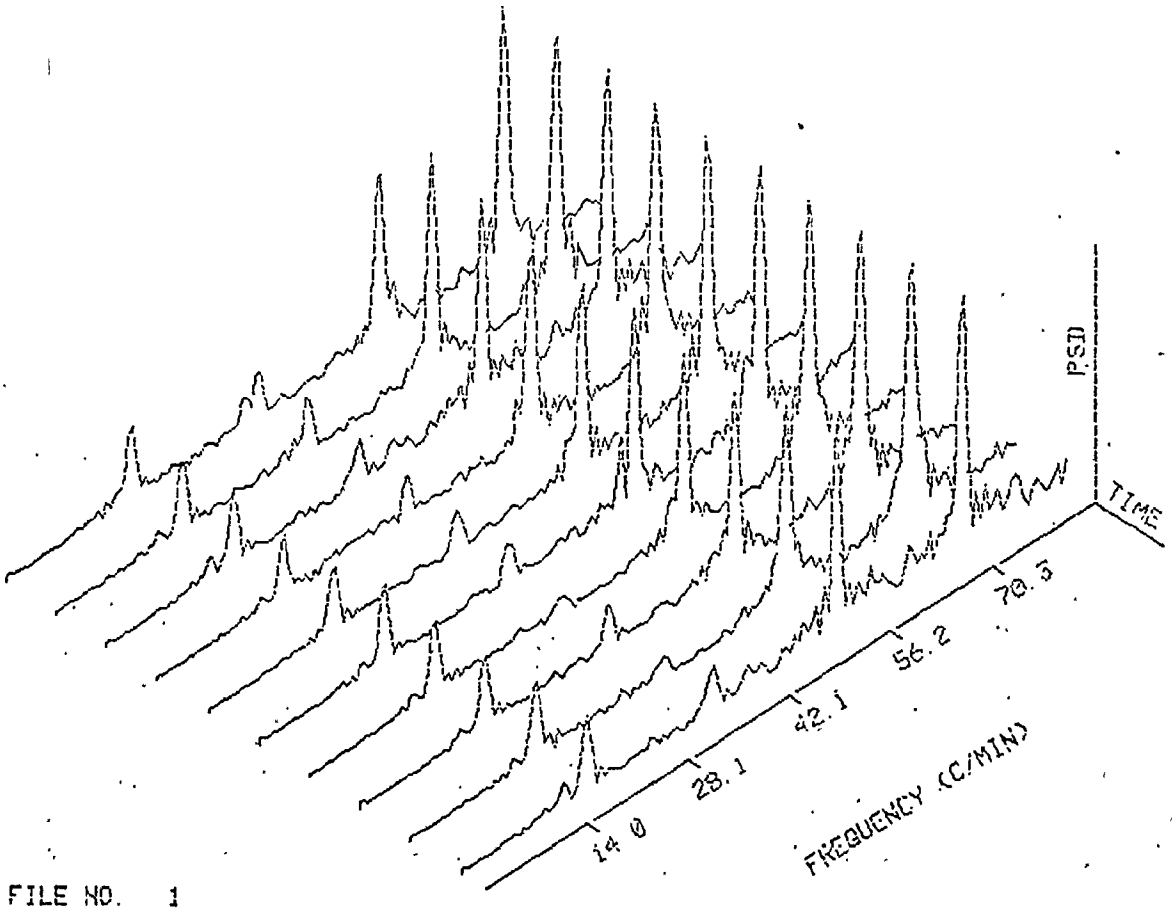


Fig 5.11 The record of the second example in section 5.4



FILE NO. 1
PERIODOGRAM
SEGMENT LENGTH = .8 MIN

Fig 5.12 Spectra of the record of fig 5.11 (periodogram)

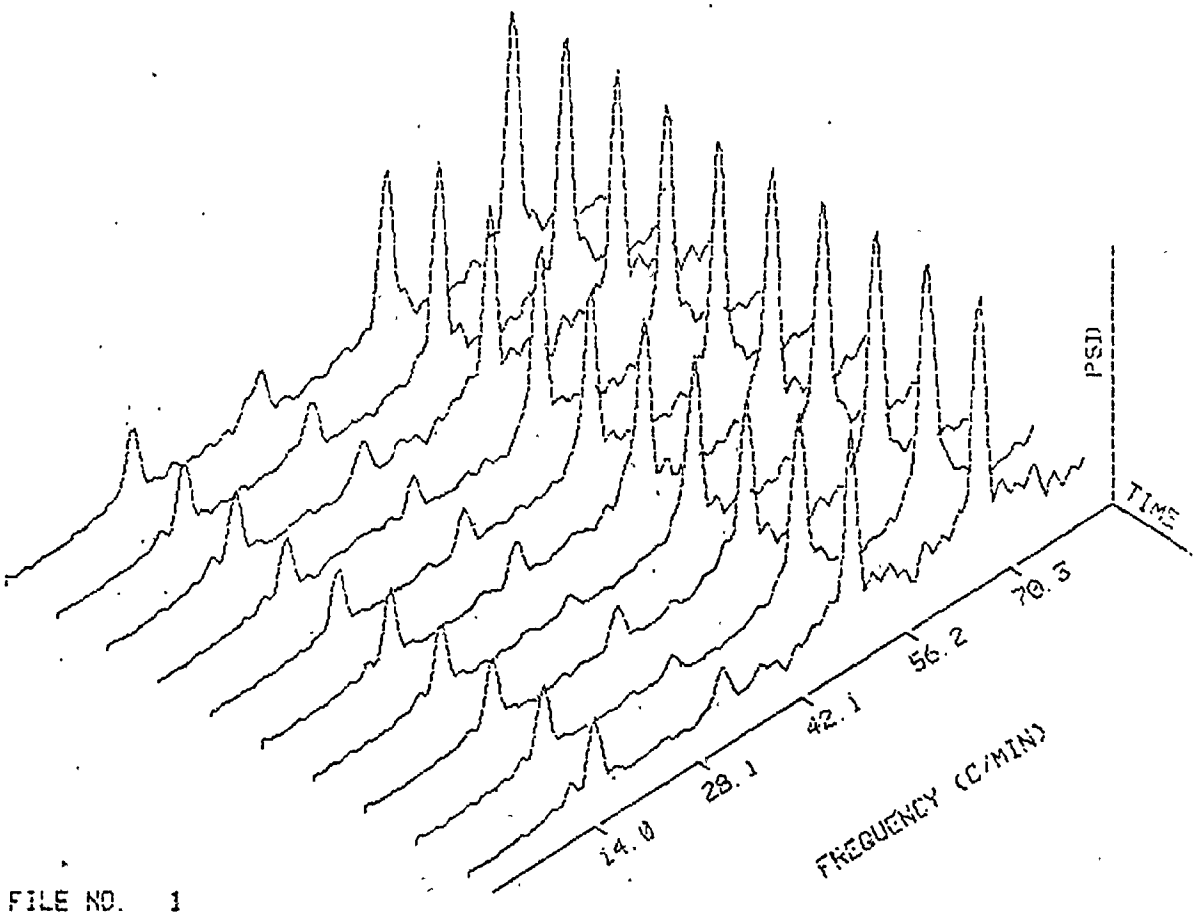
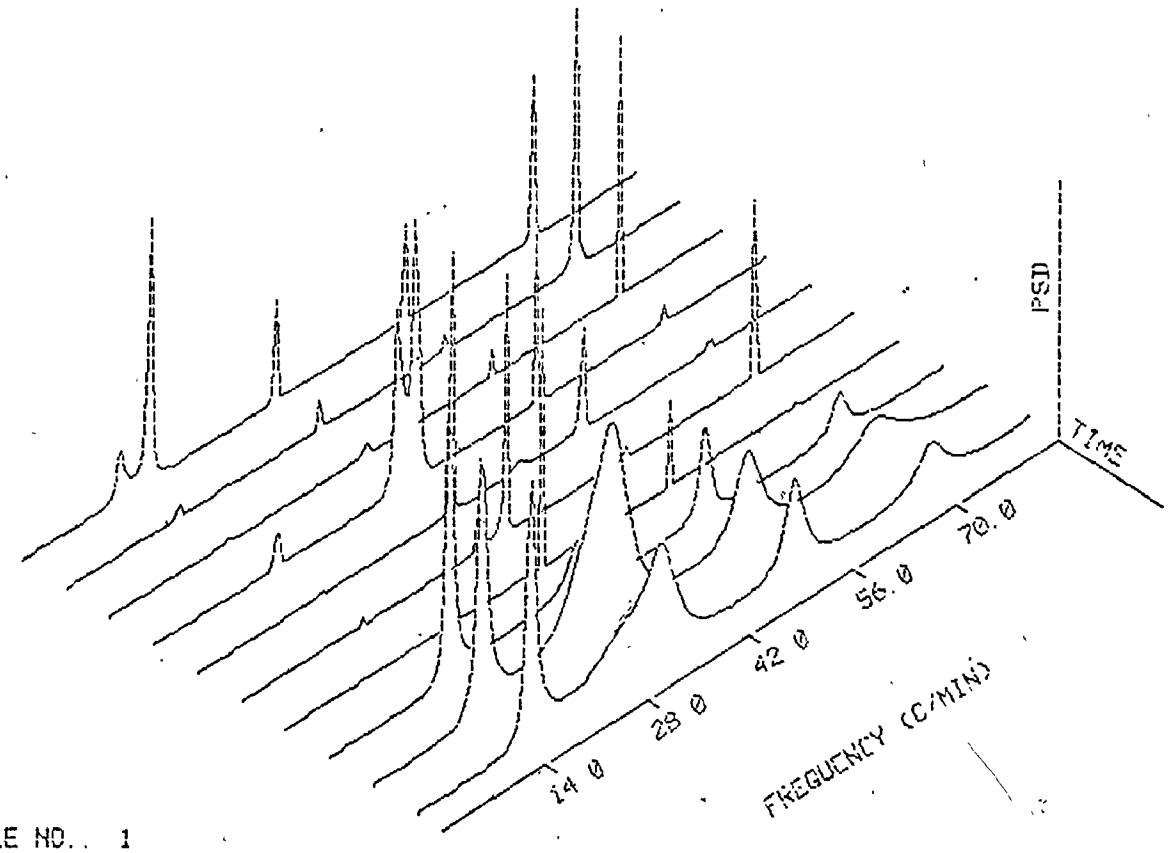


Fig 5.13 Spectra of the record in fig 5.11 (correlation method)



FILE NO.. 1
AR 40 SPECTRA
SEGMENT LENGTH = .0 MIN

Fig. 5.14 AR(40) spectra of the record of fig 5.7 showing line splitting.

5.6 Application of the One-Sided ARMA Method

The one-sided ARMA method was applied to the signal of the first example in this chapter. This gave rather an interesting result. That is, the ARMA spectral model has a better resolution than its autoregressive counterpart if all the frequency range is considered. In the example treated here, both methods agree in resolving the fundamental rhythm at about 17 c/min but the ARMA spectrum has sharper peaks at the harmonic frequencies. This can be seen by comparing the AR(20) and ARMA(16, 16) spectra in Figures 5.4 and 5.15, respectively. Also, the ARMA(16, 16) has much better resolution than the AR(16) of Figure 5.3. This improvement in resolution was obtained by including the 16 zeros of the MA part. In fact, inclusion of zeros may improve or worsen the resolution, depending on the locations of the zeros with respect to the poles of the periodic components. Kay in [40] has shown that the effect of noise on the autoregressive spectral estimates of noisy signals is caused by introducing spectral zeros. However, if the zeros are inherent in the signal, they should give a spectrum which is closer to the true spectrum. The illustration in Figure 5.16 shows how zeros may affect the resolution. A pole close to the unit circle will produce a broader peak than if it was embedded between two pairs of zeros, but if a zero is dislocated to coincide with the pole, it will completely eliminate the peak.

The deterioration in performance holds also for the ARMA method as the model order is reduced to (12, 12). This can be seen in Figure (5.17) where only a few segments in the middle still have good resolution;

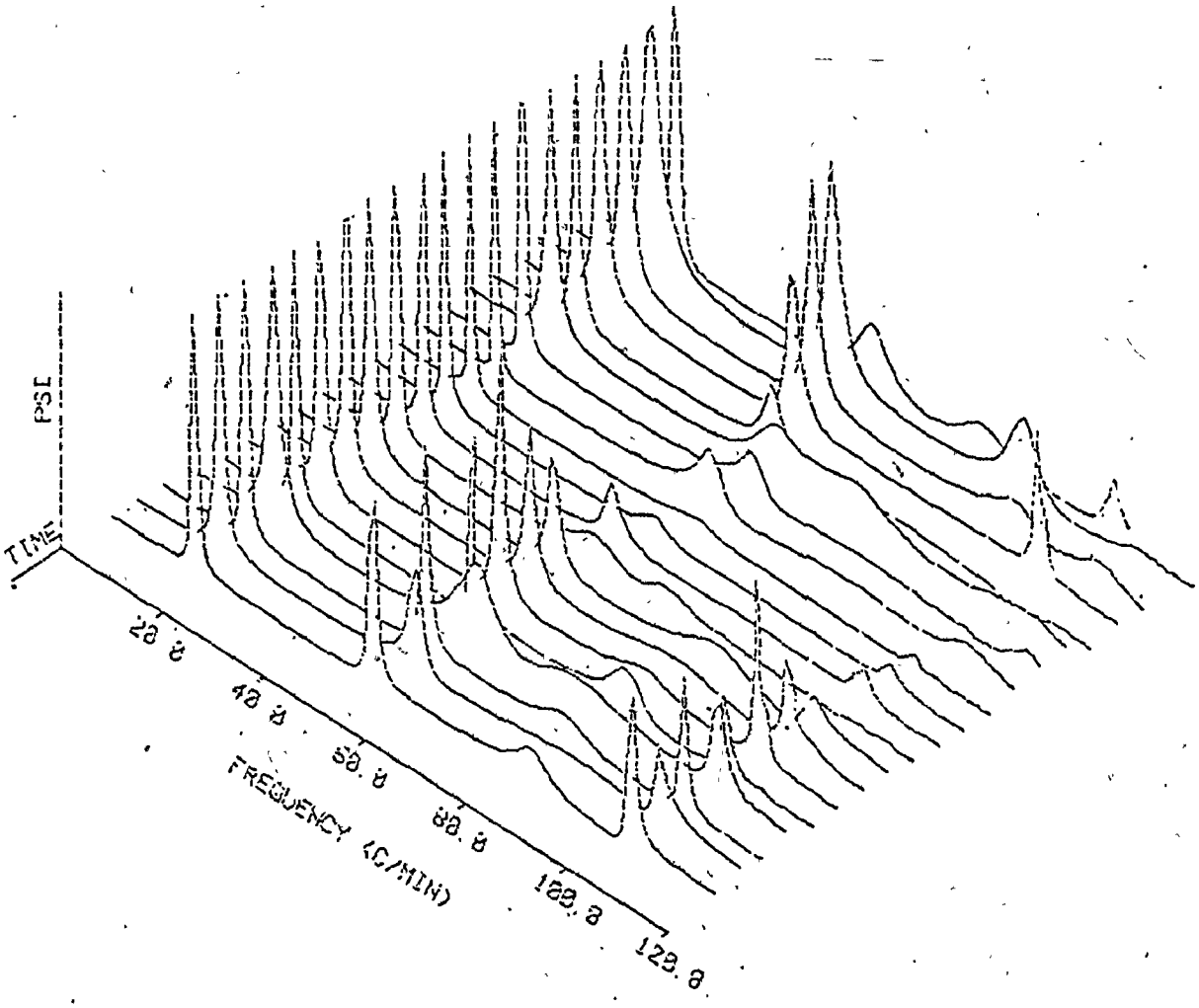
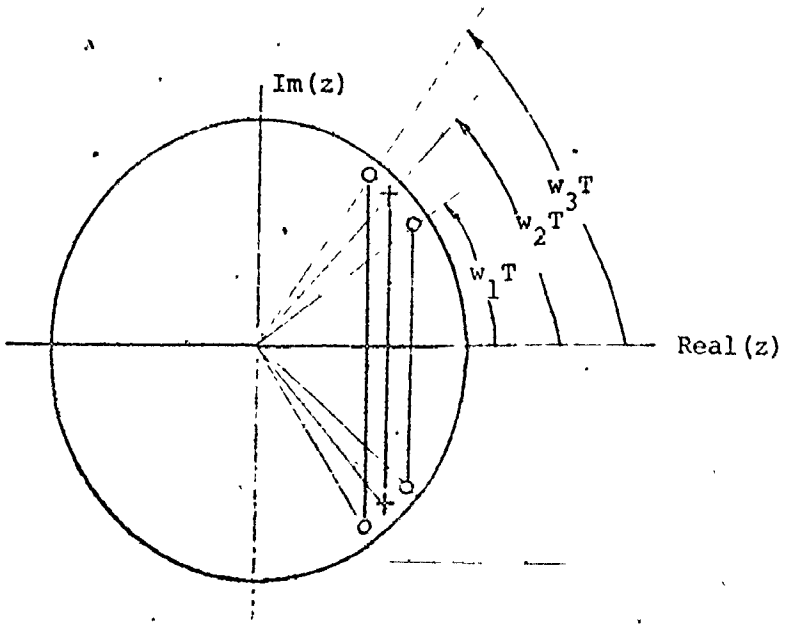
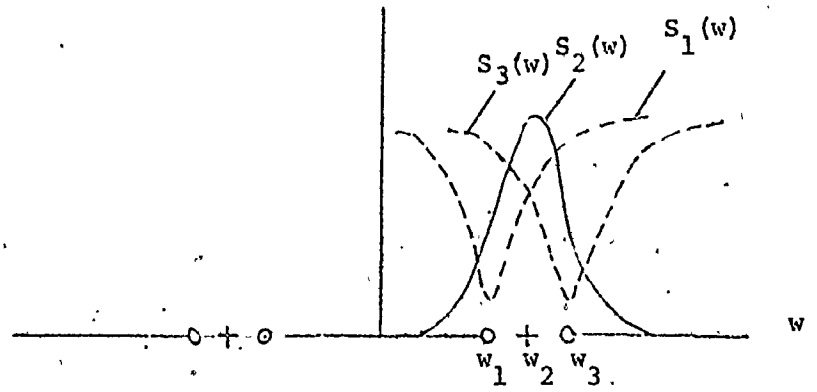


Fig 5.15 ARMA(16,16) spectra of the example of section 5.6





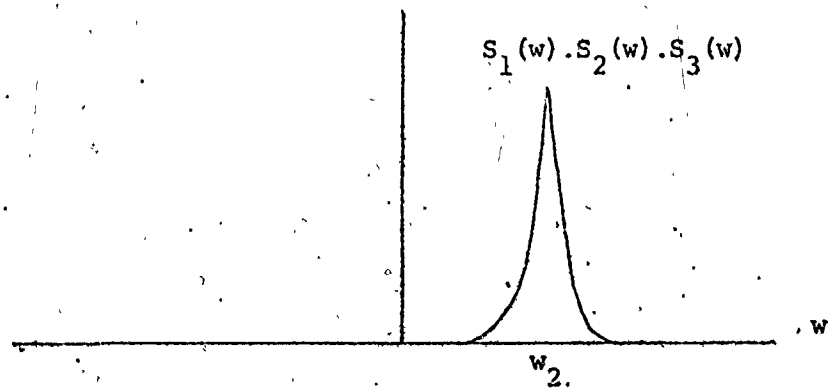


Fig 5.16 Effect of zeros on the resolution

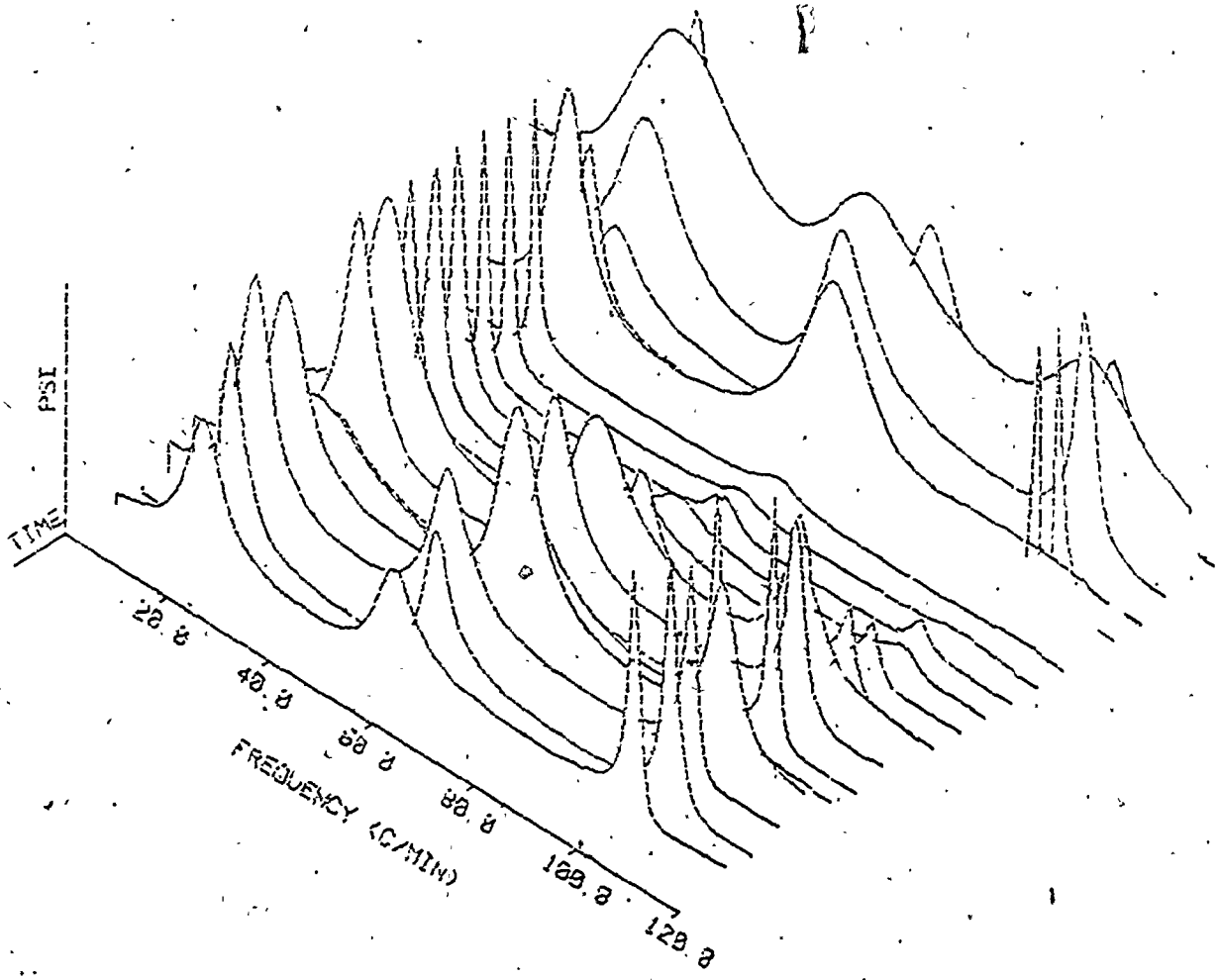


Fig 5.17 ARMA(12,12) spectra of the example of section 5.6

MARK ON ORIGINAL

whereas, all others have become very smooth. Roughly speaking, the estimation in the ARMA method is based on extracting the spectral poles to locate the periodic components, then estimating the zeros to provide additional spectral shaping. However, it suffers from the lack of a procedure for choosing a good order.

As mentioned in Chapter 3, the one-sided spectrum algorithm does not guarantee the positive definitness of the spectrum. An example for the negative estimates which may be obtained is shown in Figure 5.18. Seven segments among the analyzed twenty segments show reversed peaks with negative amplitudes. Almost all the negative estimates obtained were in the form of peak reversal or the existence of a new negative peak adjacent to the original peak.

5.7 Temporal Variation of the Small Intestinal ECA

All the examples presented so far were taken from signals recorded from proximal sites in the small intestine. They all exhibit a stationary fundamental rhythm at about 17 c/min. This agrees with the ECA characteristics in the frequency plateau region of the small intestine [2]. The spectra of a twenty minute record of a signal taken from the distal small intestine (170 - 180 cm from the pylorus) are shown in Figure 5.19. A typical one-minute segment before filtering is shown in the top of the figure. There is considerable variation in the ECA frequency with time at this site. Most signals, however, show large variations in the harmonic contents which implies a change in the waveform. This change does not necessarily have to be the same change

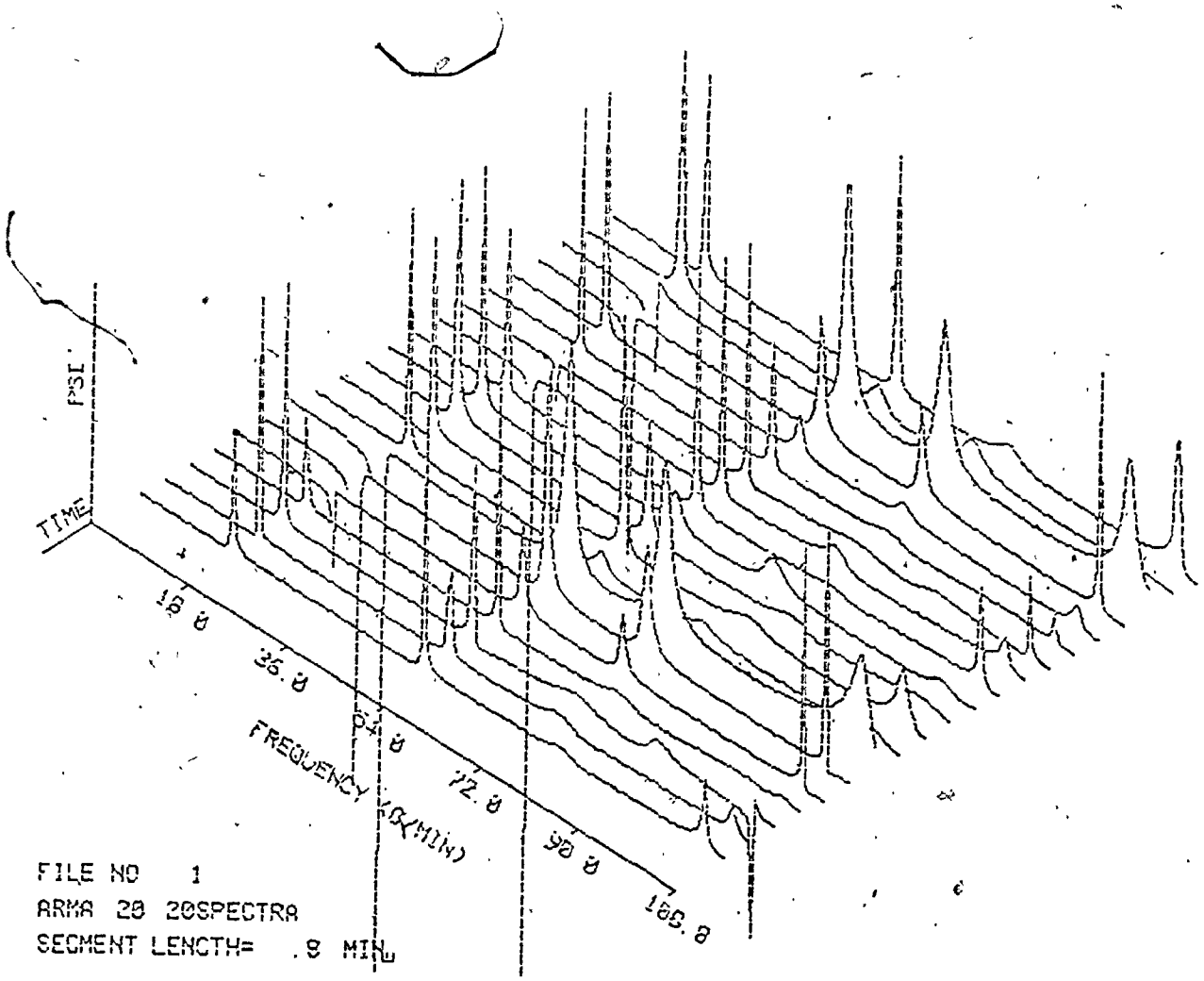


Fig 5.18 ARMA(20,20) spectra of the example of section 5.6

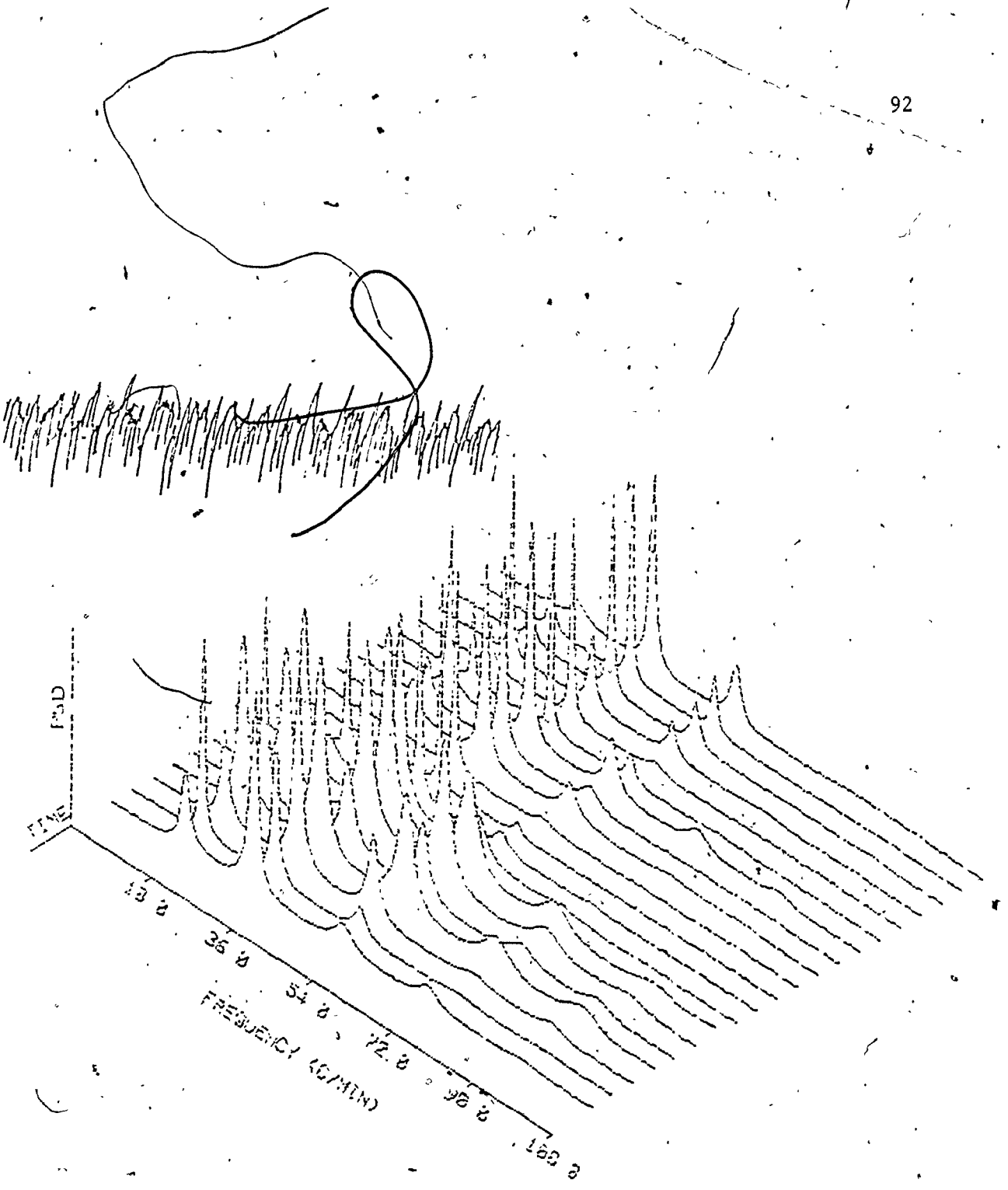


Fig 5.19 AR(28) spectra of the example of section 5.7 showing temporal variation of a distal record. A typical one minute segment is shown in the top.

across the membrane, but it could also be due to a change in the phase relationship of the ECA at adjacent sites or a change in the driving influence from proximal sites. The muscle contractions, however, occur at a rate equal to the fundamental frequency.

It should be noted that the peak position has some bias and statistical variance which depends on the algorithm used for estimating the autoregressive parameters. Two new algorithms invented by Marple [43] and Lang and McClellan [44] are reported by their inventors not to suffer much from large variations in the peak position.

5.8 Spatial Distribution of the Small Intestinal ECA

The ECA was analyzed for one minute of data recorded by all the 17 electrodes using the autoregressive method. The records analyzed are shown in Figure 5.20 after being filtered with a low-pass filter with a cutoff frequency of 4 Hz then sampled at a rate of 10 Hz. The sampled signals were plotted using the same incremental plotter used for plotting the spectra. The resulting spectra are shown in Figure 5.21. A frequency gradient from 17 c/min to 13 c/min was observed for this dog. The fundamental frequency was constant at the first few electrodes and then it became variable for the rest of the distal electrodes. The harmonics were present in an unpredictable fashion.

The spatial and temporal patterns of the ECA are major determinants of the small intestinal motility. The control waves in the proximal small intestine had a constant frequency within biological limits. This implies that the control waves at different sites in this

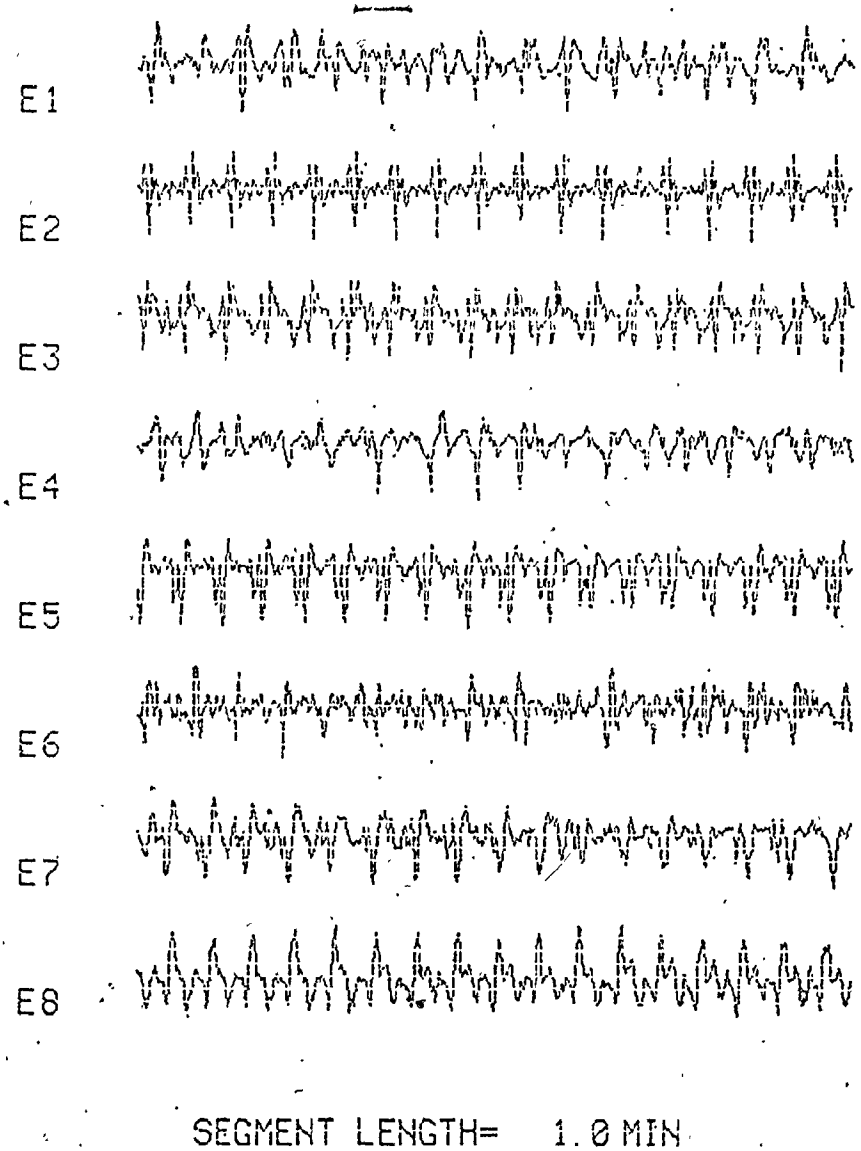


Fig 5.20 A one minute segment of all the 17 electrodes

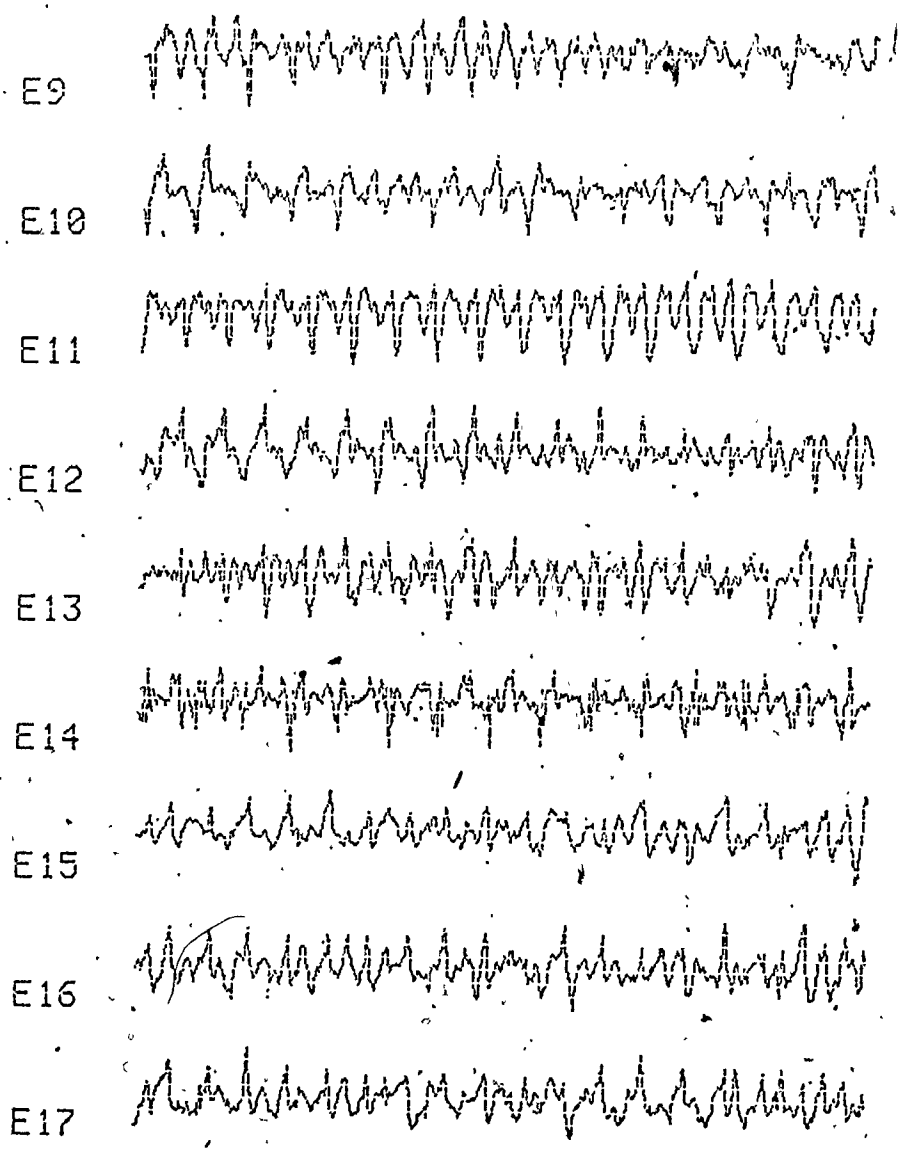
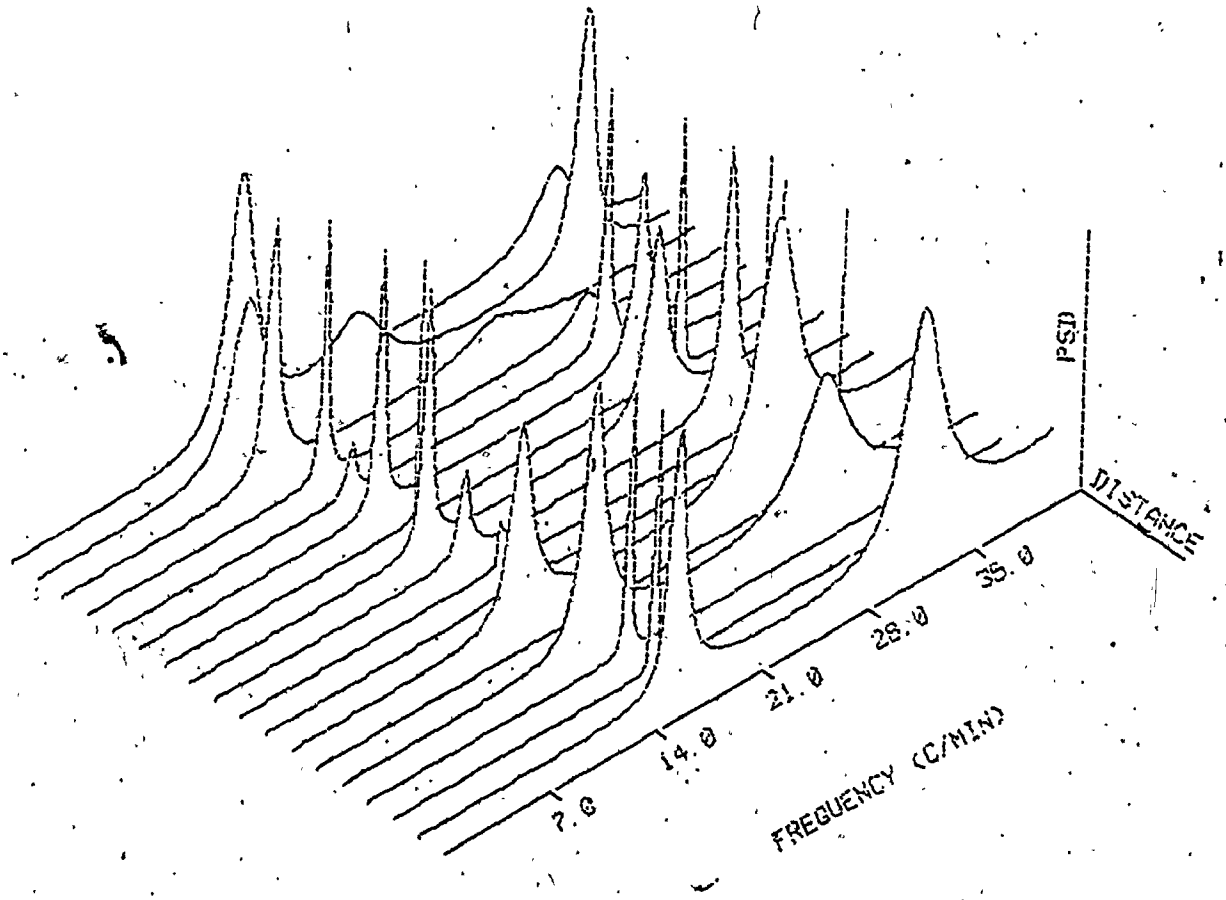


Fig 5.20 - Cont'd



AR(35) Spectra

Fig 5.21 Spatial distribution of the spectra of the small intestinal ECA

part of the small intestine were phase locked. This pattern of the ECA allows peristaltic contractions to occur which results in rapid propulsion which is desirable, since the duodenal contents must be removed rapidly and spread over the rest of the small intestine so that they do not act as an obstruction to further gastric emptying. In the distal small intestine, the ECA frequency was variable which means that the control waves were not phase locked. The contractions in this part of the small intestine would not, therefore, be peristaltic, resulting in a longer transit time.

There is, however, a frequency gradient for the ECA in this region which means that the frequency of contractions also has a gradient. The uncoordinated contractions with a distal frequency decrease may be responsible for largely mixing movements and slow distal propulsion.

CHAPTER 6

CONCLUSIONS

The achievements in this thesis can be categorized as follows:

1. Studying some of the currently available techniques for spectral estimation, with emphasis on the new ARMA methods which are still under investigation, where it has been shown that:
   i. Additive interference between autoregressive signals produces an ARMA process with equal MA and AR orders.
   ii. The two ARMA methods termed "cascade representation" and "one sided spectrum" methods are equivalent if the length of the sequence $\{c_k\}$ in the one-sided spectrum method is taken twice as long as the MA autocorrelation sequence in the cascade method.
2. The implementation of four methods in a minicomputer based general program for spectral estimation which provides many facilities for analyzing GI signals, such as:
   handling large amounts of data,
   easiness in use through an interactive dialogue,
   three-dimensional graphics of the spectra, and,
   a well-documented output.
3. Demonstrating the performance of the four methods by examples taken from the small intestinal ECA. Application of these methods has led us to the following conclusions:

The Periodogram: Is for deterministic periodic functions of time. Application of this method requires periodic extension of the data. It worked well with signals which were clearly periodic, but behaved badly with noisy signals where it showed large variations and spurious peaks.

The Autocorrelation Method: Produces smoother spectra than the periodogram does but on the account of the resolution. It needs a large number of samples to become effective. However, the low resolution feature is not serious in almost all the examples treated since the periodic components are well separated. Application in a noisy environment was not successful in identifying the rhythmic components.

The Autoregressive Method: The autoregressive approximation of the small intestinal ECA has been found to be reasonable up to order 36 for 512 samples at 10 Hz sampling rate. The order has been shown to be dependent on the sampling rate. Meanwhile, a sampling rate of 10 Hz was demonstrated to be unnecessarily high for some signals, whereas, a sampling rate of 5 Hz did not cause any aliasing. This should reduce the order to less than 20. Line splitting in the autoregressive spectral estimates occurred in only one case when the order was increased to 40. Application in a noisy environment lowered the resolution but successfully identified the rhythmic components in the case of well separated frequencies.

The One-Sided Spectrum ARMA Method: This method has been shown to have better resolution than the autoregressive method and at a lower order,

but with more computational effort. However, it suffers from two major drawbacks that need to be investigated further. These are:

i. The lack of a method for identifying the order.

ii. The occurrence of negative estimates, mostly in the form of peak reversal or the occurrence of a negative peak adjacent to the original peak.

The compensation for the negative estimates is a point that has to be studied.

The ARMA representation in general has been shown to be a good representation for the intestinal ECA assuming that the sources of this muscular activity are periodic with true autoregressive representation.

4. Investigating the temporal and spatial patterns of the small intestinal ECA. It was found that the ECA frequency was constant with time in the proximal small intestine and variable in time in the distal parts. The ECA was also found to have a constant frequency plateau region in the proximal small intestine and a frequency gradient which causes the frequency to decrease distally from 17 c/min to 13 c/min in the distal small intestine. These patterns are consistent with the motility function of the small intestine since the contraction pattern is such that rapid peristaltic contractions occur in the duodenum, then follows a region of uncoordinated contractions which have a net distal frequency decrease.

Suggestions for Future Work:

Through our practice with the mathematical basis for spectral estimation and applying it to the intestinal ECA, it is believed that the following points need to be investigated:

1. Applying the AR parameter estimation algorithms referred to in [17, 43, 44]. These methods avoid calculating the autocorrelation function and calculate the parameters from the data directly which is desirable when the order is a large fraction of the number of samples.
2. Studying the statistical properties of the estimates used where the variance problem becomes twofold:
   i. spectral amplitude variance
   ii. peak position variance
   The former is responsible for the periodogram spurious peaks and the latter causes dependence of the peak position on the initial phase in the AR method.
3. Studying the order determination problem and the negative estimates in the one-sided ARMA method.
4. Generalization to obtain a multivariable autoregressive representation of the Gastrointestinal ECA. This is particularly important regarding two points.
   i. as a spectral estimator
   ii. for studying the structural properties of the ECA spatial pattern to investigate the mutual effects between different parts in the digestive tract.

REFERENCES

1. T. Nelsen and J. Becker, "Simulation of the electrical and mechanical gradient of the small intestine", Am. J. Physiol. 214, No. 4, pp. 749-757, April 1968.

2. S. K. Sarna, E. E. Daniel and Y. J. Kingma, "Simulation of slow wave electrical activity of small intestine", Am. J. Physiol. 221, No. 1, pp. 166-175, July 1971.

3. S. K. Sarna, E. E. Daniel and Y. J. Kingma, "Simulation of the electrical control activity of the stomach by an array of relaxation oscillators", Am. J. Digest. Dis., Vol. 17, pp. 299-310, 1972.

4. S. K. Sarna, "Control mechanisms for gastric emptying", Proc. Symp. on Comp. Elect. and Control, Calgary, Alberta, Canada, pp. IV. 1-5, May 1973.

5. B. L. Bardakjian and S. K. Sarna, "Interactive processors for the analysis and modelling of kiological rhythms", Med. & Biol. Eng. & Comput., Vol. 18, pp. 149-200, 1980.

6. S. K. Sarna, "Gastrointestinal electrical activity: Terminology", Gastroentrology, Vol. 68, No. 6, pp. 1631-1635, 1975.

7. A. Papoulis, "Probability, Random Variables and Stochastic Processes", McGraw Hill, 1965.

8. B. C. Kuo, "Analysis and synthesis of sampled data control systems", Prentice Hall, 1963.

9. R. B. Blackman and J. W. Tukey, "The measurement of power spectra from the point of view of communications engineering", Dover Publications, New York, 1958.

10. R. K. Otnes and L. Enochson, "Applied time series analysis", John Wiley and Sons, 1978.

11. A. H. Gray and J. D. Markel, "A spectral flatness measure for studying the autocorrelation method of linear prediction of speech analysis", IEEE Trans., ASSP-22, pp. 207-216, June 1976.

12. R. V. Hogg and A. T. Craig, "Introduction to mathematical statistics", McMillan Publishing Co., New York, 1970.

13. A. V. Oppenheim and R. W. Schafer, "Digital signal processing", Prentice Hall, New Jersey, 1975.

14. C. K. Yuen, "Comments on modern methods for spectral estimation", IEEE Trans., ASSP-27, pp. 298-299, June 1979.

15. J. P. Burg, "Maximum Entropy Spectral Analysis", presented at the 37th Annual Meeting Soc. of Exploration Geophysicists, Oklahoma City, Oklahoma, 1967.

16. S. Haykin, Editor, "Nonlinear methods of spectral analysis", Springer-Verlag, New York, 1979.

17. J. P. Burg, "A new analysis technique for time series data", presented at the NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, August 12-23, 1968.

18. G. E. P. Box and G. M. Jenkins, "Time series analysis: Forecasting and control", Holden-Day, San Fransisco, 1976.

19. H. Akaike, "Power spectrum estimation through autoregressive model fitting", Ann. Inst. Statist. Math., Vol. 21, pp. 407-419, 1969.

20. A. Van den Bos, "Alternative interpretation of maximum entropy spectral analysis", IEEE Trans. Inform. Theory, Vol. IT-17, pp. 493-494, July 1971.

21. T. J. Ulrych and M. Ooe, "Autoregressive and mixed autoregressive-moving average models and spectra", in "Nonlinear methods of spectral analysis", S. Haykin, Editor, Springer Verlag, New York, 1979.

22. N. K. Sinha, "Modelling and Identification of Dynamic Systems", Course Notes, Electrical Engineering Department, McMaster University, Hamilton, Canada.

23. J. Durbin, "The fitting of time series models", Rev. Int. Inst. Stat., Vol. 28, pp. 233-244.

24. H. Akaike, "A new look at statistical model identification", IEEE Trans. Automatic Control, Vol. AC-19, pp. 716-723, Dec. 1974.

25. S. A. Tretter and K. Steiglitz, "Power spectrum identification in terms of rational models", IEEE Trans. Automatic Control, Vol. AC-12, pp. 185-188, April 1967.

26. L. H. Koopmans, "The spectral analysis of time series", Academic Press, New York, 1974.

27. J. F. Kinkel, J. Perl, L. Scharf and S. Subberud, "A note on covariance invariant digital filtering and autoregressive-moving average spectrum analysis", IEEE Trans. ASSP-27, pp. 200-202, April 1979.

28. M. Kaveh, "High resolution spectral estimation for noisy signals", IEEE Trans. ASSP-27, pp. 286-287, June 1979.

29. J. Cadzow, "High performance spectral estimation: A new ARMA Method", IEEE Trans. ASSP-28, pp. 524-529, Oct. 1980.

30. K. J. Astrom and P. Eykhoff, "System identification - A survey", Automatica, Vol. 7, pp. 123-162, 1971.

31. J. W. Bandler, "Computer-aided circuit optimization", Chp. 6 in "Modern filter theory", John Wiley and Sons, Inc., 1973.

32. W. Gersh, "Estimation of autoregressive parameters of a mixed autoregressive-moving average time series", IEEE Trans. Automatic Control, Vol. AC-15, pp. 593-585, 1970.

33. M. Morf, D. T. Lee, J. R. Nickolls and A. Vieira, "Classification of algorithms for ARMA models and ladder realizations", presented in the IEEE International Conference on Acoustics, Speech and Signal Processing, May 9-11, 1977.

34. S. M. Kay, "A new ARMA spectral estimator", IEEE Trans., ASSP-28, pp. 585-588, Oct. 1980.

35. J. Cadzow, "Autoregressive moving average spectral estimation: A model equation error procedure", IEEE Trans. Geoscience and Remote Sensing, Vol. GE-19, pp. 24-29, Jan. 1981.

36. A. S. Willsky, "Digital signal processing and control and estimation theory", MIT Press, Cambridge, Massachusetts, 1979.

37. Y. T. Chan, J. M. M. Lavoie, and J. B. Plant, "A parameter estimation approach to estimation of frequencies of sinusoids", IEEE Trans. ASSP-29, pp. 214-219, Apr. 1981.

38. Y. S. El-Cherif and S. K. Sarna, "Parametric spectral estimation of Gastrointestinal signals", to appear in the Proc. of the 3rd Annual Conference of the IEEE-EMBS, Frontiers of Engineering in Health Care, 19-20 Sept., 1981, Houston, Texas,

39. Y. S. El-Cherif, "Minicomputer implementation of a general program for spectral estimation", Internal Report GMRG-5, Dept. of Surgery, McMaster University Medical Centre, Hamilton, Ontario, Canada.

40. S. Kay, "The effect of noise on the autoregressive spectral estimator", IEEE Trans. ASSP-27, pp. 478-485, Oct. 1979.

41. S. M. Kay and S. L. Marple, Jr., "Sources and remedies for spectral line splitting in autoregressive spectrum analysis", in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, 1979, pp. 151-154.

42. R. W. Herring, "The cause of line splitting in Burg Maximum Entropy spectral analysis", IEEE Trans., ASSP-28, pp. 692-701, Dec. 1980.

43. L. Marple, "A new autoregressive spectrum analysis algorithm", IEEE Trans., ASSP-28, pp. 441-454, Aug. 1980.

44. S. W. Lang and J. H. McClellan, "Frequency estimation with maximum entropy spectral estimators", IEEE Trans., ASSP-28, pp. 716-725, Dec. 1980.