A study of the application of modern techniques to speech waveform analysis

This item was submitted to Loughborough University’s Institutional Repository by the/an author.

Additional Information:


Metadata Record: https://dspace.lboro.ac.uk/2134/28015

Publisher: © Khaldoon Atta Ghaidan

Rights: This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 2.5 Generic (CC BY-NC-ND 2.5) licence. Full details of this licence are available at: http://creativecommons.org/licenses/by-nc-nd/2.5/

Please cite the published version.
This item was submitted to Loughborough University as a PhD thesis by the author and is made available in the Institutional Repository (https://dspace.lboro.ac.uk/) under the following Creative Commons Licence conditions.

For the full text of this licence, please go to:
http://creativecommons.org/licenses/by-nc-nd/2.5/
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synopsis.</td>
<td>1</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>iii</td>
</tr>
<tr>
<td>List of symbols and Acronyms.</td>
<td>iv</td>
</tr>
<tr>
<td>Table of contents.</td>
<td>vi</td>
</tr>
<tr>
<td><strong>Chapter 1</strong></td>
<td></td>
</tr>
<tr>
<td>1.1 Introduction.</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Thesis summary.</td>
<td>3</td>
</tr>
<tr>
<td><strong>Chapter 2</strong></td>
<td></td>
</tr>
<tr>
<td>2.1 Speech production.</td>
<td>5</td>
</tr>
<tr>
<td>2.2 Speech parameters.</td>
<td>6</td>
</tr>
<tr>
<td>2.3 Acoustic analysis of speech.</td>
<td>12</td>
</tr>
<tr>
<td>2.4 Speech characteristics.</td>
<td>13</td>
</tr>
<tr>
<td>2.5 Spectrographs.</td>
<td>15</td>
</tr>
<tr>
<td>2.6 Spectrogram production.</td>
<td>16</td>
</tr>
<tr>
<td><strong>Chapter 3</strong></td>
<td></td>
</tr>
<tr>
<td>3. General Analysis.</td>
<td></td>
</tr>
<tr>
<td>3.1 Traditional methods.</td>
<td>17</td>
</tr>
</tbody>
</table>
A study of the application of modern techniques to speech waveform analysis

by Khaldoon Atta Ghaidan B.Sc. M.Sc.

A Doctoral thesis

submitted in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy of the University of Technology Loughborough.

May 1986.

Supervisor: Professor J.W.R. Griffiths.

Dept. of Electronics & Electrical Engineering.
University of Technology.
Loughborough.
England.

© by Khaldoon Atta Ghaidan
To BAN And YAZEN
SYNOPSIS

Spectrograms are perhaps the most commonly used method for studying the characteristics of speech waveforms. Producing a spectrogram can conveniently be divided into two parts, the analysis and the display, and this thesis describes a study of both these aspects.

Analysis is normally carried out using the swept filter technique or, in recent years, the Fast Fourier Transform. However more recently there has been much interest in what are called modern spectral estimation methods, methods which can offer significant advantages over traditional techniques. The use of such methods has been examined thoroughly and the results compared with the more traditional methods.

A recent commercial product for producing spectrograms is based on a swept filter technique but uses a complex digital filter. This system is also analysed and the results compared with the other methods.

The analysis was carried out on real speech data. A system based on the BBC micro-computer was developed to obtain the
digitised samples of real speech. The digitised speech could be processed on the BBC micro itself but more usually was fed to a main-frame computer in the computer centre (a Prime 750) for analysis.

Several different techniques for displaying the analysed data were examined, some using the Prime graphic facilities and others which were based on the BBC micro. The processed data from the Prime could be fed back to a Framestore which was interfaced to the BBC micro so enabling the spectrogram to be viewed on a video monitor. Hard copies of the spectrograms could be obtained on a colour printer connected to the BBC micro.
Acknowledgements

It gives me great pleasure to express my deep gratitude to my supervisor, Professor J.W.R. Griffiths, for his guidance and encouragement throughout this work. His invaluable suggestions and comments were of great help in the development of this work.

Many thanks are due to the Government of the Republic of Iraq for providing the Scholarship.

My thanks to the Loughborough Sound Images Ltd. for the prints of the spectrogram produced on their LSI-sound spectrograph.

I would like to thank my colleagues in the department, Avtar Gida for his valuable comments and suggestions, Dr. T. Chen for reading the draft, C. Carey-Smith and F. Lee for the useful discussions during the research.

I would like to thank Mrs. S. Clarson for helping out with part of the typing.

Special thanks are due to my friend Marwan Al-Akaidi for his help with the diagrams.

I wish to express my sincere thanks to my wife Ban for her constant encouragement, for without her patience it would not have been possible to see this work through to completion. To her and our son Yazan I dedicate this work.
# LIST OF SYMBOLS AND ACRONYMS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Input sample</td>
</tr>
<tr>
<td>@</td>
<td>Filter coefficient</td>
</tr>
<tr>
<td>XF</td>
<td>Fourier Transform of X</td>
</tr>
<tr>
<td>e</td>
<td>Error</td>
</tr>
<tr>
<td>R(n)</td>
<td>Auto-correlation function</td>
</tr>
<tr>
<td>R'(n)</td>
<td>Estimated auto-correlation</td>
</tr>
<tr>
<td>F,f</td>
<td>Frequency f;</td>
</tr>
<tr>
<td>Y^T</td>
<td>Transpose of Y</td>
</tr>
<tr>
<td>j</td>
<td>Square root of -1</td>
</tr>
<tr>
<td>ESD</td>
<td>Energy Spectral Density</td>
</tr>
<tr>
<td>PSD</td>
<td>Power Spectral Density</td>
</tr>
<tr>
<td>\alpha</td>
<td>Filter real coefficient</td>
</tr>
<tr>
<td>\beta</td>
<td>Filter imaginary coefficient</td>
</tr>
<tr>
<td>Mn</td>
<td>Mean value</td>
</tr>
<tr>
<td>*</td>
<td>Conjugate pair</td>
</tr>
<tr>
<td>'**'</td>
<td>Estimated conjugate</td>
</tr>
<tr>
<td>s</td>
<td>Complex frequency plane</td>
</tr>
<tr>
<td>z</td>
<td>z-plane</td>
</tr>
<tr>
<td>S,S',S''</td>
<td>Residual sum of squared errors</td>
</tr>
<tr>
<td>MEM</td>
<td>Maximum Entropy Method</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum Likelihood</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>LP</td>
<td>Linear Prediction</td>
</tr>
<tr>
<td>AR</td>
<td>Auto-Regressive</td>
</tr>
<tr>
<td>YW</td>
<td>Yule-Walker</td>
</tr>
<tr>
<td>$\lambda_{\text{min}}$</td>
<td>Minimum eigen value</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>Final residual power</td>
</tr>
<tr>
<td>LS</td>
<td>Least Square</td>
</tr>
<tr>
<td>PHD</td>
<td>Pisarenko Harmonics Decomposition</td>
</tr>
<tr>
<td>FFT</td>
<td>Fast Fourier Transform</td>
</tr>
<tr>
<td>CFT</td>
<td>Continuous Fourier Transform</td>
</tr>
<tr>
<td>DFT</td>
<td>Discrete Fourier Transform</td>
</tr>
<tr>
<td>CPU</td>
<td>Central Processing Unit</td>
</tr>
<tr>
<td>D.R</td>
<td>Dynamic Range</td>
</tr>
</tbody>
</table>
3.1.1 The swept-filter method. 18
3.1.2 The FFT methods.
   3.1.2.1 Indirect FFT. 21
   3.1.2.2 Direct FFT (Periodograms). 22
   3.1.2.3 The use of windows. 23
   3.1.2.4 Segmental averaging. 25
3.1.3 Complex filters.
   3.1.3.1 Complex filters from Real filters. 29
   3.1.3.2 Complex pole representation. 32
3.2 The Modern methods. 39
   3.2.1 Maximum entropy method (MEM) 39
      3.2.1.1 The Levinson recursion. 43
      3.2.1.2 The Burg algorithm. 46
      3.2.1.2 The Anderson method. 48
   3.2.2 The autoregressive method. 52
      3.2.2.1 The Yule-Walker estimates. 56
   3.2.3 Linear prediction method. 60
      3.2.3.1 The Algorithm. 63
      3.2.3.2 Forward algorithm. 64
      3.2.3.3 Backward algorithm. 64
      3.2.3.4 Both fwr and bkwr algorithm. 65
   3.2.4 Pisarenko harmonics decomposition 66
Chapter 4

4.1 Speech recording.

4.2 Computer Simulations and flow-charts.
   4.2.1 Traditional methods.
      4.2.1.1 The complex filter.
      4.2.1.2 The FFT methods.
   4.2.2 Modern methods.
      4.2.2.1 The maximum entropy method.
      4.2.2.2 The autoregressive method.
      4.2.2.3 The linear prediction method.
      4.2.2.4 The Pisarenko method.

Chapter 5

5.1 The display.
   5.1.1 The Graphics display.
   5.1.2 The Framestore display.

5.2 Discussions.

Chapter six

6.1 The Discussions.
   6.1 Evaluation of the Bessel filter.
   6.2 Comparison of the FFT methods.
   6.3 Evaluation of the modern methods.
Chapter 7

7.1 Conclusions. 126

7.2 Suggestions for further works. 129

References. 130

Appendix -A- Notes on the Framestore 137
1.1 Introduction

Spectrograms are the most commonly used methods of studying the characteristics of speech waveforms. A spectrogram is a three dimensional representation of speech with time, frequency and intensity as the three parameters. The production of a spectrogram can be divided into two main parts, the analysis and the display. Traditionally the analysis is carried out using the swept filter technique where the recorded speech is repeatedly played through a band-pass filter having a fixed pass-band with the centre frequency being moved in steps along the frequency range. Recently the use of the Fast-Fourier Transform has been introduced to carry out speech analysis. The most recent commercial system for producing spectrograms is the LSI-sound spectrograph which employs the swept filter technique using a complex digital filter. A model of the complex filter method was simulated using a Bessel fourth-order low-pass complex digital filter. Modern spectral estimation techniques have been firmly established and they can offer significant advantages over some of the traditional methods. A review of some of these
methods was carried out and their use for speech analysis was evaluated.

Real speech records were used in the production of the spectrograms. A system based on the BBC micro interfaced by a special circuit to an A/D-D/A board and a Framestore, was used to produce digitised speech sampled at 8 kHz which was fed back and stored on the main computer Prime 750 located at the Loughborough University Computer Centre. Short records of 2 sec. duration were used to produce spectrograms by the traditional and modern methods of spectral analysis.

The graphic facilities of the Prime 750 were used to display the analysed data of the different spectrograms and the results were investigated.

A system based on the BBC micro interfaced to a Framestore was set up to display the analysed data. The processed data was quantized and then fed back from the Prime 750 to the Framestore through the BBC micro where it can be viewed on a video display. Hard copies of the spectrograms can be obtained on a colour printer connected to the BBC micro.

A spectrogram of the 2-sec. of speech was produced on the LSI-commercial system using a complex digital filter. The spectrogram was then used as the reference to compare and evaluate the different spectrograms.
1.2 Thesis summary

Chapter 1 is an introduction to the work carried out in this thesis and a summary of the thesis.

Chapter 2 includes a brief description of the speech production by the human mechanism, the speech acoustical properties and the speech characteristics. Also a note on spectrograms and their production is included.

Chapter 3 is a literature survey of the spectral estimation techniques used for time series analysis. These methods are discussed under two main headings.

The first is a survey of the traditional methods which means that they have been used for speech spectrographs, such as the swept filter method and the FFT methods. An outline of the complex digital filter theory is also included.

The second is a survey of the modern methods such as the Maximum Entropy, Linear Predication, Auto-Regressive and Pisarenko Harmonics decomposition.

Chapter 4 describes the system that was set-up to record real speech samples. The arrangement used is based on an interface of the BBC microcomputer to the framestore and the A/D-D/A board in order to produce digitized speech samples. The simulations of the analysis methods are briefly
outlined. A set of the flowcharts for the different computer programs is included in this chapter.

Chapter 5 describes the different methods of displaying the spectrogram that are investigated. A brief description of the arrangement using the BBC micro and the Framestore.

Chapter 6 include discussions on the different stages of the work. The discussions are conducted in three stages. The first stage is to evaluate the computer simulation model and to justify the use of the Bessel lowpass filter for the complex digital filter by comparing the different results with that of the LSI-sound spectrograph.

The second is to compare the different methods of using the FFT for the analysis with that of the complex filter method.

The third is conducted to evaluate the use of the modern methods for the spectrogram by comparing the results to that of the complex filter method. A comparison between the modern methods and the FFT methods is also included.

Chapter 7 is conclusions on the work carried out with suggestions for further work.

Appendix A contains technical data on the B.T framestore. The Framestore was used in the systems for recording the speech and for the display of the spectrogram.
CHAPTER TWO

Speech is man's primary method of communication. He has developed the vocal means for coding and conveying information beyond a rudimentary stage. Wave sounds are produced by the vocal apparatus and consist of rapid and significantly erratic fluctuations in air pressure.

2.1 Speech Production

Fig. (2.1) shows a schematic diagram of the human speech production mechanism. In normal speech production, the chest cavity expands and contracts to force air from the lungs out through the trachea past the glottis. If the vocal cords are tensed they will vibrate in the mode of a relaxation oscillator, modulating the air into discrete puffs or pulses. If the cords are spread apart, the air stream passes through the glottis and is unaffected. The air stream then passes through the pharynx cavity, past the tongue, and depending on the position of the trap door velum, through the mouth and/or nasal cavity. The air stream is expelled at either the mouth or the nose, or both and is perceived as speech.
FIG 2.1
A SCHEMATIC DIAGRAM OF THE HUMAN SPEECH PRODUCTION MECHANISM
2.2 **Speech parameters**

The sound pressures which are generated and radiated by the vocal apparatus come under two main categories, so called "voiced" and "unvoiced" sounds. Voiced sounds are produced by forcefully expelling air from the lungs through the vocal cords. The vocal cords "chop" the airflow, producing a repetitive pressure impulse. The spectrum of the waveform is as shown in fig. (2.2a). It is a substantially flat line spectrum having components at the fundamental and harmonics of the vocal cord vibration frequency. Above a few kilohertz the amplitude of the harmonics falls about 12 dB per octave.

The frequency can be consciously controlled, and is an important speech parameter, known as the speech "pitch". Its variation imparts "intonation" to the utterance, or emphasis. In any given individual, pitch varies typically by only a fraction of an octave during normal speech, and has a frequency of about 50 Hz in a deep voiced male up to about 500 Hz in a small child.

The pressure waveform at the larynx is therefore the source of "voiced" sounds, and is known as the excitation. The different voiced sounds such as "oo", "ah" and "ee" are produced by varying the position of the lips, jaw, tongue and velum. This complex shaped cavity has resonances at
FIG. 2.2a
MAGNITUDE SPECTRUM OF
"VOICED" EXCITATION

FIG. 2.3 a
MAGNITUDE SPECTRUM OF
"UNVOICED" EXCITATION

FIG. 2.2 b
TYPICAL MAGNITUDE SPECTRUM
OF "VOICED" SOUND
(EXCITATION MODIFIED BY VOCAL
TRACT RESONANCES )

FIG. 2.3 b
TYPICAL MAGNITUDE SPECTRUM
OF "UNVOICED" SOUND
(EXCITATION MODIFIED BY VOCAL
TRACT RESONANCES )
frequencies determined by its instantaneous shape. The broadband excitation signal is therefore modified to give (usually) a spectrum of four resonances, fig. (2.2b). These resonances are called "formants" and their frequency, amplitude and bandwidth determine the "voiced" sound produced.

The unvoiced sounds differ in not using the vocal cords. Air from the lungs is expelled through a constriction which increases the air speed to the point where turbulent flow occurs. This produces a noise like pressure waveform or excitation which again has a substantially flat but continuous spectrum fig. (2.3a). This waveform is again modified by the vocal tract resonances, producing differing "unvoiced" sounds, such as "ff", "ss", "sh" and "s" as in snow. Generally fewer resonances are involved in "unvoiced" sounds fig. (2.3b).

The basic parameters described, voiced, unvoiced, excitation, pitch (if voiced) and formant (or spectrum) structure are to a first approximation, independent linear processes which fully define the utterance. We may therefore approximate the voice production process by the block diagram shown in fig (2.4)\textsuperscript{18}. There are two further aspects that should be noted. The physical limitations of the vocal tract imposes severe constraints on what sounds can be produced. The human voice cannot duplicate any
FIG. 2.4
LINEAR MODEL OF THE SPEECH PRODUCTION PROCESS
arbitrary audio sound. There are limits on how quickly we can change the sound being uttered. As a general rule no more than ten distinguishable sounds per second can be made, and even for these the four main speech parameters tend to change smoothly and continuously. No more than a couple of widely differing sounds per second can be uttered.

These limitations are extremely important when we consider the rate at which information can be sent using speech. Shannon's Information theorem states that a 3 kHz wide-signal having a signal/noise ratio of 30 dB can transmit information at a maximum rate of about 30 kbit/s. But it has been shown that from considering both the limitations of the physical process and the text equivalent of speech, the actual information rate of speech is of the order of 50 bit/s. The reason for the enormous disparity is that the speech signal is highly redundant, i.e. the essential information is in effect, repeated in the signal many times, so that it can be altered, parts can be deleted, it can be distorted and yet it can be still understood. This is fortunate as otherwise we could not communicate over a telephone line prone to clicks, we could not talk in a noisy environment or in a room having an echo, or transmit clear speech over a selective fading multiple skywave HF radio circuit. It is the redundancy of speech which makes it such an effective and robust means of communication. But
there are also further levels of redundancy associated with language. We can understand perfectly well even if a whole word is lost or misprinted. The redundancy of speech arises from its structure, namely the repetition, organisation and pattern of the signal.

2.3 Acoustic analysis of speech

During any stretch of natural speech the condition in the larynx and the vocal tract are changing continuously, the period of the opening and closing of the vocal folds varies literally from cycle to cycle and the articulators are in movement all the time, never maintaining the same configuration for much more than one tenth of a second. Hence the main characteristics of the acoustic output of the speech mechanism is its variation with time and any technique for examining this output has to provide means of following this variation if it is to be really useful. One of the reasons for the wide spread use of the sound spectrography in acoustic phonetics is that in its most commonly used form it does this.

An equally important matter, of course, is what measurement or quantities one is to follow the variations of. The waveform of speech sounds, that is, the record of air particles displacement in the path of the wave, contains within itself all the available information about the sound,
and if it is registered in suitable conditions it provides an accurate record of time variations in speech. This information is not very easily accessible, however, until it has been broken down in some stages of acoustic analysis. The choice of analytical operations has been greatly influenced by our view of the action of the human hearing mechanism. The ear is generally looked upon as being basically a harmonics analyser, which is to say that, when it is presented with a complex sound, it analyses it into the various harmonics or at least the narrow bands of which it is composed and the impression which the brain receives of the sound is largely determined by the relative amounts of sound energy carried in successive harmonics or bounds throughout the whole of the frequency range.

2.4 Speech characteristics

Vowels and other voiced sounds possess periodic or rather quasi-periodic waveforms and accordingly display harmonic spectra. This fine structure originates from the opening and closing movements of the vocal cords periodically modulating the volume of the exhaled air during phonation at a rate of F0 Hz which is the voice fundamental frequency. In the narrow-band spectrograms F0 is the harmonic spacing and in the broad-band spectrograms 1/F0 is the time interval between successive sections each reflecting a single voice
cycle.
The train of successive air-pulses emerging from the vibrating glottis is the primary source of voiced sounds. The air cavities with the vocal tract act as a multi-resonant filter on the transmitted sound and impress upon it a corresponding formant structure. The frequencies of the three formants, F1, F2, F3 are the main determinants of the phonetic quality of the vowel.

The resonance frequencies of the vocal tract, F1, F2, F3, F4 conceptually contained in the term F-pattern, vary continuously across often sharply time localised breaks in the spectrographic time-frequency-intensity picture. Such breaks may for instance indicate shifts from voice to noise source or vice-versa.

Spectrographic pictures convey an overview which are non-essential for descriptive purposes. This redundancy is in part a matter of inter-relations, repetitions, and continuities within the signal structure, in part the presence of a fine structure the details of which carry very little or no information. When processing the spectrographic data on connected speech the first object is to identify the boundaries of successive sound segments. A sound segment generally carries information on more than one phoneme of a sequence. Conversely each phoneme may be physically encoded to a smaller or greater extent in the
pattern aspect of several adjacent sound segments. The number of successive sound segments of a piece of connected speech is generally larger than the number of phonemes. For example, stop sounds can be considered to be made up of at least two typical sound segments, the occlusion and the burst and the latter phase may in some instances be split up into three successive and partly overlapping phases, the explosion transient, a short fricative, and an "h" sound. The description of a sound segment for the purpose of identification may be based on the following parameters, previously mentioned and summarised as:

a-Duration
b-Intensity
c-Energy (Area under Intensity-Time curve)
d-Voice fundamental frequency FO
e-The F-pattern structure (frequency-intensity distribution)
f-The fine structure, referring to the speech production source (voiced, unvoiced, muttered or silence).

2.5 Spectrograms

The speech waveform is complex in its frequency and also varies rapidly in time. Numerous methods have been employed in the past to try to show changing energy-frequency distribution. One such method is to analyse the data by a set of bandpass filters. The output of each band is
rectified and recorded so that it shows the variation of amplitude with time.
The most common representation is the three dimensional representation known as the spectrogram, with Time, Frequency and energy level (intensity) as the three main parameters. This type of representation is most efficient for the study of speech characteristics as it highlights the important acoustic and perceptual features such as formant structure, voicing, friction, stress and pitch.

2.6 Spectrogram production

The production of spectrograms involves two main operations. The first is the analysis of the speech record by one of the spectral estimation methods discussed in chapter 3. The processed data represents the interrelations between time, frequency and power that exist within the speech waveform. The second part is the display of the processed data in such a way that it will convey all the essential information relevant to the study of speech, eg., speech characteristics, phonemes etc.
CHAPTER THREE

3. General Analysis

The spectral analysis of the speech waveform is the main part of producing a spectrogram. The swept filter method is the most widely used method for spectrogram production. Also, the Fast Fourier transform is used in some sona-graphs. Most recently, a system based on the use of complex digital filter scanning technique has been produced commercially. These methods are reviewed under the heading traditional methods.

Modern spectral analysis techniques such as the Maximum Entropy Method MEM, the Auto-Regressive AR, the Linear Prediction LP, and Pisarenko harmonic decomposition have been fully utilised for use in different fields. A theoretical survey of these methods is carried-out in this section.

3.1 Traditional methods

The word traditional indicates that these methods have been actually used for spectrogram production. The swept filter technique is the most widely used method in the sona-graphs.
The Fast Fourier Transform has been used in a spectrograph system for speech analysis. Most recently the commercial type LSI-sonagraph does the analysis by using the swept filter technique with the filter being a complex digital.

3.1.1 The swept filter method\textsuperscript{18,20}

A speech sample is recorded on a magnetic disc and repeatedly played through a bandpass filter and is, in effect, scanned slowly across the frequency band of the signal. The result is therefore equivalent to an analysis by many such filters. This is practically accomplished by modulating the signal onto a high frequency carrier and sliding one side band of the signal past the fixed bandpass filter. The frequency of the carrier is varied in steps in order to translate the filter scan. The passband of the fixed filter is either the narrow band at 45 c/s or the wide band at 300 c/s.

3.1.2 The FFT Methods

Traditional spectrum estimation\textsuperscript{26} as currently implemented using the FFT is characterised by many trade-offs in an effort to produce statistically reliable spectral estimates. There are many trade-offs in windowing, time-domain averaging, frequency-domain averaging of sampled data.
obtained from a random process in order to balance the needs to reduce side-lobes, to perform effective ensemble averaging, and to ensure adequate spectral resolution.

Consider the deterministic analogue wave $X(t)$ that is a continuous function of time for $t=1,\ldots,n$.

If $X(t)$ is absolute integrable; i.e, the signal energy $E$ is finite

$$E = \int_{-\infty}^{\infty} |X(t)|^2 dt < \infty$$  \hspace{1cm} (1)

then the continuous Fourier Transform (CFT) $XF(f)$ of $X(t)$ exists and is given by

$$XF(f) = \int_{-\infty}^{\infty} X(t) \exp(-j2\pi ft) dt$$  \hspace{1cm} (2)

The squared modulus of the Fourier Transform is often termed the spectrum energy, $E(f)$ of $X_t$

$$E(f) = |XF(f)|^2$$  \hspace{1cm} (3)

Parseval's energy Theorem, expressed as

$$\int_{-\infty}^{\infty} |X(t)|^2 dt = \int_{-\infty}^{\infty} |XF(f)|^2 df$$  \hspace{1cm} (4)

is a statement of the conservation of energy; the energy of the Time-domain signal is equal to the energy of the Frequency-domain transform $E(f)df$. Thus $E(f)$ is an energy spectral density (ESD) in that it represents the distribution of energy as a function of frequency. If the process $X(t)$
is a wide sense stationary, stochastic process rather than a deterministic, finite-energy waveform, then the energy of such processes are usually infinite, so that the quantity of interest is the power (time average of energy) distribution with frequency. Also integrals such as (2) normally do not exist for a stochastic process. For the case of a stationary random process, the auto-correlation function,

\[ R(k) = \mathbb{E}[X(t+k) \cdot X^*(t)] \]  \hspace{1cm} \text{(5)}

provides the basis for spectrum analysis, rather than the random process \( X(t) \) itself. The Wiener-Khinchinn theorem relates \( R(k) \) via the Fourier Transform to \( S(f) \), the PSD,

\[ S(f) = \int (R(k) \exp(-j2\pi fk) dk) \]  \hspace{1cm} \text{(6)}

As a practical matter the statistical auto-correlation function is unknown. Thus an additional assumption is that the random process is ergodic in the first and second moments. This property permits the substitution of time averages of ensemble averages. For an ergodic process, then, the statistical auto-correlation function may be equated to

\[ R(k) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} [X(t+k) \cdot X^*(t)] dt \]  \hspace{1cm} \text{(7)}

It is possible by using (7) to show that (6) may be
expressed as

\[ S(f) = \lim_{T \to \infty} \left[ \frac{1}{2T} \int_{-T}^{T} [X(t) \exp(-j2\pi ft) dt]^2 \right] \tag{8} \]

Difficulties may arise if (8) is applied to finite data sets without regards to the expectation and limiting operations. Statistically inconsistent (unstable) estimates result if no statistical averaging is performed; i.e. the variance of the PSD estimate will not tend to zero as T increases without bound.

3.1.2.1 Indirect FFT method (auto-correlation)\textsuperscript{26,38}

With a finite data sequence, only a finite number of discrete autocorrelation function values, or lags, may be estimated. Blackman and Tuckey proposed the spectral estimate,

\[ S(f) = \Delta t \sum_{m=-M}^{M} R'(m) \exp(-j2\pi fm\Delta t) \tag{9} \]

based on the available autocorrelation lag estimates \( R'(m) \), where \(-1/2\Delta t < f < 1/2\Delta t\) and ('') denotes estimate. This spectral estimate is the discrete-time version of the Wiener-Khinchinn expression. An obvious auto-correlation estimate based on (7) is the unbiased estimator,

\[ R'(m) = \frac{1}{N-m} \sum_{n=m}^{N-1} x_n x_n^* \tag{10} \]

\( n=0 \ldots N-m-1 \) for \( m=0 \ldots M \) where \( M < N-1 \).
The negative lag estimates are determined from the positive lag estimates as follows:

\[ R'(-m) = R'^*(m) \]  \hspace{1cm} \text{(11)}

in accordance with the conjugate symmetric property of the autocorrelation function of a stationary process. Instead of equation (10) a slightly modified estimate,

\[ R'(m) = \frac{1}{N} \sum_{n=0}^{N-m-1} X(n+m)X^*(n) \]  \hspace{1cm} \text{(12)}

defined for \( m=0 \ldots M \) which tends to have less mean square error than (10) for many finite data sets. \( R'(m) \) is a biased estimator since \( E(R'(m)) = ((N-m)/N)R(m) \). The mean value is a triangular window weighting of the true auto-correlation function.

The periodogram(spectrum) is the estimate by the use of equation (8).

3.1.2.2 Direct FFT method (Periodograms)\textsuperscript{26,38}

The direct method of spectrum analysis is the modern version of Schuster's periodogram. The Fourier Transform of the data record is computed for \( N \) number of frequencies along the frequency range. The power spectral estimation is then computed by taking the magnitude squared of the FFT output. Equation(8) is used to compute the power output of the periodogram. This method can be viewed as a least square
fit of a harmonic set of complex sinusoids to the data.

3.1.2.3 The use of Windows

Often a periodogram of N data samples is computed using (8) when the measured process has deterministic components embedded in white noise. Care must be taken since statistically inconsistent result can occur if (10) is used literally without regard to the expectation operation. This need some sort of ensemble averaging, or smoothing of sample spectrum. The use of windows in the Time domain or the Frequency domain or both can reduce the side-lobe effects. Many of the problems of periodogram PSD estimation can be traced to the assumptions made about the data outside the measured interval. The finite data sequence may be viewed as being obtained by windowing an infinite length sample sequence with a boxcar function. The use of only this data implicitly assumes the unmeasured data to be zero, which is usually not the case. This multiplication of the actual time series by a window function means the over-all transform is the convolution of the desired transform with the transform of the window function. If the true power of a signal is concentrated in a narrow bandwidth, this convolution operation will spread that power into adjacent frequency regions. This phenomena, termed leakage, is consequence of the tacit of windowing inherent in the
computation of the periodogram.

In addition to the distorting effect of leakage on the spectral estimate,\textsuperscript{26} leakage has a detrimental impact on power estimation and detectability of sinusoidal components. Sidelobes from adjacent frequency cells add in a constructive or destructive manner to the main lobe of a response in another frequency cell of the spectrum, affecting the estimate of power in that cell. In extreme cases, the sidelobes from strong frequency components can mask the main lobe of weak frequency components in adjacent cells.

Data windowing is also the fundamental factor that determines the frequency resolution of the periodogram. The convolution of the window transform with that of the actual signal transform means that the most narrow spectral response of the resultant transform is limited to that of the main-lobe width of the window transform, independent of the data. For rectangular window,\textsuperscript{22} the main-lobe width between 3-dB levels of the resulting $(\sin Nf)/Nf$ transform is approximately the inverse of the observation time of $N\Delta t$ seconds. Other windows may be used, but the resolution will always be proportional to $1/N\Delta t$ Hz. Leakage effects due to data windowing can be reduced by the selection of windows with non-uniform weighting.
3.1.2.4 Segmental Averaging

One method for reducing the variance of the power spectrum estimate is segmental averaging. In this case the data record is decimated into independent segments, and the auto-correlation function of each segment is estimated from which the average is calculated. The power spectrum estimate is the Fourier transform of the average of the auto-correlation function. The expected value of these power spectrum estimates is biased which is easily seen in the frequency domain where the expected value of the spectral estimate is the convolution of the spectrum of the segmental window with the true or actual power spectrum. The bias of these estimates exceeds that of the periodogram of the unsegmented or complete data record since the data segments are shorter than the complete data record, thus implying that the main lobe of the segmental spectral window is broader than the window used for the periodogram. However, the variance of these estimates is less than the variance of the periodogram by a factor equal to the number of segments if the segments can be considered uncorrelated. But these estimates have less resolution than the periodogram of the entire record.

Another procedure used in power spectrum estimation for reducing the variance of the estimate is the Bartlett method which calculates the periodogram of individual
segments of the data record. These periodograms are then averaged.

Welch's\textsuperscript{43} method segments the data, windows each segment, calculates the periodogram of each windowed segment and then determines the average periodogram. The data segments may be overlapped. It is shown that the variance of the spectral estimate is reduced, but of course at the expense of the lost spectral resolution.

Let $X_n \ n=0..N-1.$ be a sample from a stationary, second order stochastic sequence and let $E(X)=0.$ Let $X_n$ have a spectral density $P(f)$ where $-1/2<f<1/2$

We take segments, possibly overlapping, of length $L$ with starting points of these segments $G$ units apart,

let $X_1(n)=X(n)$ \hspace{1cm} $n=0...L-1$

$X_2(n)=X(n+G)$ \hspace{1cm} $n=0...L-1$

$\vdots$

\hspace{5cm} \text{(13)}

$X_k(n)=X(n+(K-1)G)$ \hspace{1cm} $n=0...L-1$

Suppose that these segments from $X_1$ to $X_k$ cover the entire record, i.e.

$(K-1)G+L=N$
This segmenting is as follows

For each segment of length $L$ we calculate a modified periodogram. We select a window $W(n)$ $n=0..L-1$, and form the sequences,

$X_1(n)W(n) \ldots \ldots X_k(n)W(n)$.

We then take the finite Fourier Transforms

$X_{F1}(m) \ldots \ldots X_{Fk}(m)$

where

$$X_{Fk}(m) = \frac{1}{L} \sum X_k(n)W(n)e^{-j2\pi nm/L}$$
Finally the K periodograms

\[ I_k(f_m) = \frac{L}{U} \sum_{k=1}^{K} |X_F(k)|^2 \]

where \( f_m = \frac{m}{L} \) \( m = 0, \ldots, \frac{L}{2} \)

and \( U = \frac{1}{L} \sum W^2(n) \) \( n = 0, \ldots, L-1 \)

The spectral estimate is the average of these periodograms, i.e.

\[ P'(f_m) = \frac{1}{K} \sum I_k(f_m) \ \ \ \text{for} \ \ k = 1, \ldots, K \]

where

\[ h(f) = \frac{1}{L} \sum W(n) \exp(-j2\pi fn) |^2 \]

and \( \int_{-1/2}^{1/2} h(f) df = 1 \)

Hence we have a spectral estimator \( P'(f) \) with a resultant spectral window whose area is unity and whose width is of the order of \( 1/L \).
3.1.3 Complex Filters

These are filters\textsuperscript{14} which can be used with complex signals. They are derived from the transformation of Real low-pass prototype filters and are characterised by the fact that their coefficients are complex numbers. These complex coefficients arise because the filter is not constrained to have negative frequency transmission which is the conjugate of its positive transmission, i.e., the impulse response is not constrained to be real.

3.1.3.1 Complex-filter from real filter

Consider a real continuous filter characteristic which centres at \( w=0 \). The centre frequency is shifted to a new frequency \( w_c \), without introducing a conjugate filter at \( -w_c \) as would be the result of conventional low-pass into band-pass transformation. Also there is no constraint to have \( w_c \) greater than the filter band-width. Mathematically, the shift is accomplished by replacing \( s \) by \( s-jw_c \) in the real filter transfer function \( H(s) \). The equivalent to multiplying the impulse response by \( \exp(jw_c t) \). To relate this \( s \)-plane transfer along the \( jw \)-axis to a \( z \)-plane shift of a digital filter whose transfer function is \( H'(z) \), we use the mapping relationship \( z=\exp(sT) \), where \( T \) is the sampling
A typical frequency response of a Bessel 4th-order complex digital filter of 400-Hz bandwidth (BW) at 0-centre frequency and at 2000Hz shifted centre frequency.
period, to determine a shifted z ($z_{shifted}$) that will be used to replace z in $H'(z)$. By substituting $(s-j\omega c)$ for $(s)$ we get,

$$z_{shifted} = \exp(s-j\omega c)T = \exp(-\omega c T)z$$

Thus the shift may be accomplished by replacing z by $\lambda z$ in the transfer function $H$, where $\lambda$ is a complex constant of magnitude 1 defined by the relationship,

$$\lambda = \exp(-j\omega c T)$$

To see what this shift means in relation to the z-plane poles and zeros, consider the typical pole zero for a real filter. Each pole-zero is described by the relation $z-a=0$, where the location of the pole zero is $(a)$. Replacing $z$ by $\lambda z$ and dividing by $\lambda$ we get

$$(z-\lambda^{-1}a) = 0$$

The shifted pole is at $\lambda^{-1}a$ which is the location of the previous pole rotated by $\omega c T$ radians. It can also be considered as a modulation of the complex filter input by a frequency $\omega c$ radians per second, to translate the centre frequency of the complex envelope to zero frequency, the centre of the low-pass prototype. To complete this analogy the filter output would have to be modulated by a frequency of $\omega c$ to translate it back to its proper centre frequency. To implement the rotation of real sampled data filters we note that all linear, time invariant, finite, sampled data filters may be written as a ratio of polynomials, i.e.
The shifted transfer function $H'(z)$ is therefore also a ratio of polynomial,

$$
\frac{a'_0 + a'_1 z + \cdots + a'_n z^n}{b'_0 + b'_1 z + \cdots + b'_m z^m}
$$

where the new polynomial coefficients ($'$) have been derived from the old relationships.

$$
a'_i = \frac{i}{1} a_i \quad \text{and} \quad b'_i = \frac{i}{1} b_i
$$

In actually implementing the difference equations for the filters we need to express the transfer function in terms of polynomials in the delay operator $(z^{-1})$,

$$
\frac{c_0 + c_1 z^{-1} + \cdots + c_p z^{-p}}{d_0 + d_1 z^{-1} + \cdots + d_q z^{-q}}
$$

The recursive difference equation for simulating the digital filter having the transfer function is

$$
Y_m = \sum_{i=p}^{q} Y_{m-i}(1/d_0) \sum_{j=0}^{p} c_{j} X_{m-j} \quad i=1..q \quad \text{and} \quad j=0..p
$$

where $X$ and $Y$ are the filter inputs and outputs respectively. Applying the same argument, we find the coefficients of the shifted filter to be determined by,

$$
c'_i = \frac{i}{1} c_i \quad \text{and} \quad d'_i = \frac{i}{1} d_i
$$
3.1.3.2 Complex pole representation

A single-pole complex filter, with a compensating zero at the origin, is described by the flow chart in fig(3.1).

and has a sampled-data transfer function

\[
H(z) = \frac{z}{z - z_o} = \frac{1}{1 - z_o z^{-1}} \tag{21}
\]

The response to a unit pulse at the time origin is

\[
h(n) = z_o^n = \exp((\pi n T + j\omega n T) \tag{22}
\]

Expressing \( h(n) \) as a sum of real and imaginary parts, we get from equation (22) that

\[
h(n) = h_{\text{real}}(n) + jh_{\text{imag}}(n)
\]

\[
= \exp(\pi n T) \cos(\omega n T) + j \exp(\pi n T) \sin(\omega n T) \tag{23}
\]

If the chart is expanded to show signal flow for the real and imaginary parts of the complex signal then fig(3.2) is obtained.
FIG. 3.1

FIG. 3.2
To simulate this filter for one sampling interval requires one multiplication for each amplifier and one addition for each adder input. A total of 4-multiplications, 6-additions and 2-delays are required for one representation of a complex pole.

The equations defining the filter outputs as implemented in fig(3.3) are:

\[
Y_1(N) = X_1(N) + \alpha Y_1(N-1) - \beta Y_2(N-1)
\]

\[
Y_2(N) = X_2(N) + \beta Y_1(N-1) + \alpha Y_2(N-1)
\]
By taking the z-transform of both equations, we get the following equations

\[ Y_1 = X_1 + \alpha z^{-1} Y_1 - \beta z^{-1} Y_2 \]
\[ Y_2 = X_2 + \beta z^{-1} Y_1 - \alpha z^{-1} Y_2 \]  

(24)

where \(X\) is an input and \(Y\) is an output, \(\alpha\) and \(\beta\) are defined as the real and the imaginary parts of the filter coefficients. \(z^{-1}\) represents a delay of one sampling period.

By arranging equations (24) we get

\[ Y_1 = H_1(z) \cdot X_1 - H_2(z) \cdot X_2 \]
\[ Y_2 = H_1(z) \cdot X_2 - H_2(z) \cdot X_1 \]  

(25)

where

\[ H_1(z) = \frac{1 - \alpha z^{-1}}{1 - 2\alpha z^{-1} + (\alpha^2 + \beta^2) z^{-2}} \]
\[ H_2(z) = \frac{\beta z^{-1}}{1 - 2\alpha z^{-1} + (\alpha^2 + \beta^2) z^{-2}} \]  

(26)

If the constraint

\[ X_2 = jX_1 \]

is imposed then equation (25) becomes,

\[ Y_1 = (H_1(z) - jH_2(z)) X_1 = H(z) \cdot X_1 \]
\[ Y_2 = (H_1(z) - jH_2(z)) X_2 = H(z) \cdot X_2 \]  

(27)

and thus,
\[ Y_2 = jY_1 \]

hence if the inputs to the filter are equal in magnitude, but quadrature in phase, the output will also exhibit this property. In addition, the transfer function of the filter can be written

\[ H(z) = H_1(z) - jH_2(z) \]

\[ \frac{1}{1 - (\alpha - j\beta)z^{-1}} \]

for a single complex pole which has the properties

\[ \beta = r \sin(\theta) \quad \text{and} \quad \alpha = r \cos(\theta) \]

\[ H(z) = \frac{1}{1 - r \exp(j\theta)z^{-1}} \]

where \( r < 1 \) for a stable filter, that is, the pole is inside the unit circle on the \( z \)-plane. By substituting \( z^{-1} = \exp(-sT) \) and considering the steady state so that \( s = j\omega \) then,

\[ H(j\omega) = \frac{1}{1 - r \exp(j(\theta - \omega)T)} \]

and

\[ H(j\omega) = \frac{1}{1 - 2r \cos(\theta - \omega) + r^2} \]

The pole at \( z = 1/r \exp(j\theta) \) transfers to
\[ \exp(sT) = \frac{1}{r} \exp(je) \quad \text{or} \quad s = \frac{1}{T} \left( \log \left( \frac{1}{r} \right) + je \right) \]

giving \( r = \exp(-wcT) \) and \( e = woT \) where

- \( wc \) = half the bandwidth
- \( wo \) = center frequency
- \( T \) = sampling period

and

\[ H(jw) = \frac{1}{1-r} = \frac{1}{1-\exp(-wcT)} \quad \text{for} \quad w = wo \]  -----(31)

Several interesting properties may be deduced from the expressions, resulting from the filter with only one complex pole when driven by a quadrature input signal.

(a) The centre frequency gain is a function of the bandwidth only.

(b) The response is symmetrical, that is, the 3-dB points are arithmetically equi-distant from the centre frequency.

(c) Computation of the coefficients are straightforward. However the actual input is a real signal

\[ X_1 \quad \text{and} \quad X_2 = 0 \]

from equation (25),

\[ Y_1 = H_1(z) \cdot X_1 \quad \text{and} \quad Y_2 = H_2(z) \cdot X_1 \]

Ideally

\[
\begin{vmatrix}
H_1(z) \\
\hline
H_2(z)
\end{vmatrix}
= 1
\]

and the angle

\[
\frac{H_1(z)}{H_2(z)} \quad \text{is} \quad \pi/2
\]
In order to generate correctly ordered outputs which then can be applied to the next section, thus satisfying the previous constraint for the second and all succeeding sections.

\[
\begin{align*}
H_1(z) &= \frac{1-r\cos(\theta)z^{-1}}{r\sin(\theta)z^{-1}} \\
H_2(z) &= \frac{1-rcos(\theta)[\cos(wT)-jsin(wT)]}{rsin(\theta)[\cos(wT)-jsin(wT)]}
\end{align*}
\]

as \(wT \rightarrow 0\) and \(r \rightarrow 1\), that is the in-band response of a narrow bandwidth filter both the gain and the phase relationships approach the ideal.

Each section of the filter assists in extending the region for which the outputs approach the ideal until after four sections, the region extends for many bandwidths. The rectified output that is formed by summing the squares of both outputs which therefore, should ideally not contain any second order harmonics, is in fact smooth over the pass-band. Ripple becomes progressively a larger proportion of the signal as the response of the filter falls.
3.2 The Modern Methods

The modern techniques of spectral estimation has progressed through many stages since the discovery of the Fast Fourier Transform. They include methods such as the maximum entropy method MEM, the Auto-Regressive method AR, the auto-regressive moving average ARMA, Linear prediction method LP, Pisarenko harmonics decomposition PHD, and Prony's method. A review of some of these methods is carried out in this section.

3.2.1 The Maximum Entropy Method (MEM)

Given a record $X_n$ of length $N$ from a stationary time series of zero mean, the sample auto-correlation function $R(k)$ can be computed as

$$R(k) = \frac{1}{N} \sum_{i=1}^{N-k} X_i X_{i+k} \quad 1=1, \ldots, N-k \quad ----- (32)$$

The Fourier transform of $R(k)$ is an estimate of the spectral density $S(f)$. The entropy rate of the process is,

$$H = \frac{1}{2} \log[(S(f))] df \quad ----- (33)$$

where $S(f) = \sum_{-\infty}^{\infty} R(k) \exp(-j2\pi kfT)$

Here $T$ is the sampling period and it is assumed that the time series is essentially limited to the frequency band $-B<f<B$. Because of the finite number of samples, only a finite number say $2K+1$ with $K<N$ of auto-correlation values
are known. We can maximize equation (32) with respect to the unknown $R(k)$, $k > K + 1$ subject to the constraint that the spectral density denoted by $S'(f)$ must be consistent with the known auto-correlation.

$$R(0), R(1), \ldots, R(K-1) \quad \text{i.e.}$$

$$R(k) = \int_{-B}^{B} S'(f) \exp(-j2\pi fkT) df \quad 0 \leq k \leq K$$

Here the estimated $S'(f)$ called the maximum entropy spectral density estimate, expresses maximum uncertainty with respect to the unknown information but is consistent with the known information.

$$S'(f) = \frac{P_m}{B[1 + \sum_{k=1}^{m} \hat{a}_k \exp(-j2\pi fkT)]^2} \quad -(34)$$

Where $P_m$ and $\hat{a}_k$ are the output power and the filter coefficients of a prediction error filter.

Suppose that the time series $X_n$ of length $N$ is applied as an input to a linear digital filter of impulse response $h_n$ of length $m$. Let $y_n$ and $d_n$ represent the resulting output and the desired output respectively which both have length $m+N+1$. The error between the filter output and the desired output is given by,

$$e_n = d_n - y_n$$

The equation that minimizes the mean square value of the
error is known in the matrix form as

\[ R \cdot h_0 = R_d \]  \hspace{1cm} \text{(35)}

where \( h_0 \) is called the Wiener filter and

\[
\begin{bmatrix}
R(0) & \cdots & R(-m) \\
\vdots & \ddots & \vdots \\
R(m) & & R(0)
\end{bmatrix}
\begin{bmatrix}
h_0^1 \\
h_0^2 \\
\vdots \\
h_0^m
\end{bmatrix}
= \begin{bmatrix}
R_d(0) \\
R_d(1) \\
\vdots \\
R_d(m)
\end{bmatrix}
\]

with

\[ R(k) = E(X_n \cdot X_{n-k}^*) \]

\[ R_d(k) = E(d_n \cdot X_{n-k}^*) \]

The minimum value of the error power is

\[ R_{\min} = R_d(0) - R_d \cdot h_0 \]  \hspace{1cm} \text{(36)}

Consider a special case of Wiener filter that predicts the sample of \( X_n \) one unit time ahead i.e.

\[ d_n = X_{n+1} \]

Then the cross-correlation between the input and the desired output becomes,

\[ R_d(k) = E(X_{n+k+1} \cdot X_n^*) = R(k+1) \]

equation (36) can be written as

\[ \sum_{i=1}^{m} h_0^i \cdot R(k-i) = R(k) \quad i=1..m \quad k=1,2,\ldots,m \]  \hspace{1cm} \text{(37)}
With the output of the predictive filter delayed by one time unit as

\[ e_n = d_n - y_n = x_n - \sum h_0, k \cdot x_n - k \cdot e_k, x_n - k \quad k = 0 \ldots m \]  

Equation (38) defines a prediction error filter (PFE) of order \( m \) with input \( x_n \) and output \( e_n \). The filter coefficients \( e_k \) is defined by

\[ e_0 = 1, \quad e_k = -h_0, k \quad k = 2, \ldots, m \]  

For one step prediction considered above, \( d_n = x_{n+1} \) and hence

\[ R_d(0) = R(0) \]

and \( (R^*(1), R^*(2), \ldots, R^*(m)) = (R(-1), R(-2), \ldots, R(-m)) \)

Thus the prediction error power is determined from equation (36) as

\[ P_{\min} = \sum e_k \cdot R(-k) = P_m \]  

also equation (37) can be written as

\[ \sum e_k \cdot R(1-k) = 0 \quad i = 1, 2, \ldots, m \]  

Equation (40) can be expressed in matrix form as

\[
\begin{bmatrix}
R(0) & \ldots & R(m) \\
R(1) & \ldots & R(m-1) \\
\vdots & \ddots & \vdots \\
R(m) & \ldots & R(0)
\end{bmatrix}
\begin{bmatrix}
e_1 \\
e_2 \\
\vdots \\
e_m
\end{bmatrix}
= 
\begin{bmatrix}
P_m \\
\vdots \\
0
\end{bmatrix}
\]

\[ \text{------(41)} \]
3.2.1.1 The Levinson recursion: 12,28

Define a forward and a backward prediction error filter as the PEF with the input data in forward and reverse order respectively. The forward and backward filter outputs are then given as

\[ e^m_{f,n} = \left(\varphi^m_k X_{n-k}\right), \quad k=0 \ldots m \quad \text{-----}(42) \]

\[ e^m_{b,n} = \left(\varphi^m_k X_{n+k-m}\right), \quad k=0 \ldots m \quad \text{-----}(43) \]

where the subscript \( m \) is for the filter order. For a forward PEF of order \( m-1 \) equation (40) is given by,

\[ (\varphi^{m-1}_k R(I-k) = P_{m-1} \quad i=0 \quad \text{-----}(44a) \]

\[ = 0 \quad i=1 \ldots m-1 \]

or in matrix form

\[
\begin{bmatrix}
R(0) & R(1-m) \\
\vdots & \vdots \\
R(m-1) & R(0)
\end{bmatrix}
\begin{bmatrix}
1 \\
\varphi^{m-1}_1 \\
\vdots \\
\varphi^{m-1}_{m-1} \\
0
\end{bmatrix}
= 
\begin{bmatrix}
P_{m-1} \\
\vdots \\
0
\end{bmatrix}
\quad \text{-----}(44b)
\]

The backward PEF of order \( m-1 \) can be obtained by taking the complex conjugate of both sides of equation (40) and
recognizing that the prediction error power $P_{m-1}$ is a real quantity and that $R^*(k) = R(-k)$.

Also substituting $m-1-k$ for $k$ and $m-1-1$ for $l$, we have

$$
\sum_{k=0}^{m-1} R(-k) = P_{m-1}
$$

or in matrix form

$$
\begin{bmatrix}
R(0) & \cdots & R(1-m) \\
\vdots & \ddots & \vdots \\
R(m-1) & \cdots & R(0)
\end{bmatrix}
\begin{bmatrix}
1 \\
\vdots \\
1
\end{bmatrix}
= \begin{bmatrix}
0 \\
0 \\
P_{m-1}
\end{bmatrix}
$$

we can combine the two equations (44b) & (45b) to expand the number of PEF equations by one as follows

$$
\begin{bmatrix}
R(0) & R(-m) \\
R(1) & R(1-m) \\
\vdots & \vdots \\
R(m-1) & R(-1) \\
R(m) & R(0)
\end{bmatrix}
\begin{bmatrix}
1 \\
1 \\
0 \\
0
\end{bmatrix}
= \begin{bmatrix}
P_{m-1} \\
d^m_m \\
0 \\
0
\end{bmatrix}
$$

Then comparing (46) with (41) we obtain
The recursive procedure based on equation (47) is called the Levinson's recursion. The parameter $r_m$ is called the partial reflection coefficient as it plays similar role as the reflection coefficient in a transmission line that has both forward and reflected waves. In the recursive formula of equation (47) note that for all values of $m$, 

$$\begin{align*}
\begin{cases}
1, & k=0 \\
\hat{a}_m = r_m, & k=m \\
0, & k>m
\end{cases}
\end{align*}$$

Thus the Levinson's recursion algorithm is given by

$$\hat{a}_m^* = \hat{a}_{m-1}^* + \hat{a}_m \cdot \hat{a}^*_{m-1}$$

Eliminating $d_m$ from equations (48) & (49) we have the following recursive formula for the prediction error power

$$P_m = P_{m-1}[1-(r_m)^2]$$

which indicates that $P_m$ is a nondecreasing function of $m$ the filter order.
3.2.1.2 The Burg algorithm

For PEF to be minimum phase, Burg has shown that the reflection coefficients must all lie in the range $0_m < 1$. This is also the condition for the filter to be stable. The main problem with the methods described above is that the coefficients are not always less than one in magnitude and thus the filter stability is not guaranteed. One solution to this problem suggested by Burg is to minimize the mean prediction error power taken as the average of the forward and the backward error powers i.e.

$$P_m = 1/2(P_f, m + P_b, m)$$

where

$$P_f, m = 1/(N-m) \sum e_{f, n}^m e_{f, n}^*$$

and

$$P_b, m = 1/(N-m) \sum e_{b, n}^m e_{b, n}^*$$

The Burg's algorithm starts with $m=0$ for which

$$P_0 = R(0) = 1/N \sum X_n X_n^*$$

and then computes $P_1$ from equation (48). The value of $r_1$ for which $P_1$ is a minimum is then determined by solving $dP_1/dr_1 = 0$. Continuing this process for higher integer values of $m$, we can deduce a general expression for the reflection coefficient given by
where the forward and backward prediction-error terms are given by equations (42) & (43). Equation (53) indicates that, except for the minus sign, the reflection coefficient at stage m is equal to the normalised value of the cross-correlation of the forward prediction error $e_{m-1}^{f,n}$ and the delayed backward prediction error $e_{m-1}^{b,n-1}$, with both errors evaluated at stage $m-1$. The value of $r_m$ computed in equation (49) is always less than or equal to one in magnitude as

$$ (e_{m-1}^{f,n} - e_{m-1}^{b,n-1})(e_{m-1}^{*f,n} - e_{m-1}^{*b,n-1}) > 0 $$

Thus the use of Burg's method for the reflection coefficients always assures the stability of the resulting prediction error filter.
3.2.1.3 The Anderson Method

For a set of equal spacing (\(\Delta\)), a fast method is obtained for the MEM spectral estimation derived from Burg's algorithm. The MEM power spectrum

\[
P_{m\Delta}(f) = \frac{1}{\left[1 - \sum \theta_m \exp(-j2\pi fn\Delta)\right]^2}
\]

The frequency is limited to the Nyquist interval \(-1/2(\Delta)<f<1/2(\Delta)\). The power \(P\) and the coefficients are determined from equation (41).

Consider Burg estimation with \(m=0 \rightarrow m=1\). The two-point prediction error filter \((1,-\theta^1)\) is determined as a filter having minimum power output, where the forward power should be averaged with the power observed when the filter is operated in reverse. Minimizing this average power,

\[
b_1 = 1/2(N-1)[(X_t-\theta^1_1X_{t-1})^2+(X_{t+1}-\theta^1_1X_t)^2]
\]

as a function of \(\theta\) gives

\[
\theta^1_1 = \frac{2 \sum X_tX_{t+1}}{\sum [X_t^2+X_{t+1}^2]} \quad t=1, \ldots, N
\]

We now consider the generalization of this procedure for the step \(m-1 \rightarrow m\). The average output power of the \(m+1\) long
prediction error filter

\[ b_m = \frac{1}{2(N-m)} \sum [(X_{t+k} - \sum_{k=1}^{m} \varepsilon_k X_{t+k})^2 + (X_{t+m+k} - \sum_{k=1}^{m} \varepsilon_k X_{t+m+k})^2] \]

\[ t=1, \ldots, m \quad k=1, \ldots, m \quad \text{-----}(55) \]

is minimized with respect to the single parameter \( \varepsilon_m \). The dependence of the filter coefficients on \( \varepsilon_m \) is determined by the \( m \) lower equations in (49) having the well-known solutions

\[ \varepsilon_m = \varepsilon_{m-1} - \varepsilon_m \varepsilon_{m-1} \quad k=1, \ldots, m-1 \quad \text{-----}(56) \]

If we put \( \varepsilon_0 = -1 \) and \( \varepsilon_k = 0 \) for \( k > m \) then equation (56) holds for all \( m \). Equation (55) can now be preformulated by use of equation (56)

\[ b_m = \frac{1}{2(N-m)} \sum [(\sum_{k=1}^{m} \varepsilon_k X_{t+k})^2 + (\sum_{k=1}^{m} \varepsilon_k X_{t+m+k})^2] \]

\[ t=1, \ldots, m \quad k=1, \ldots, m \]

\[ = \frac{1}{2(N-m)} \sum [\sum_{k=1}^{m} \varepsilon_{m-1} X_{t+k} - \varepsilon_m \varepsilon_{m-1} X_{t+k}]^2 \]

\[ + [\sum_{k=1}^{m} \varepsilon_{m-1} X_{t+m+k} - \varepsilon_m \varepsilon_{m-1} X_{t+m+k}]^2) \]

\[ = \frac{1}{2(N-m)} \sum [(b_{mt} - \varepsilon_m b'mt)^2 + (b'mt - \varepsilon_m b'mt)^2] \quad \text{-----}(57) \]

here we have introduced the quantities \( b_{mt}, b'mt \) for \( t=1, \ldots, N \)

where

\[ b_{mt} = \sum_{k=1}^{m} \varepsilon_{m-1} X_{t+k} = \sum (\varepsilon_{m-1} m-k, X_{t+m-k}) \quad \text{-----}(58a) \]
\[ b'mt = \sum (\varepsilon_{m-1} \cdot x_{t+m-k}) = \sum (\varepsilon_{m-1} \cdot X_{t+m-k}) \]  

Since \( bmt \) and \( b'mt \) are independent of \( \varepsilon_m \) the condition

\[ \frac{d}{d\varepsilon_m} bmt = 0 \]

gives

\[ \frac{d^2}{d\varepsilon_m^2} bmt \geq 0 \]

Furthermore

\[ \frac{d^2}{d\varepsilon_m^2} bmt \geq 0 \]

so that the extremum of \( bmt \) for \( \varepsilon_m \) given by equation (57) is a minimum.

From equations (56) & (57) we can derive

\[ bmt = b_{m-1,t} - \varepsilon_{m-1} \cdot b_{m-1,t} \]  

\[ b'mt = b'_{m-1,t+1} - \varepsilon_{m-1} \cdot b_{m-1,t+1} \]

It is seen that the arrays \( bmt \) and \( b'mt \) are constructed from the arrays \( b_{m-1,t} \) and \( b'_{m-1,t} \) by a simple linear operation.

The starting values are \( (t=1, \ldots N) \)

\[ b0,t = b'0,t = x_t \]

Since these arrays are not used in the calculation procedure, the values for \( m=1 \) are used. For \( (t=1, \ldots N-1) \)

\[ b1,t = x_t \]
The recursion formula for $P_m$ is derived by inserting (56) into (41):

$$\begin{bmatrix} R'(0) & \ldots & R'(m) \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ -\alpha_1 \\ \vdots \\ -\alpha_m \end{bmatrix} = \begin{bmatrix} -\alpha_1 \\ -\alpha_2 \\ \vdots \\ -\alpha_m \\ -\alpha_m \end{bmatrix}$$

$$= \begin{bmatrix} R'(0) & \ldots & R'(m) \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ -\alpha_1 \\ \vdots \\ -\alpha_m \end{bmatrix} = \begin{bmatrix} -\alpha_1 \\ -\alpha_2 \\ \vdots \\ -\alpha_m \end{bmatrix}$$

$$= \begin{bmatrix} P_{m-1} \\ 0 \\ \vdots \\ 0 \\ * \end{bmatrix} \begin{bmatrix} \mathbf{1} \\ -\alpha_m \\ \vdots \\ -\alpha_m \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} P_m \\ 0 \\ \vdots \\ 0 \\ \mathbf{0} \end{bmatrix}$$

(\text{the explicit value of } \ast \text{ is not needed}) \quad P_m = P_{m-1}[1-(\alpha_m)^2]$$

It follows from equation (57) that $\alpha_m < 1$, so that

$$0 < P_m < P_{m-1}$$
3.2.2 Auto-Regressive Method

A linear random process can be described by a linear filter model of the form,

\[ X(t) = \mu + a_t + V_1 a_{t-1} + W_2 a_{t-2} \]

The sequence \( a_t \) is a white noise series with zero mean and \( \sigma^2 \) variance. The \( \mu \) is a parameter that describes the level of the process \( X(t) \). The linear filter is represented by

\[ W(z) = 1 + W_1 z + W_2 z^2 + \ldots \] where \( z \) is the unit delay operator.

In this case \( \mu \) is the mean of \( X(t) \).

Assume that \( \mu \) has been estimated and removed from the series. Process \( X(t) \) will therefore describe a zero mean process.

If we take the \( z \)-transform of (62)

\[ X(z) = W(z)A(z) \] letting \( Q(z) = W^{-1}(z) \)

\[ Q(z) X(z) = A(z) \]

In the time domain (62) can be written as

\[ X(t) = \theta_1 X(t-1) + \theta_2 X(t-2) + \ldots + a_t \]

In deriving equation (64) we have assumed that the linear process is invertible, which is satisfied by \( (z) \) converging on or within the unit circle. A stochastic process can well
be presented by a finite version of (63)

\[ X(t) = \theta_1 X(t-1) + \theta_2 X(t-2) + \ldots + \theta_p X(t-p) + \alpha_t \quad \text{-----}(65) \]

equation (65) is a discrete representation of an AR-Process of order \( p \). Since \( p \) is finite, invertibility is assumed.

For stationarity the roots of the polynomial

\[ 1 - \theta_1 z + \ldots + \theta_p z^p \]

must lie outside the unit circle.

It has been shown by Smylie et al. (1973) that an expression for the entropy of a discrete stationary Gaussian process

\[ X(t) \quad t = 1, \ldots, m+1 \]

is given by

\[ H = \frac{1}{2} \log \det(C(m)) \quad \text{-----}(66) \]

where \( C(m) \) is the semi-positive definite Toeplitz auto-covariance matrix of the process \( X_t \).

\[
C(m) = \begin{bmatrix}
R(0) & R(1) & \ldots & R(m) \\
R(1) & R(0) & \ldots & R(m-1) \\
\vdots & \vdots & \ddots & \vdots \\
R(m) & R(m-1) & \ldots & R(0)
\end{bmatrix} \quad \text{-----}(67)
\]

Let us assume that the first \( m+1 \) lags \( R(0), \ldots, R(m) \) are exactly known. The idea behind MEM is to choose the unknown auto-covariance coefficients \( R(m+1), \ldots, R(m+2) \) etc...
in such a manner that the entropy of the process is maximized at each step. It follows from (62) that the coefficient $R(m+1)$ is determined by maximizing $\text{det}(C(m))$ with respect to $R(m+1)$. The fact that $\text{det}(C(m+1))$ has a single maximum as a function of $R(m+1)$ follows the semi-positive definite property of $C(m+1)$ by applying the product rule of differentiation. Once $R(m+1)$ has been obtained $R(m+2)$ is determined by substituting the value of $R(m+1)$ just found into $C(m+2)$ and maximizing with respect to $R(m+2)$. This procedure is repeated for the other coefficients.

The solution to the maximization of $\text{det}(C(m+1))$ with respect to $R(m+1)$ is shown to be,

\[
\begin{vmatrix}
R(1) & R(0) & \ldots & R(m-1) \\
R(2) & R(1) & \ldots & R(m-2) \\
\vdots & \vdots & \ddots & \vdots \\
R(m+1) & R(m) & \ldots & R(1)
\end{vmatrix}
\]

by using equation (65) consider an AR-process of order $m$

\[X(t)=\theta_1 X(t-1)+\ldots+\theta_m X(t-m)+\epsilon_t \quad \text{-------(69)}\]

Let us multiply throughout in (69) by $X(t-k)$ and take expected values; then since $E(X(t-k)\cdot \epsilon_t) = 0$ for $k > 0$

\[R(k)=\theta_1 R(k-1)+\theta_2 R(k-2)+\ldots+\theta_m R(k-m) \quad k > 0 \quad \text{-------(70)}\]
Substituting \( k=1,2,\ldots,m+1 \) in equation (70) we obtain the set of equations that are commonly known as the Yule-Walker equations

\[
\begin{align*}
R(1) - \theta_1 R(0) - \ldots - \theta_m R(m-1) &= 0 \\
R(2) - \theta_1 R(1) - \ldots - \theta_m R(m-2) &= 0 \\
&\vdots \\
R(m+1) - \theta_1 R(m) - \ldots - \theta_m R(1) &= 0
\end{align*}
\]  -----(71)

Suppose now that the first \( m+1 \) lags \( R(0), R(1), \ldots, R(m) \) of the auto-covariance function are known. Substituting these values into the first \( m \) equations of (71) yields the coefficients \( \theta_1, \ldots, \theta_m \).

The unknown coefficient \( R(m+1) \) may now be determined from (71) by solving

\[
\begin{vmatrix} 
R(1) & R(0) & R(m-1) \\
R(2) & R(1) & R(m-2) \\
& & \vdots \\
R(m+1) & R(m) & R(1) 
\end{vmatrix} = 0 \quad -----(72)
\]

A comparison between (64) & (68) shows that exactly the same value for \( R(m+1) \) is determined by maximizing the entropy of the process. This equivalence extends to all the
extrapolated auto-covariance coefficients.

The spectrum of the AR-process for unit sampling may be determined from (70) by taking the Z-transform, thus

\[
|X(z)|^2 = \frac{\frac{1}{A(z)}}{1 - \theta_1 z^{-1} - \theta_2 z^{-2} \ldots - \theta_m z^{-m}}
\]

substituting \( z = \exp(-j2\pi f) \) in (73)

\[
S(f) = \frac{2\sigma^2}{1 - \sum_{i=1}^{m} \exp(-j2\pi fi)}
\]

In order to compute MEM using equation (74) we must determine the length of the required prediction filter \( m \) (ie the order of the AR-process), then the coefficients \( \theta_1, \theta_2 \ldots \theta_m \), since the method of determining \( m \) assumes the knowledge of the coefficients.

3.2.2.1 The Yule-Walker Estimate

A method of obtaining the coefficients is by solving the equations in (74) with estimates \( R'(k) \) substituted for autocovariance \( R(k) \). The estimates \( R'(k) \) are computed as

\[
R'(k) = \frac{1}{N} \Sigma (X(t+k) - M_n)(X(t) - M_n)
\]

where \( M_n = \frac{1}{N} \Sigma X(t) \quad t=1\ldots \)
The YW equations are now written

\[
\begin{bmatrix}
  R'(0) & R'(m-1) \\
  R'(1) & R'(m-2) \\
  \vdots & \vdots \\
  R'(m-1) & \cdots \cdots \cdots \cdots \cdots R'(0)
\end{bmatrix}
\begin{bmatrix}
  \@m,1 \\
  \@m,2 \\
  \vdots \\
  \@m,m
\end{bmatrix}
= 
\begin{bmatrix}
  R'(1) \\
  R'(2) \\
  \vdots \\
  R'(m)
\end{bmatrix}
\]

where \@m,k is the kth coefficient of the order AR-process.

A recursive solution to (76) may be obtained using Levinson recursion formula equation (43). The estimation of coefficients is illustrated as follows:

Let us obtain \@3,k k=1,2,3 from \@2,k k=1,2.

By using the first two equations of (76) with m=3, the estimates \@3,1 & \@3,2 may be expressed in terms of \@3,3 in the form,

\[
\begin{bmatrix}
  R'(0) & R'(1) \\
  R'(1) & R'(0)
\end{bmatrix}
\begin{bmatrix}
  \@3,1 \\
  \@3,2
\end{bmatrix}
= 
\begin{bmatrix}
  R'(1) \\
  R'(2)
\end{bmatrix} = \@3,3 \begin{bmatrix}
  R'(1) \\
  R'(2)
\end{bmatrix}
\]

however substituting m=2 in equation (76) gives

\[
\begin{bmatrix}
  \@2,1 \\
  \@2,2
\end{bmatrix}
= R'(1) \begin{bmatrix}
  \@2,1 \\
  \@2,2
\end{bmatrix}
\]

\[
= C'(1) \begin{bmatrix}
  R'(1) \\
  R'(2)
\end{bmatrix}
\]

57
It follows that

\[
\begin{bmatrix}
\theta_{3,1} \\
\theta_{3,2}
\end{bmatrix} = \begin{bmatrix}
\theta_{2,1} \\
\theta_{2,2}
\end{bmatrix} \theta_{3,3} - \begin{bmatrix}
\theta_{2,2} \\
\theta_{2,1}
\end{bmatrix} \quad \quad \text{(77)}
\]

In order to evaluate (77) we need an expression for \( \theta_{3,3} \).

In deriving equation (76) we used the fact

\[
E(X(t-k).at) = 0 \quad \text{for} \ k > 0 \quad \text{when} \ k = 0 \ \text{then}
\]

\[
E(X(t-k).at) = E(X(t).at) = E(at)^2 = \sigma^2
\]

\[
R'(0) = \theta_1 R'(1) + \theta_2 R'(2) + \cdots + \theta_m R'(m) + \sigma^2
\]

Augmenting (71) by (78), we can write

\[
\begin{bmatrix}
R(0) & R'(1) & R'(m) \\
R'(1) & R'(0) & R'(m-1) \\
\vdots & \vdots & \vdots \\
R'(m) & R'(m-1) & R'(0)
\end{bmatrix} \begin{bmatrix}
1 \\
-\theta_1 \\
\vdots \\
-\theta_m
\end{bmatrix} = \begin{bmatrix}
\sigma^2 \\
0 \\
\vdots \\
0
\end{bmatrix}
\quad \quad \text{(79)}
\]

The correspondence between the AR-process \(^2\) (79) and prediction of \( X(t) \) from the knowledge of its past values identifies the constant \( P_{m+1} \) as the prediction error resulting from the convolution of \( X(t) \) with the \( m+1 \) point prediction error filter \( Y_t \). We express the recursion in (77) using the formulation in (79).
Similarly the recursion of the right hand side of (74) becomes

\[
\begin{bmatrix}
P_4 \\
0 \\
0
\end{bmatrix} =
\begin{bmatrix}
P_3 \\
D_3
\end{bmatrix}
\begin{bmatrix}
D_3 \\
P_3
\end{bmatrix}
\]

----- (81)

from equations (80) & (81)

\[
D_3 = R'(3) - @2,1.R'(2) - @2,2.R'(1)
\]

\[
P_3 = R'(0) - @2,1.R'(1) - @2,2.R'(2)
\]

finally we evaluate @3,3 from (81) as

\[
@3,3 = \frac{D_3}{P_3}
\]
3.2.3 Linear prediction method

Given the time series \(X_1, X_2, ..., X_n\) let us assume that \(X_n\) can be estimated by \(X'n\) where

\[
X't = \sum_{j=1}^{m} \beta_j X(t-j) \quad \text{for } t=m+1, ..., n \quad ------(82)
\]

for given values of the parameters or filter coefficients \(\beta_j\) the non-recursive filter (82) provides a forward prediction of \(X_t\), putting \(e't = X_t - X't\) and adopting least squares (LS) criterion, the parameters \(\beta_j\) are determined by minimizing the residual sum of squares

\[
\sum (e't)^2 \quad t=m+1, ..., n
\]

Eventually an LS solution is sought for the (AR) scheme of order \(m\) defined by

\[
X_t = \sum_{j=1}^{m} \beta_j X(t-j) + e't \quad \text{for } t=m+1, m+2, ..., n \quad ------(83)
\]

the filter (83) provides a forward prediction of \(X_t\), and its parameter \(\beta_j\).

Conversely, suppose that \(X_t\) can be estimated by

\[
X'' = \sum_{j=1}^{m} \alpha_j X(t+j) + a \quad \text{for } t=1, 2, ..., n \quad ------(84)
\]

The filter in (84) provides a backward prediction of \(X_t\) and its parameter \(\alpha''\) can be determined by minimizing
\[ \sum (e'_t)^2 \quad t=1..n-m \quad \text{where} \quad e'_t = X_t - X'^t \]

the forward prediction in equation (84) is described by the 

\((n-m) \times (m)\) matrix

\[ X' \cdot \theta' = Y' \]

\[ \text{------(85)} \]

where

\[
X' = \begin{bmatrix}
X_m & X_{m-1} & \ldots & X_1 \\
X_{m+1} & X_m & \ldots & X_2 \\
\vdots & \vdots & \ddots & \vdots \\
X_{n-1} & X_{n-2} & \ldots & X_{n-m}
\end{bmatrix} \quad \theta' = \begin{bmatrix}
\theta_1 \\
\theta_2 \\
\vdots \\
\theta_m
\end{bmatrix} \quad Y' = \begin{bmatrix}
X_{m+1} \\
X_{m+2} \\
\vdots \\
X_n
\end{bmatrix} \quad \text{------(86)}
\]

If \(n > 2m\), this system is over determined and no \(\theta'\) will 
satisfy solution for equation(86).

It is useful to define residual vectors

\[ e' = Y' - X' \cdot \theta' \]

as an LS solution (\(\theta'^*\)) to equation (85) minimizing the 
residual sum of squares,

\[ S = S(\theta'^*) = e'^T \cdot e' \]

where \(T\) denotes transpose operation.

The minimum value of residual sum of squares \(S\) is achieved 
when all its first partial derivatives are zero. It follows
that $S$ is minimized when

$$X'T, e' = 0$$

and so $@'$ must satisfy the equation

$$X'T(Y' - X', @') = 0$$  \hspace{1cm} (67)

The LS solution to equation (67) is characterised by the $m \times m$ system of normal equations,

$$X'T, X', @' = X'T, Y'$$

$$R', @' = S'$$  \hspace{1cm} (68)

where

$$R' = (r_{i,j}) = X'T, X'$$

and

$$S' = (S'1, S'2, \ldots S'm) = X'T, Y'$$

Similarly the backward prediction of equation (64) described by the $(n-m) \times m$ matrix equation

$$X'', @'' = Y''$$

where

$$X'' = \begin{bmatrix} X2 & Xm+1 \\ \vdots & \vdots \\ X_{n-m+1} & X_n \end{bmatrix}, \quad @'' = \begin{bmatrix} @''_1 \\ \vdots \\ @''_m \end{bmatrix}, \quad Y'' = X1$$

$$Xn-m$$

62
The LS solution $\theta^{**}$ is characterised by the $m \times m$ system of normal equations

$$R'' \cdot \theta^{**} = S''$$

where $R'' = (r''_{i,j}) = X''^T X''$

and $S'' = (S''_1, S''_2, \ldots, S''_m) = X''^T Y''$

For both backward and forward we can assume the solution

$$(R' + R'') \theta^* = S' + S''$$

by comparing equations (88), (89) & (90) we can say that MEM vector $\theta^*$ is not related in any simple fashion to $\theta'^*$ and $\theta''^*$. Thus the normal equations (88), (89) & (90) pertain to three distinct AR-schemes of order $m$.

3.2.3.1 The Algorithm:

The general scheme is so that for a forward prediction, $R'm+1$ and $S'm+1$ can be obtained from $R'm$ and $S'm$. When a backward prediction is needed, a similar scheme is available to generate $R''m+1$ and $S''m+1$ from $R''m$ and $S''m$. 

63
3.2.3.2 Forward Algorithm

\[ R^{'}m+1,j \leftarrow S^{'}m - X_{n-m}X_n \]
for \( j=2 \ldots m+1 \) do
\[ R^{'}m+1,j \leftarrow R^{'}m,j-1 - X_{n-m}X_{n+1-J} \]
for \( i=1 \ldots m \) do
for \( j=1 \ldots i \) do
\[ R^{'}i,j \leftarrow R^{'}i,j - X_{m+1-I}X_{m+1-J} \]
for \( i=1 \ldots m \) do
\[ S^{'}i \leftarrow S^{'}i - X_{m+1-I}X_{m+1} \]
\[ S^{'}m+1 \leftarrow (X_tX_{m+1}+t) \quad t=1 \ldots n-m-1 \]

3.2.3.3 Backward Algorithm

\[ R^{"}m+1 \leftarrow S^{"} - X_1X_{m+1} \]
for \( j=2 \ldots m \) do
\[ R^{"}m+1,j \leftarrow R^{"}m,j-1 - X_jX_{m+1} \]
for \( i=1 \ldots m \) do
for \( j=1 \ldots i \) do
\[ R^{"}i,j \leftarrow R^{"}i,j - X_{n-m+1-I}X_{n-m+J} \]
for \( i=1 \ldots m \) do
\[ S^{"}i \leftarrow S^{"}i - X_{n-m}X_{n-m+i} \]
\[ S^{"}m+1 \leftarrow (X_tX_{m+1}+t) \quad t=1 \ldots n-m-1 \]
3.2.3.4 Both Forward and Backward prediction Algorithm.

\[ R_{m+1} \leftarrow S_m - X_{n-m} \cdot X_n - X_1 \cdot X_{m+1} \]
for \( J = 2 \ldots m \)
\[ R_{m+1,J} \leftarrow R_{m,J-1} - X_{n-m} \cdot X_{n+1-J} - X_J \cdot X_{m+1} \]
for \( i = 1 \ldots m \) do
for \( j = 1 \ldots i \) do
\[ R_{i,J} \leftarrow R_{i,J} - X_{m+1-i} \cdot X_{m+1-j} - X_{n-m+i} \cdot X_{n-m+j} \]
for \( i = 1 \ldots m \) do
\[ S_{i} \leftarrow S_{i} - X_{m+1-i} \cdot X_{m+1} - X_{n-m} \cdot X_{n-m+1} \]
\[ S_{m+1} \leftarrow 2 \cdot (X_t \cdot X_{m+1+t}) \quad t=1\ldots n-m-1 \]

All three algorithms produce the normal equations of order \( m+1 \) by forming the \((m+1)\)th row, then overwriting the coefficient matrix and right-hand side vector of order \( m \), and finally calculating the required inner product.
If a stochastic process consists solely of sinusoids in additive noise, then it is possible to model it as a special case ARMA process. This model assumes the sinusoids are in general non-harmonically related. The mathematical properties of this special ARMA process lead to eigenanalysis for the estimation of its parameters.

Sinusoids in additive noise is a frequently used test process for evaluating spectrum analysis techniques. To motivate the selection of an ARMA process as the appropriate model for sinusoids in white noise, consider the following trigonometric identity:

\[
\sin(n\omega) = 2\cos(\omega)\sin((n-1)\omega) - \sin((n-2)\omega) \quad (92)
\]

for \(-\pi < \omega < \pi\)

By letting \(\omega = 2\pi f \Delta t\) where \(-1/2\Delta t < f < 1/2\Delta t\)

\(\sin(n\omega)\) represents a sinusoid sampled at increments of \(\Delta t\) secs. By setting \(X(n) = \sin(n\omega)\), equation (92) may be rewritten as a second order difference equation

\[
X(n) = 2\cos(\omega)X(n-1) - X(n-2) \quad (93)
\]

Permitting the current sinusoid value to be recursively computed from the two previous values.
If the z transform of (93) is taken then

\[ X(z)[1-2\cos(n)Z^{-1}+Z^{-2}] = D(z) \]

where \( D(z) \) is a polynomial of second degree that reflects the initial conditions. It has the characteristic polynomial

\[ 1-2\cos(n)Z^{-1}+Z^{-2} \]

or equivalently

\[ Z^2-2\cos(n)Z+1 \]

with roots at

\[ Z_1 = \exp(-2\pi f_1 \Delta t) \quad \text{and} \quad Z_2 = \overline{Z_1} = \exp(2\pi f_1 \Delta t) \]

The roots are of unit modulus \( |Z_1| = |Z_2| = 1 \), and the sinusoidal frequency in Hertz is determined from the roots as follows:

\[ f_1 = \frac{\tan^{-1}[\text{Im}(Z_1)/\text{Real}(Z_1)]}{2\pi} \quad 1=1,...,p \]

Note that \( f_1 = -f_2 \). Observe that (93) is the limiting case of AR(2) process in which the driving noise variance tends to zero and the poles tend to the unit circle. Also with only two coefficients and knowledge of two samples (93) makes it possible to perfectly predict the sinusoidal process for all time. In general a 2pth order difference equation of real coefficients of the form

\[ X(n) = \sum_{m=1}^{2p} \beta_m (X_{n-m-1}) \]

can represent a deterministic process consisting of \( p \) real sinusoids of the form \( \sin(2\pi f_1 \Delta t) \). In this case the \( \beta_m \) are...
coefficients of the polynomial

\[ Z^{2p} + a_1 Z^{2p-1} \ldots a_{p-1} Z^p + a_p Z + a_{2p-1} Z + a_{2p} = \]

\[ \sum (Z-Z_1)(Z-Z_1^*) = 0 \] \hspace{1cm} \text{-----(97)}

with unit modulus roots that occur in complex conjugate pairs of the form \( Z_i = \exp(j2\pi\Delta t) \), where the \( f_i \) are arbitrary frequencies such that \( 1/-2\Delta t < f_i < 1/2\Delta t \) and \( i=1\ldots p \).

For this purely harmonic process it can be shown that

\[ a_i = a_{2p-1} \quad i=0\ldots p \]

For sinusoids in additive white noise \( w(n) \), the observed process is

\[ Y(n) = X(n) + w(n) = -\Omega_m X(n-m) + w(n) \] \hspace{1cm} \text{-----(98)}

where \( E[w(n)w(n+k)] = \delta_k \delta_k \)

\[ E[w(n)] = 0 \quad \text{and} \quad E[Z_n w(n)] = 0. \]

Since the noise is assumed to be uncorrelated with the sinusoids. Substituting \( X(n-m) = Y(n-m) - w(n-m) \) into (98) it is possible to rewrite (98) as

\[ \sum \Omega_m Y(n-m) = \sum \Omega_m w(n-m) \] \hspace{1cm} \text{-----(99)}

where \( \Omega_0 = 1 \) by definition.
Expression (99) represents the sinusoids in white noise process in terms of the noise \( w_n \) and the noisy observations \( Y_n \); it has the structure of an ARMA\((p,p)\). However this ARMA has a special symmetry in which the AR parameters are identical to the MA portion of the model.

If the autocorrelation function of \( Y(n) \) is known the ARMA parameters can be found as the solution to an eigen equation.

An equivalent matrix expression for (99) is

\[
Y^T \Sigma - \Sigma Y^T
\]  
-----(100)

where \( Y^T = [Y_n \ Y_{n-1} \ ... \ Y_{n-2}] \)

\( \Sigma^T = [1 \ \theta_1 \ \ldots \ \theta_{2p-1} \ \theta_p] \)

\( \Sigma = [w_n \ w_{n-1} \ \ldots \ w_{n-2p}] \)

Premultiplying both sides (100) by the vector \( Y \) and taking the expectation yields

\[
E[YY^T] = E[Y \Sigma Y^T]
\]  
-----(101)

Defining \( X^T = [X(n) \ \ldots X(n-2p)] \)

then

\[
E[YY^T] = \Sigma Y = \begin{bmatrix}
R_y(0) & R(-2p) \\
\vdots & \vdots \\
R_y(2p) & R(0)
\end{bmatrix}
\]  
-----(102)
\[ E[YW^T] = E[(X+W)W^T]E[W^T] = 0^T I \tag{103} \]

\[ \text{Ry is the toeplitz autocorrelation matrix for the observed process and I is the identity matrix.} \]

The fact that \( E[XW^T] = 0 \) follows from the assumption that the sinusoids are uncorrelated with the noise. Expression (101) is then rewritten as

\[ \text{Ry \lambda} = \omega^2 \lambda \tag{104} \]

which is an eigen equation where the noise variance \( \omega^2 \) is an eigenvalue of the autocorrelation matrix \( \text{Ry} \). The ARMA parameter vector \( \lambda \) is the eigen vector associated with the eigen value \( \lambda \) scaled so that the first element is unity.

Equation (104) will yield the ARMA parameters when the lags are known.

Once the ARMA coefficients are found, the roots \( Z_n \) of the polynomial in (97) formed from the coefficients will yield the sinusoid frequencies since the roots are of unit modulus with

\[ Z_n = \exp(j2\pi fn \Delta t) \tag{105} \]

The minimum eigenvalue and associated eigenvector may be found by the power method in which the sequence of vectors

\[ \lambda(k+1) = \text{Ry}^{-1} \lambda(k) \quad k=0,1.. \tag{106} \]
converges in the limit to the eigenvector of the minimum eigenvalue for some initial guess $A(0)$. Equation (106) can be rewritten as

$$R_y A(k+1) = A(k)$$

which can be solved for the unknown vector $A(k+1)$, given $A(k)$.

Once $A$ is found, the minimum eigenvalue (and therefore the noise variance estimate) is given by

$$\lambda_{\text{min}} = \frac{A^T R_y A}{A^T A}$$

Once the frequencies have been determined from the polynomial rooting of $A$ the sinusoid powers can be determined. The autocorrelation lags $R_y(1) \ldots R_y(p)$ may be expressed in matrix form.

$$FP = R_y$$

$$P = \begin{bmatrix} P_1 \\ \vdots \\ P_p \end{bmatrix}, \quad F = \begin{bmatrix} \cos(2\pi f_1) & \cos(2\pi f_p) \\ \vdots & \vdots \\ \cos(2\pi f_1p) & \cos(2\pi f_pp) \end{bmatrix}$$
The matrix $F$ is composed of terms that depend upon the sinusoid frequencies as determined from polynomial rooting. The sinusoid powers are found by solving the simultaneous equation set (109) for the power vector $P$. The noise power can also be determined from

$$\sigma_w^2 = R_y(0) - \sum_{i=1}^{P} P_i$$
CHAPTER FOUR

Real time speech records were used in this work to compare the different methods of spectral estimation techniques and to evaluate the use of the modern methods for the production of the speech spectrogram.

4.1 Speech recording

A system based on the use of the BBC micro with a frame store which was interfaced to a special A/D-D/A board was used to record the speech samples. A schematic diagram of this system is shown in Fig. (4.1). The speech was recorded at a sampling frequency of 8 kHz and was stored on the framestore. The samples were stored in two 6-bit bytes with the most significant byte MSB on the first level of the framestore while the LSB was stored on the second level. A record of up to 8-seconds of speech can be recorded, played back, retrieved or stored on a tape or a floppy disc. The digitised speech samples were fed through the BBC micro back to the main frame computer Prime 750. Short records of 2-secs duration were used to produce spectrograms by the different methods discussed in chapter (3). These spectrograms are then used for the comparison and evaluation of the different methods.
The speech records that were used in the evaluation throughout this work for a female voice recorded with very little noise in the background. The microphone was an RS-Electret condenser microphone 600 -dual impedance type L230 with sensitivity of -60dB at 1KHz (Ref. 0dB=1/ubar). The anti-aliasing filter was a PCM TRANSMIT/RECEIVE FILTER type intel-2912A. The speech was recorded at a sampling frequency of 8KHz and 12-bit Linear PCM. The sampled data was stored into two 6-bit bytes with the MSB stored on level 0 and the LSB stored on level 1 of the framestore.
FIG. 4.1
Schematic Diagram of the Speech Recording System
4.2 Computer simulations

The main frame computer Prime 750 was used throughout this work and all the programs were written in FORTRAN. Computer models of the different traditional spectrographs were simulated on the Prime 750 to produce speech spectrograms of short records of the speech sample. Computer models for the production of the spectrograms by the modern methods were also simulated and then used to produce spectrograms for the same speech records.

4.2.1 Traditional methods

Computer models of two of the most widely used traditional spectrographs were simulated on the Prime 750. They include the scanning filter technique using complex digital filter and the FFT methods.

4.2.1.1 The Complex filter method

A computer model of the scanning filter technique using complex digital filter was set up on the Prime 750. The filter used was a complex digital filter based on the Bessel 4th order low-pass filter. The pole transformation from the S-plane to the z-plane was carried out by the relation \( z = \exp(ST) \) where \( T \) is the sampling period. The filter was
chosen because it exhibits linear phase characteristics up to the cut-off frequencies which ensures minimum phase distortion.

Fig. (4.2) shows a flow chart of the program SPECT.CMLXFTR which simulates the scanning filter technique using a complex digital filter. The speech record was passed through the low pass digital filter at a centre frequency of zero and the filter output power was recorded. The filter centre frequency was then shifted along the frequency axis by small steps comparable with the filter bandwidth, and the filter output power was recorded for each step. The total output power for half a sampling period was used to form the speech spectrogram.

4.2.1.2 The FFT methods

A computer model of the FFT methods to produce speech spectrograms was set up on the Prime 750. Fig. (4.3) shows a flow chart of the program SPECT.FFT which can be used to produce speech spectrograms by the different techniques of the FFT methods. They include the Direct FFT methods, the Indirect FFT, the use of time windows such as Hamming and Blackman’s. The speech record was segmented and the power spectral estimation of each segment was computed using any one of the FFT methods. The total power estimations of the segments was used to form the spectrogram.
4.2.2 Modern Methods

Computer models to simulate the production of speech spectrograms by the modern methods of spectral estimation were formed on Prime 750. These methods include the Maximum Entropy based on Burg Algorithm, auto regressive based on solving the Yule Walker equations, Linear prediction based on forward and backward prediction, and Pisaranko harmonic decomposition.

4.2.2.1 The Maximum Entropy Method

The speech record was segmented and the power spectral estimation of the short segment was computed by the Maximum Entropy Method - based on the Burg algorithm. The order of the prediction filter was governed by the Aklake criterion where the minimum prediction error power was computed to determine the right order of the error filter. Fig. (4.4) shows a flow chart of the program SPECT.BURG which computes the spectrogram by the Burg algorithm.

Fig. (4.5) shows the flow chart of the program SPECT.ANDRSON[^3] which computes the spectrogram using Anderson efficient method for sampled data based on the Burg algorithm.
4.2.2.2 The Autoregressive Method

Fig. (4.7) shows a flow chart for the program SPECT.AR which computes the spectrogram by solving the Yule-Walker set of equations. The AR-model order was found using the Akiakke minimum residual power error as in the Burg algorithm.

4.2.2.3 The Linear Prediction

A program was set up to compute the spectrogram by the Linear prediction method. Fig. (4.8) shows the flow chart of the program SPECT.LP which computes the spectrogram by assuming an error filter and then finding a Least-square solution to (AR) scheme of order m. The order of the filter was determined by the Akiakke criterion.

4.2.2.4 Pisarenko Harmonics Decomposition

A computer simulation of computing the speech spectrum using Pisarenko harmonic decomposition was formed on the Prime750. Fig (4.9) is a flow chart of program SPECT.PISAR. The minimum eigenvalue \( \lambda_{\text{min}} \) was found using the autocorrelation matrix, then the eigen vector associated with \( \lambda_{\text{min}} \) was then used to form the polynomial. The roots of the polynomial were then found and were used to compute the power of the sinusoids determined.
READ DATA
\( X_i \ i = 1 \ldots N \)
FILTER BANDWIDTH
FILTER ORDER

CALCULATE DIGITAL FILTER
COMPLEX POLES ON Z-PLANE
FROM ANALOG FILTER S-PLANE
\( Z = e^{\text{st}} \)
\( W_c = 0 \)

\( W_c = W_c + \Delta W_c \)

INCREMENT FREQUENCY AT \( \Delta W_c \)
\( \Delta W_c = \frac{\pi}{N} \)
CALCULATE NEW POLE POSITION.
FIND FILTER NORMALISED POWER OUTPUT

FIG. 4.2
SPECT . CMPLXFLTR
START

READ DATA

\[ X_i \quad i=1\ldots N \]

IS INDIRECT FFT REQUIRED?

\[ R_x(m) = \frac{1}{N} \sum_{n=0}^{N-1} X_n X_{n+m} \]

FIND FFT FOR N POINTS

COMPUTE AUTOCORRELATION FUNCTION

STOP

FIG. 4.3
SPECT. FFT
START

READ DATA
\[ X_m \quad n = 1 \ldots N \]
\[ M \quad \text{MAXIMUM ORDER} \]

INITIALIZE
\[ \mathcal{E}_e = \sum_{k=1}^{N} |X_k|^2 \]
\[ \text{DENOM.} = 2 \mathcal{E}_e \]

\[ j = i + 1 \]

COMPUTE REFLECTION COEFFICIENT
\[ a_{ii} = \hat{k}_i = \frac{-2 \sum_{k=3}^{N-1} b_{i-1,k-1} e_{i-1,k}}{\sum_{k=1}^{N-1} (|b_{i-1,k-1}|^2 + |e_{i-4,k}|^2) \}
\]
\[ a_{ki} = a_{k-1,i} + a_{kk} a_{k-1,k-i} \]
\[ \sigma_{k}^2 = (1 - |a_{kk}|^2) \sigma_{k-1}^2 \]

\[ e_{mn} = X_n + \sum_{k=1}^{m} a_{mk} X_{n-k} \]
\[ b_{mn} = X_{n-m} + \sum_{k=1}^{m} a^{*}_{mk} X_{n-m+k} \]

NO

IS \[ m = 1 \]

YES

CALCULATE SPECTRUM
\[ Q(f) = \frac{\sigma^2 \Delta t}{|1 - \sum_{k=1}^{m} a_k \exp(-j2\pi f \Delta t)|^2} \]

STOP

FIG. 4.4
SPECT. BURG
START

READ $X_i$, $i = 1$ -- $N$

$M$ MAXIMUM

$P_0 = 1/N \sum_{i=1}^{M} X_i$

$b_1(1) = X(1)$

$b_2(1) = X(N)$

$b_1i = b_2i-1 = Xi$

$m = 1$

$a_a_m = \frac{2}{N} \sum_{i=1}^{N} (b_1i - b_2i)

\sum_{i=1}^{N} b_1i^2 + b_2i^2

p_m = p_m-1(1 - a_m^2)$

IF $m = 1$

YES

IF $m > 1$

NO

$\bar{a}_i = a_i - a_a_m a_{m-1}$

IF $\bar{a}_i < 0$

YES

CALCULATE SPECTRUM $\tilde{S}$

$S_{AR} = \frac{\sigma^2}{\Delta t} \left| 1 - \sum_{k=1}^{M} a_k \exp(-2\pi j f k \Delta t) \right|^2$

STOP

FIG. 4.5

SPECT. ANDERSON

82
COMPUTE AUTOCORRELATION FUNCTION

\[ R_{xx}(m) = \frac{1}{N} \sum_{n=0}^{N-m-1} x_{n+m} x_n \]

\[ P = 1 \]

INITIALIZE RECURSION

\[ a_u = \frac{R_{xx}(1)}{R_{xx}(0)} \]

\[ (1 - |a_{uu}|^2) R_{xx}(0) \]

\[ K = K + 1 \]

\[ a_{kk} = \frac{R_{xx}(k)}{\sum_{e=1}^{k-1} a_{k-e} R_{xx}(k-e) / \sigma^2_{k-1}} \]

\[ a_{ki} = a_{k-bi}^* a_{kk} a_{k-1,k-i} \]

\[ k = (1 - |a_{kk}|^2) \sigma^2_{k-1} \]

NO

YES

CALCULATE SPECTRUM

\[ Q_{AR} = \frac{a^2 \Delta t}{1 - \sum_{k=1}^{M} a \exp(j2\pi f k \Delta t) \sigma^2} \]

STOP

FIG. 4.6
SPECT. A.R
READ INPUT DATA

\[ X_i \quad i = 1 \ldots n \]

\[ m = 1 \]

\[ m = m + 1 \]

**COMPUTE FORWARD ERROR**

\[ \hat{e}_t^f = X_t - \sum_{j=1}^{m} j \cdot X_{t-j} \]

**COMPUTE BACKWARD ERROR**

\[ \hat{e}_t^b = X_t - \sum_{j=1}^{m} j \cdot X_{j-t} \]

\[ e^2 = e^2_t^f + e^2_t^b \]

**FIND LEAST SQUARE SOLUTION FOR MINIMUM ERROR POWER**

**COMPUTE RESIDUAL POWER**

**COMPUTE FBE**

**IS FBE MINIMUM?**

**STOP**

**FIG. 4.7**

SPECT. LP
START

READ DATA
X_i \text{ i=1------N} 
P = 1

COMPUTE 2P + 1 BIASED ESTIMATE R(m) = \frac{1}{N} \sum_{n=0}^{N-1} X_n X_{n+m}

FIND \lambda \text{ MIN.}
MINIMUM EIGEN VALUE AND ASSOCIATED EIGEN VECTOR \vec{\lambda}

\text{ IS} \lambda \text{ MIN.} 
\text{NO}

\text{YES}

ROOT POLYNOMIAL FORMED FROM \vec{\lambda}.
SOLVE FOR ROOTS
Z_i \text{ i = 1----P}

COMPUTE FREQUENCIES FROM ROOTS OF POLYNOMIAL
F_i \text{ i = 1----P}

DETERMINE THE P SINUSOIDS POWER FP = R_{xx}

STOP

FIG. 4.8
SPECT. PISARENKO
5.1 The display.

The spectrogram display is regarded as an equally important aspect of the spectrogram production as the analysis method. It provides a visible interpretation of the speech parameters which is essential in the field of speech research. The efficiency of any spectrogram production method is evaluated by the ease at which the essential information about the speech parameters can be acquired from its display.

Two different techniques of displaying the spectrogram data are investigated in this chapter. The first technique is based on the use of the graphic facilities provided by the main frame computer Prime750 situated at the Loughborough University computer centre. The second technique uses a BBC micro interfaced to a Framestore.

5.1.1 The Graphic display.

The graphic facilities available on the main frame computer Prime 750 were used to display the spectrogram. Three-dimensional data as an isometric projection with time on the x-axis, frequency on the y-axis and power amplitude on the z-axis.

Another way of displaying the spectrogram data was as a contour map which is a two dimensional diagram with time on the X-axis.
FIG. 5.1
A SCHEMATIC DIAGRAM OF
THE DISPLAY CIRCUIT
frequency on the Y-axis, and the power amplitude as equi-level contours.

5.1.2 The Framestore display

The second technique is based on the use of a circuit arrangement which interfaced the BBC microcomputer with the Framestore. The Framestore is the British-Telecom R16.4.2/RBH field store which is a monochrome digital field store with hardware selectable resolution. It provides 64-grey levels of intensity. The spectrogram data were quantized into 64-discrete levels of intensity and stored on the Framestore. A direct viewing of the spectrogram was provided on a Video-monitor which was connected to the Framestore. Hardcopies of the spectrogram can be obtained on a colour printer connected to the Framestore through the BBC-micro. Fig (5.1) shows a schematic diagram of this arrangement.

5.2 Discussion

A spectrogram of the 2-secs speech record of the sentence "MARY HAD A LITTLE LAMB" was produced by the computer model using the digital complex filter scanning method. The filter used was a bessel 4th order low-pass filter with a bandwidth of (80) Hz. The data was displayed by the different techniques discussed in this chapter. A spectrogram of the same speech record was produced on the LSI-sound spectrograph for the same filter
bandwidth of 80 Hz which was then used as the reference to evaluate these techniques.

The isometric projection display as in Fig. (5.2) presents a very poor translation of the speech spectrogram and offers very limited information about the speech waveform. The important acoustic features are hardly visible which imposes severe limitations on the use of this technique for any serious study of the speech characteristics.

The second display of Fig (5.3) is the contour mapping of the same data. The diagram presents more information than that of Fig (5.2) but some important features are still missing. A comparison between these diagrams and the diagram of Fig (5.5) of the same spectrogram data produced by the LSI-sound spectrograph shows clearly that the intensity representation is the most efficient method for displaying the speech spectrogram.

The second technique of displaying the data as intensity levels using a framestore is shown in fig (5.4). The display is very clear and all the main features are visible. A comparison between this diagram and the diagram of fig (5.5) is favourable. This method actually provides a clear display of all the important acoustic and perceptual features of the speech such as formant structure, voicing, friction, stress and pitch.
MARY HAD A LITTLE LAMB

FIG. 5.2

COMPLEX FILTER BW=80Hz D.R=40dB
MARY HAD A LITTLE LAMB

FIG 5.3 CONTOUR MAPPING
COMPLEX FILTER
BW=80Hz D. R=40dB
MARRY HAD A LITTLE LAMB

Fig 5.4 Complex Filter
BW=80Hz D.R=40dB
MARY HAD A LITTLE LAMB
4kHz

FIG 5.5 LSI-SOUND SPECTROGRAPH
BW=80Hz D.R=40dB
CHAPTER SIX

6. Discussions

This work was carried out in three stages. The first stage was to simulate the complex digital filter method using the Bessel filter. The use of the Bessel 4th-order lowpass digital filter was evaluated by comparing its results with that of the LSI-sound spectrograph for similar bandwidths and dynamic range.

The second stage was to evaluate the FFT methods by comparing their results with that of the Bessel complex filter.

The third stage was to evaluate the use of the modern methods by comparing their results with that of the FFT and the Bessel filter.

A short speech record of 2secs. duration with the sentence "MARY HAD A LITTLE LAMB" was used throughout the evaluation.

6.1 The Complex filter evaluation

A spectrogram of the speech record was produced on the LSI sound spectrograph for a filter bandwidth of 80 Hz and dynamic range of 40dB shown in fig. (6.1a). A spectrogram
of the same speech record produced by the Bessel 4th-order complex filter for the same bandwidth and dynamic range. An examination of the two spectrograms shows a close resemblance in the formant structure and the energy distribution pattern of both spectrograms. The similarities between the two spectrograms show the validity of using the Bessel 4th-order filter for producing the speech spectrograms. Spectrograms were produced by the Bessel filter method for various bandwidths which covers the practical range of the spectrograph, and are shown in figs(6.2, 6.3, 6.4). An examination of the various results that were obtained, clearly justify the choice of the Bessel 4th-order filter for the computer model simulation of the complex filter method.

The production of colour hard copies of the spectrograms improved their quality and made it easier to distinguish between different energy levels than was the case with the black and white shadings.

Colour hard copies of the spectrograms for filter bandwidths of 400 Hz and 80 Hz were produced on the LSI-sound spectrograph. Colour hard copies of the Bessel spectrograms for filter bandwidth of 80 Hz and 400 Hz were produced in two different ways. The first was produced using eight distinct colours.
Comparisons between the spectrograms are shown in figs (6.5) and (6.6). The spectrograms show noticeable improvement on the black-white spectrograms of fig. (6.1). Another colour hard copy of the spectrogram was produced using the BBC micro 8-colours display and is shown in fig(6.7). This printout was found to be satisfactory when compared with that of fig(6.3a).

The introduction of colours to the spectrogram display has improved the overall quality and made the interpretation of the information on the spectrogram display easier to explain.

The increase in the dynamic range of the Bessel filter method to 60dB was found to show more energy than that of the 40dB without severely affecting the clarity of the frequency resolution as shown in fig(6.4).

6.2 Evaluation of the FFT methods

The use of the FFT method for the spectrogram production was simulated on the mainframe computer. Various spectrograms were produced by using the different FFT methods, and then compared with the spectrograms produced by the Bessel 4th-order filter. The first spectrogram was produced by using the direct FFT method with no windowing of the data and for a dynamic range of 40dB. A comparison of this spectrogram with that produced by the Bessel filter for a
bandwidth of 160Hz and dynamic range of 40dB as in fig(6.8) shows that reasonable similarity in the formant pattern exists between the two spectrograms. The resolution of the FFT spectrogram was lower than that of the Bessel. When the dynamic range was increased to 60dB the difference in the resolution of both methods was clearly noticeable as shown in fig(6.9).

The indirect FFT method produced the same result as that of the direct method when the length of the autocorrelation function covered the whole length of the segment data as shown in fig(6.10). When the length of the auto-correlation function was reduced to a smaller value than that of the length of the data, then the resolution was found to be lower than that of the full length, and the quality of the spectrogram deteriorated as shown in fig(6.11).

The application of a time window to the data prior to the FFT produced a little improvement in the results and is shown in fig(6.12).

The overlapping of the data by introducing a 1/2 data length delay produced the best resolution of the FFT methods and is shown in fig(6.13).

A comparison between the overlapping FFT method and the Bessel complex filter for different values of filter bandwidths shows that the frequency resolution of the first method depends on the length of the data record, while the
second method can produce a wide range of frequency resolution determined by the filter bandwidth and is independent of the data record length.

When the dynamic range was increased to 60dB for the FFT methods the clarity of the displayed data was affected in each case because of the poor frequency resolutions of the FFT methods when compared with that of the complex filter method for similar constraints.

6.3 Evaluation of the modern methods

The modern methods were used to produce spectrograms of the speech record. An examination of fig(6.17) show that the MEM-Burg method produced spectrograms with such sharp frequency resolution that some of the information at the upper half of the frequency range were masked. When the dynamic range was increased from 40dB to 60dB, the Burg-spectrogram showed an improved display with the formant pattern clearly visible, and the energy distribution extending across the whole frequency range. The resolution achieved by the MEM method at 60dB dynamic range is comparable with the resolution achieved by the complex filter for a 40dB dynamic range.

The AR-method produced a spectrogram fig(6.19) with reasonable resolution at the upper half of the frequency
range but the information at the lower half of the frequency range was masked.

The increase of the dynamic range to 60dB fig(6.20) did not improve the spectrogram display.

The LP-method produced a spectrogram for a filter order (M=52) and dynamic range of 40dB which was similar to the spectrogram produced by the Burg method, fig(6.21). The spectrogram produced for (M=52) and dynamic range of 60dB shown in fig(6.22) was also similar to the results obtained by the Burg-algorithm.

The MEM-Burg algorithm and the LP-method produced reasonable results while the AR presented a very poor display especially at the lower end of the frequency scale. The resolution of the methods depends on the model order (i.e. the 'filter length') and it is independent of the data length. A smaller order (M=30) filter was used and the results compared with a Bessel filter of bandwidth 160 Hz fig(6.23). Another spectrogram was produced for M=15 and the results are comparable with those of the Bessel filter bandwidth of 400 Hz fig(6.24).

The Burg method and LP method offer reasonable resolution which depends on the order of the model M (i.e. the length
of the error filter). When the length of the filter was decreased the resolution decreased and was comparable with the wider bandwidth of the complex filter. Fig(6.23b) shows a spectrogram produced by the Burg method of order $M=30$ compared with a filter bandwidth of 160 Hz and fig(6.24a) shows a spectrogram produced by the Burg method of order $M=15$ which is comparable with the spectrogram produced by a Bessel complex filter bandwidth of 400 Hz.

The resolution of the FFT method depends on the length of the data record, while the resolution of the MEM Burg or LP method depends on the order of the model (i.e. the filter length) and is independent of the data record length as shown in fig(6.25).

In the modern methods the frequency resolution depends on the model order of the filter and it is independent of the length of the signal record. The methods do not assume that the data outside the record are zero as the case is in the FFT methods and hence, no windows are needed. Very high frequency resolutions can be obtained by some of the modern methods for very short signal record. The frequency resolution can be controlled by the value of the model order independent of the number of samples in the signal record. The realisation of these methods does not require any complex mathematical operations and hence they can be computed easily.
The Pisarenko method was first used to compute the spectral estimate of the signal used in (Marple and Kay)\textsuperscript{26}. The signal consisted of three sine waves at different fractional frequencies with some noise components. The results obtained for the spectral estimate depended mainly on the size of the Toeplitz auto-correlation matrix. The minimum Eigen value was computed to define the size of the matrix. The minimum Eigen value was decreasing gradually as the size of the matrix was increased. No actual minimum value of the Eigen value was ever reached although the computation was extended up to the maximum size of the auto-correlation matrix which is equal to the number of samples in signal.

A constraint was set to find the smallest possible change in the value of the computed Eigen value for a matrix size from that of the one previously computed for a lower matrix size.

The size of that constraint had a great effect on the number of components of frequencies that were actually calculated. When the constraint was set to about 2\% which is the difference between the two values, the main frequency components in the signal were found in their correct position within about 5\% error. There were other components on the frequency axis that were not in the noise region which had considerable power and could easily have been confused with the main components. When the constraint was set to a lower value of 1\% more components were found and the main frequency components were shifted from their actual
position. There was no straightforward method of defining the minimum Eigen value.

The method was also found to use a very high CPU time on the main frame computer which was about 200 times that of the FFT method. When the method was used for the speech analysis, the results were very disappointing and could not be translated in any sense or related in any way to the speech signal.
"MARY HAD A LITTLE LAMB"

4KHz

Fig(6.1a) LSI-SPECTROGRAPH
BW=80Hz  D.R=40dB

Fig(6.1b) COMPLEX FILTER
BW=80Hz  D.R=40dB
MARY HAD A LITTLE LAMB

**Fig (6.2a)** COMPLEX FILTER
BW = 800Hz  D.R = 40dB

**Fig (6.2b)** COMPLEX FILTER
BW = 400Hz  D.R = 40dB
MARY HAD A LITTLE LAMB

Fig (6.3a) COMPLEX FILTER
BW=40Hz D. R=40dB

Fig (6.3b) COMPLEX FILTER
BW=160Hz D. R=40dB
MARY HAD A LITTLE LAMB

4KHz

Fig (6.4a) Complex Filter
BW=80Hz D.R=40dB

4KHz

Fig (6.4b) Complex Filter
BW=80Hz D.R=60dB
MARY HAD A LITTLE LAMB

**Fig(6.5a)** LSI-SPECTROGRAPH
BW = 80Hz D.R = 40dB

**Fig(6.5b)** COMPLEX FILTER
BW = 80Hz D.R = 40dB
MARY HAD A LITTLE LAMB

Fig (6.6a) LSI-SPECTROGRAPH
BW=400Hz D.R=40dB

Fig (6.6b) COMPLEX FILTER
BW=400Hz D.R=40dB
MARY HAD A LITTLE LAMB

Fig(6.7) COMPLEX FILTER
BW=80HZ  D.R=40dB
"MARY HAD A LITTLE LAMB"

**Fig (6.8a) DIRECT FFT**

- **N = 128**
- **D. R = 40 dB**

**Fig (6.8b) COMPLEX FILTER**

- **BW = 160 Hz**
- **D. R = 40 dB**
"MARY HAD A LITTLE LAMB"

**Fig (6.9a) DIRECT FFT**

- $N = 128$
- $D. R = 60$ dB

**Fig (6.9b) COMPLEX FILTER**

- $BW = 160$ Hz
- $D. R = 60$ dB
"MARY HAD A LITTLE LAMB"

**Fig. (6.11a) DIRECT FFT**

N = 128  
D. R = 40 dB

**Fig. (6.11b) INDIRECT FFT (short)**

N = 128  
D. R = 40 dB
"MARY HAD A LITTLE LAMB"

4KHz

**Fig (6.12a) DIRECT FFT**
N=128  D. R=40dB

4KHz

**Fig (6.12b) WINDOW HAMMING**
N=128  D. R=40dB
"MARY HAD A LITTLE LAMB"

**Fig. 6.13a** DIRECT FFT  
N=128  D. R=40dB

**Fig. 6.13b** OVERLAPPING  
ND=128  D. R=40dB
"MARY HAD A LITTLE LAMB"

4kHz

Figure 6.15a: Overlapping FFT
ND=128, D.R=40dB

4kHz

Figure 6.15b: Complex Filter
BW=80Hz, D.R=40dB
"MARY HAD A LITTLE LAMB"

**Fig (6.16a) Overlapping FFT**
- ND = 128
- D. R = 60 dB

**Fig (6.16b) Complex Filter**
- BW = 160 Hz
- D. R = 60 dB
"MARY HAD A LITTLE LAMB"

Fig (6.18a) COMPLEX FILTER
BW = 80  D.R = 60 dB

Fig (6.18b) MEM - BURG
M = 52  D.R = 60 dB
"MARY HAD A LITTLE LAMB"

**Fig. 6.19a** COMPLEX FILTER
BW = 80  D.R = 40 dB

**Fig. 6.19b** AR-YULE WALKER
M = 52  D.R = 40 dB
"MARY HAD A LITTLE LAMB"

Fig. (6.20a) COMPLEX FILTER
BW=80    D.R=60dB

Fig. (6.20b) AR-YULE WALKER
M=52      D.R=60dB

120
"MARY HAD A LITTLE LAMB"

**Fig (6.22a)** COMPLEX FILTER
- BW = 80
- D. R = 60 dB

**Fig (6.22b)** LINEAR PREDICTION
- M = 52
- D. R = 60 dB
"MARY HAD A LITTLE LAMB"

Fig (6.23a) COMPLEX FILTER
BW = 160  D.R = 60 dB

Fig (6.23b) MEM - BURG
M = 30  D.R = 60 dB
"MARY HAD A LITTLE LAMB"

**Fig (6.24a) Complex Filter**
- $f = 4000\text{ Hz}$
- Duration: $2\text{ secs}$
- Filter specification: $BW = 400\text{ Hz}$, $D. R = 60\text{ dB}$

**Fig (6.24b) MEM - BURG**
- $M = 15$
- Duration: $2\text{ secs}$
- Filter specification: $D. R = 60\text{ dB}$
"MARY HAD A LITTLE LAMB"

**Fig (6.25a)** DIRECT FFT
N = 128  D. R = 60 dB

**Fig (6.25b)** MEM-BURG
M = (52)  D. R = 60 dB
CHAPTER SEVEN

7.1 Conclusions

The spectrograms produced by the complex digital filter method were for filter bandwidth range of 40-400Hz and the dynamic range values were 40dB and 60dB. The results obtained were very good with the quality and the resolution being similar to those of the spectrograms produced by the LSI spectrograph. The complex filter simulation using the Bessel 4th-order filter was therefore considered to be an acceptable method to be used as the reference for the comparison and evaluation of the other methods.

The different displays of the spectrograms suggest that the Framestore arrangement is highly acceptable for the display of the data. The use of the Framestore for the recording circuit and the display circuit with the BBC micro provides a very practical arrangement for the spectrograms to be produced in the laboratory with standard equipment, then hard copies of the spectrogram can easily obtained. It also suggest a very practical use of the system for any future work on a spectrograph implementation.

The complex digital filter method is found to be very practical and efficient. The method provides a wide range of filter bandwidth which extends between 40-500Hz. The dynamic range is variable and can be set to any desirable value. The normal values of the dynamic range is between 40-60dB. The frequency resolution of this method can be
easily controlled and depends on the filter bandwidth and not on the record length. The use of a 4th-order filter is also less costly than the special linear phase filters that are needed for minimum distortion.

The use of the FFT methods produced reasonable results for dynamic range of 40 dB but the spectrograms are very poor for the dynamic range of 60 dB. Also the frequency resolution of the methods depends mainly on the length of the data segment, which makes it impractical for producing spectrograms of short utterances.

The modern methods of Burg MEM and the LP produce spectrograms with very good frequency resolution. The frequency resolution of these methods is independent of the length of the data segment. It is possible to produce spectrograms for short utterances with reasonable frequency resolution using these methods. The AR-method was found to have poor frequency resolution at the lower scale of the frequency range.

The complex filter method was found to use the highest CPU time on the computer while the FFT uses the lowest CPU time. The Burg, LP and AR methods use a reasonable time which is more than that of the FFT but less than that of the complex
filter. Typical times for the methods are as follows:

Bessel complex filter, 28 mins
Burg, AR, LP and FFT methods 14 mins, 5 mins

Therefore the use of the Burg algorithm or the LP method would lead to the production of spectrograms with higher resolution than that of the FFT methods but as good as the spectrograms that are produced by the complex filter method. The computational efficiency of the Burg algorithm or the LP was found to be noticeable as we compare the CPU time taken by each method on the main frame computer Prime 750.

The Pisarenko method was not very successful for the speech analysis. The results obtained were not acceptable and the CPU time used was very high compared with the other methods. As an example the time used for one program run was about 650 mins of CPU time.
7.2 Suggestion for further work

The implementation of the Burg method or the Linear-prediction method for the spectrogram can be very practical for future development. The different displays of the spectrograms suggest that the Framestore arrangement is highly acceptable for the display of the data. The use of the Framestore for the recording circuit and the display circuit with the BBC micro can be very practical for any future work on spectrograph implementation.
REFERENCES

1. Ables J.G


2. Ackroyd M.H

Digital filters (Book). Butterworths 1973

3. Andersen N.


4. Arsac J.


5. Barabell A.J

Improving the resolution performance of the Eigen structure based direction finding algorithm. ICASSP 83 Boston.

6. Barrodale I. & Erickson R.E.


7. Bergland G.D

8. Bignham C., Godfrey M.D., Tuckey J.W.
Modern techniques of power spectrum estimation.
June 1967.

Digital filters with complex coefficients.
Faculte' Polytechnique de Mons Mons Belgium.
1973

10. Burg J.P.
The relation between Maximum Entropy Spectra and
Maximum Likelihood Spectr. Geophysics vol.73
April 1972.

11. Cadzow J.A.
Spectral Estimation. An overdetermined rational
model equi approach. Proc. of the IEEE Vol.70
No9 Sept. 1982.

12. Chen C.H.
Non-linear Maximum Entropy spectral analysis
method for signal recognition. Research studies
Press (RSP 1982). (Book).

13. Chen W.Y & Slegen G.R.
Experiments with ME power spectra of sinusoids.
July 1974.
14. Crystal T.H.


15. Del Ro E. & Maffucci D.F.

A minimum phase FIR complex filter design for Transmultiplexer implementation. University De Firenze Italy. 1982

16. Fant G.


17. Fettwies A.


18. Flanagan J.L.


19. Flanagan J.L.

Parametric coding of speech spectra.

20. Fry D.B.

Acoustic phonetics Cambridge Univ. Press.
Cambridge 1976 (Book).

21. Griffiths J.W.R

Short notes on complex filters.
Loughborough University. 1984

22. Harris F.

On the use of windows for harmonic analysis with
the Discrete Fourier Transform. Proc. of IEEE

23. Hellen P.A.T

Military secure speech using HF. Electronics &
Power. The Journal of the Institution of
Electrical Engn. March 85.


A technical report on the BT Framestore
R16/4/2/2BH. ]BTRL Martlesham U.K

25. Holmes J.N

The JSRU channel vocoder. IEEE Proc. Vol.127

26. Marple JR S.L. & Kay S.M.

Spectrum analysis- A modern perspective.
27. Marple L.

28. McCallig M. T.

29. Meerkotten K.

30. Nishikawa K.
Complex coefficient digital all pass networks and the applications to variable delay equalizer design. ISCAS 1980 IEEE International Symposium on Circuit and Systems vol 2

31. Nuttall A. H.
Spectral analysis of a univariant process with Bad points via Maximum entropy and linear prediction technique. NUSL report no 5303 March 1976. Naval Under-water System Center.

32. Oppenheim A. V. & Willsky A. S.
33. Papoulis A.

Maximum Entropy and Spectral Estimation A Review.

34. Pisarenko V.F.

On the estimation of spectra by means of non-linear function of the covariance Matrix.

35. Pisarenko V.F.


36. Rabiner.L.R. & Gold B.


37. Robinson E.A.

A Historical prospective of spectrum estimation.
Proc. of the IEEE Vol.70 no.9 Sept. 82.

38. Schoeveld.C.V. & Prinsen.P.J.A.

39. Tufts D.W. & Kumaresan R.

40. Tufts D.W. & Kumaresan R.
    Estimation of multiple sinusoids making Linear Prediction perform like Maximum Likelihood.

41. Van Dvn Bos.
    July 1971.

42. Ulrych T.J. & T.N Bishop

43. Welch.P.
APPENDIX A

BT-FRAMESTORE

1. Introduction

The R16.4.2/EBH field store board is a monochrome digital field store with hardware selectable resolution. Included on the board is the digital field store, a micro processor interface and all the input/output analogue television circuitry including six bit A/D and D/A converters. The use of 64k 1 bit dynamic rams means that a high resolution video store could be housed on single board with modest power consumption. The highest resolution is 512 x 512 pixels, however if a lower resolution is selected then the store is partitioned to give multiple field stores on one board. If a resolution of 512 x 256 is required then two stores are available. Six planes of storage exist for all resolutions and this gives up to 64 grey levels. Circuitry to enable a real time field snatch is included on the board and has been designed to work with both interlace and random interlace video sources. A gated oscillator circuit is used to synchronise incoming video to the store, rather than a phase lock loop. The advantage of this technique is that the store is only locked...
to the incoming video for the duration of the field being
snatched. This enables the store to snatch a field from
several different video sources whose sync timings are not
locked together and may have either interlaced or non
interlaced syncs.
The store board interfaces directly to any of the motorola
6800 series of microprocessors and occupies a 1k block of
memory in the microprocessor's memory map. Individual
pixels are accessed by storing their X, Y co-ordinates in
address latches and instructing the store a read or write
operation. Interface to processors other than the 6800
family is possible with external circuitry.
1.2 Digital store
The digital store consists of 24, 64k x 1 bit dynamic RAMs.
The 24 ICs are divided into six planes with each plane
containing four ICs. A plane is in effect a one bit field
store. The six planes in parallel give 64 grey scale
levels. The access cycle time of the memory ICs is too slow
to store the sampled video from the A/D converter directly,
therefore the inputs to the four ICs in each plane are
multiplexed. This reduces the access cycle requirement to a
quarter of the video sampling rate. Multiplexing of the
outputs is also necessary to reconstruct the video. Shift
registers are used to multiplex the input/output of the
store and to interface the memory with A/D and D/A
converters.

Timing signals i.e. (RAS, CAS, etc) are derived from a master oscillator circuit using combinational logic. The oscillator runs at approximately 20 MHz and is divided by a 74163 synchronous counter. Selection switches on the output of this counter determine the timing waveforms for either 512 or 256 horizontal resolutions. The 512 resolution gives a sampling frequency of 10 MHz for the A/D converters and it follows that 256 resolutions is 5MHz sampling. The gated master oscillator can be adjusted to give other sampling rates. There is a facility on the board to plug in a crystal oscillator should the store be used for graphics only, i.e. no field snatch required.