



CANADARARY ALARMAN

MICROCOPY RESOLUTION TEST CHART NATIONAL BUREAU OF STANDARDS - 1963 - A





RADC-TR-85-45 Final Technical Report July 1985



AN ADAPTIVE APPROACH TO A 2.4 kb/s LPC SPEECH CODING SYSTEM

Oklahoma State University

Rao Yarlagadda David L. Soldan Robert D. Preuss Leslie D. Hoy Graig S. Sims Bryan L. Burns

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

DTIC FILE COPY



85 10 11 026

ROME AIR DEVELOPMENT CENTER Air Force Systems Command Griffiss Air Force Base, NY 13441-5700

This report has been reviewed by the RADC Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RADC-TR-85-45 has been reviewed and is approved for publication.

APPROVED:

linten d'agola

ANTON S. SEGOTA Project Engineer

APPROVED:

ALLAN C. SCHELL Chief, Electromagnetic Sciences Division

FOR THE COMMANDER:

John Q.

JOHN A. RITZ Acting Chief, Plans Office

If your address has changed or if you wish to be removed from the RADC mailing list, or if the addressee is no longer employed by your organization, please notify RADC (EEV) Hanscom AFB MA 01731. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

UNCLASSIFIED SECURITY CLASSIFICATION OF THIS PAGE

AD-A160312

				MENTATION I			
18 REPORT SECURITY CLASSIFICATION			16 RESTRICTIVE MARKINGS				
UNCLASSIFIED			N/A				
	CLASSIFICATION	AUTHORITY		3 DISTRIBUTION			
	CATION DOW	NGRADING SC		Approved fo	or public re	erease; di	scribution
N/A				unlimited.			
	G ORGANIZATI	ON REPORT N	UMBER(S)	5. MONITORING	ORGANIZATION R	REPORT NUMBE	R(S)
N/A				RADC-TR-85-	-45		
Sa. NAME OF	PERFORMING C	RGANIZATION		7a NAME OF MO	INITORING ORGA	NIZATION	
			(If applicable)	Rome Air De	evelopment (Center (EE	v)
	City, State, and		nputer Engineering	76 ADDRESS (City		(Code)	
	r OK 7407		mborer pustneering	Hanscom AFE	3 MA 01731		
OCTIIMULE		~					
A NAME OF	FUNDING / SPOI	NSORING	85 OFFICE SYMBOL	9 PROCUREMENT		DENTIFICATION	
ORGANIZA			(If applicable)	F19628-81-K			
Rome Air	Developmen	nt Center	EEV	117020-01-0	(=00 - 3		
BC. ADDRESS (City, State, and	ZIP Code)		10 SOURCE OF F	UNDING NUMBE	नऽ	
		• •		PROGRAM	PROJECT	*45K	NORK NIT
Hanscom A	AFB MA 017	31		ELEMENT NO	NO	NO	ACCESSION N
	ude Security C			33401F	_7820	03	21
	AUTHORIES						
13a TYPE OF	REPORT	13b T	ldan, Robert D. Pre ME COVERED	14 DATE OF REPO	RT (Year, Month,		GE COUNT
Rao Yarla 13a TYPE OF Final	REPORT	13b T FROM	ME COVERED	14 DATE OF REPO			
Rao Yarla 13a TYPE OF Final 16. SUPPLEME	REPORT	13b T FROM	ME COVERED ATO T8_SUBJECT TERMS	14 DATE OF REPO	RT (Year, Month, 1985	, Day) 15 FAI	GE COUNT 258
Rao Yarla 13a TYPE OF Final 16. SUPPLEME N/A 17	Agadda, Da REPORT INTARY NOTAT	13b TI FROM	IME COVERED	14 DATE OF REPO July (Continue on reverse 15 Sign 10n Adap	RT (Year, Month, 1985 e if necessary an hal Estimat ptive Digita	, Day) 15 FAI nd identify by t ion al Filter	GE COUNT 258
Rao Yarla 13a TYPE OF Final 16. SUPPLEME N/A 17 FIELD 12 09 19 ABSTRACT	REPORT COSATI GROUP 02 04 1 (Continue on	13b T FROM ION CODES SUB-GROU	INE COVERED I IS SUBJECT TERMS IP Speech Analys Pitch Estimat Noise Suppres essary and identify by block	(Continue on reverse is Sign ion Adap sion Spee number)	RT (Year, Month, 1985 e if necessary an hal Estimat ptive Digita ech Enhance	Day) 15 FAI	GE COUNT 258 block number) , (See Reverse
Rao Yarla 13a TYPE OF Final 16 SUPPLEME N/A 17 FIELD 12 09 19 ABSTRACT The goal Suppressi (b) Auton Various p investiga	REPORT COSATI GROUP 02 04 Continue on of this r ion and pe regressive prefilteri	13b T FROM TON CODES SUB-GROU reverse if mec esearch w rformance spectral ng techni ter coeff	ME COVERED A TO 18 SUBJECT TERMS P Speech Analys Pitch Estimat Noise Suppres essary and identify by block as to investigate (enhancement of Name estimation in noise ques for improving icients were varied	(Continue on reverse is Sign ion Adap sion Spee number) (a) Adaptive of rowband Codin sy signals for linear predic	RT (Year, Month, 1985 and Estimat the Digit ech Enhances estimation ing Systems t speech and ctive codin	Day) 15 FAI	GE COUNT 258 Wook number) (See Reverse r noise signals and lications. were
Rao Yarla Rao Yarla 13 TYPE OF Final 16 SUPPLEME N/A 17 FIELD 12 09 19 ABSTRACT The goal suppress (b) Auton Various p investiga noise from developed dimension solution	REPORT COSATI COSATI GROUP 02 04 Continue on of this r ion and pe regressive prefilteri ated. Fil om the spe efilter co d and eval nal approa of the li	13b T FROM TON CODES SUB-GROU reverse if mec esearch w rformance spectral ng techni ter coeff ech signa nsisting uated. T ch to sup ch was fo near esti	ME COVERED A TO 18 SUBJECT TERMS P Speech Analys Pitch Estimat Noise Suppres essary and identify by block as to investigate (enhancement of Name estimation in noise ques for improving icients were varied	(Continue on reverse is Sign ion Adaptive of rowband Codin by signals for linear predict it to optimize tal predictor iter approach short time for a new recurst reloped. An in	RT (Year, Month, 1985 e if necessary an hal Estimation prive Digits ech Enhances estimation in a Systems r speech and ctive coding each filte r (ADP) with was expan ourier trans al. Fast a ive linear of	Day) 15 FAG	GE COUNT 258 Nock number) (See Reverse r noise signals and lications. were to remove riod delay w to use a two in. The two for efficien suitable for
Rao Yarla 13a TYPE OF Final 16. SUPPLEME N/A 17 FIELD 12 09 19 ABSTRACT The goal suppress (b) Autor Various p investiga noise fro developed dimension dimension solution rapid es 20 DISTRIBU CMUNCLA	COSATI COSATI COSATI GROUP 02 04 (Continue on of this r ion and pe regressive prefilteriated. Fil om the spe efilter co d and eval nal approa of the li timation o OTION AVAILAB SSIFTED/UNLIMIT	13b T FROM ION CODES SUB-GROU SUB-GROU reverse if nec esearch w rformance spectral ng techni ter coeff ech signa nsisting uated. T ch to sup ch was fo near esti f a signa	ME COVERED ATO TOTO P Speech Analyse Pitch Estimat Noise Suppres estary and identify by block as to investigate (enhancement of Name estimation in noise ques for improving icients were varied 1. of an adaptive digit he one-dimension fit press noise in the und to have signifit mation problem and 1 in noise were deviation	(Continue on reverse is Sign ion Adaptive of rowband Codin sy signals for linear predict it to optimize tal predictor iter approach short time for a new recurst reloped. An i 21 ABSTRACT SE SUNCLASS	RT (Year, Month, 1985 e if necessary an hal Estimation prive Digits ech Enhances estimation in a Systems r speech and ctive coding each filte r (ADP) with was expan ourier trans al. Fast a ive linear of improved me ECURITY CLASSIFI	Day) 15 FAG	GE COUNT 258 Dock number) (See Reverse r noise signals and lications. were le to remove riod delay w to use a two in. The two for efficien suitable for ectral
Rao Yarla Rao Yarla 13a TYPE OF Final 16. SUPPLEME N/A 17 FIELD 12 09 19 ABSTRACT The goal suppressi (b) Autor Various p investiga noise fro A new pro developed dimension dimension solution rapid es 20 DISTRIBU CMUNCLA	REPORT REPORT NTARY NOTAT COSATI GROUP 02 04 (Continue on of this r ion and pe regressive prefilteri ated. Fil om the spe efilter co d and eval nal approa of the li timation o UTION AVAILAB SSIFIED/UNLIMIT DE RESPONSIBUI	13b T FROM ION CODES SUB-GROU SUB-GROU reverse if nec esearch w rformance spectral ng techni ter coeff ech signa nsisting uated. T ch to sup ch was fo near esti f a signa	ME COVERED A	(Continue on reverse is Sign ion Adaptive of rowband Codin sy signals for linear predict it to optimize tal predictor iter approach short time for a new recurst reloped. An i 21 ABSTRACT SE SUNCLASS	RT (Year, Month, 1985 e if necessary an hal Estimat: ptive Digits ech Enhances estimation in systems r speech and ctive codin each filte r (ADP) with h was expan ourier trans al. Fast a ive linear improved me ECURITY (LASSIFI IFIED (Include Area Cod	Day) 15 FAG	GE COUNT 258 Dock number) (See Reverse r noise signals and lications. were le to remove riod delay w to use a two in. The two for efficien suitable for ectral
Rao Yarla TYPE OF Final The SUPPLEME N/A T FIELD 12 09 19 ABSTRACT The goal suppressive (b) Auton Various print investiganoise from A new produced dimension dimension solution rapid est 20 DISTRIBU CALLAN 22a NAME C Anton S.	REPORT REPORT NTARY NOTAT COSATI GROUP 02 04 (Continue on of this r ion and pe regressive prefilteri ated. Fil om the spe efilter co d and eval nal approa of the li timation o UTION AVAILAB SSIFIED/UNLIMIT DE RESPONSIBUI	13b T FROM ION CODES SUB-GROU SUB-GROU reverse if nec esearch w rformance spectral ng techni ter coeff ech signa nsisting uated. T ch to sup ch was fo near esti f a signa	ME COVERED A	(Continue on reverse is Sign ion Adaptive of rowband Codin sy signals for linear predict it to optimize tal predictor iter approach short time for cant potentia a new recurst reloped. An i 21 ABSTRACT SE SUNCLASS 225 TELEPHONE (617) 80	RT (Year, Month, 1985 e if necessary an hal Estimation ptive Digits ech Enhances estimation in a Systems r speech an ctive codin each filte r (ADP) with h was expand ourier trans al. Fast a ive linear of improved me ECUPITY CLASSIFI IFIED (include Area Coo 61-4248	Day) 15 FAN d identify by t ion al Filter methods fo for speech alysis app g systems r techniqu h pitch pe nded upon sform doma lgorithms estimator thod of sp (CATION fe) 22c OFFICE RAD	GE COUNT 258 Work number) (See Reverse signals and lications. were to remove riod delay w to use a two in. The two for efficien suitable for ectral

·-- /

والمراجع والمتعاد والمتعاد والمعاد وال

UNCLASSIFIED

estimation is described. The method treats the problem of estimating autoregressive (AR) process parameters from sequential discrete time observation corrupted by additive independent noise with known power spectral density. The method is very general and applies both to wideband and narrowband noise environments. The algorithms have been tested using simulated gaussian and speech signals with known spectral characteristics corrupted by simulated gaussian noise of known power spectral density. The results obtained by this method are compared with the results obtained by currently popular methods of AR spectral estimation in noise.

17. COSATI CODES (Continued)

Field	Group
17	02

18. SUBJECT TERMS (Continued)

Spectral Estimation	Acoustic Noise
Signal Processing	Linear Predictive
Narrowband Coding	Coding

Accession For	
NTIS GRA&I	
Py Distribution/	
Avaliable it Codes	
Averation Color Dist Special	DEIC COPY

UNCLASSIFIED

EXECUTIVE SUMMARY

This report consists of three self contained parts. Part A, entitled "Noise Suppression and Signal Estimation," presents a summary of various adaptive noise-cancelling prefilter techniques that were investigated. Part B, entitled "Autoregressive Spectral Estimation in Noise for Speech Analysis Applications," presents a fundamentally new "weighted information" approach to the spectral estimation problem. Finally, Part C presents a summary of the publication activities associated with this research effort.

Each part of this report is self-contained with separate contents, list of figures, references, etc. To assist the reader, a brief summary of parts A and B along with a general table of contents is provided below.

A. NOISE SUPPRESSION AND SIGNAL ESTIMATION

The goal of this research effort was to investigate adaptive estimation methods for noise suppression and performance enhancement of narrowband coding systems for speech signals. In the original proposal special reference was made to combining pitch tracking adaptive filters with linear predictive coding algorithms and these methods are discussed in chapter II of part A. The results of this initial

effort led to the investigation of several topics which are briefly summarized below.

In chapter 11 of part A various prefiltering techniques for improving linear predictive coding systems are evaluated. This includes the examination of a prefilter consisting of an adaptive digital predictor (ADP) with pitch Two adaptive algorithms, the least mean period delay. squared and the sequential regression, are evaluated for The method proved successful in suppressing white ADP. noise in voiced speech sounds but did not work well when the noise was narrowband such as a single sine wave. This is due to the fact that interaction between the pitch period and the narrowband noise produces a bias error in the adaptation of the filter.

Chapter III of part A discusses a new robust pitch estimation procedure which was developed as an outgrowth of the efforts described in chapter II. The performance of the pitch tracking filter depends on the quality of the pitch period estimate used to set the input delay. It was found that a tapped delay line adaptive digital filter provides a robust pitch estimate. This method allows for better resolution of the pitch frequency than the traditional techniques such as autocorrelation and harmonic analysis and has a better noise tolerance than these technqiues.

Failure of the pitch tracking adaptive filters to suppress narrowband noise prompted the investigation of several other prefiltering methods. The most successful of the

filters evaluated was the spectral subtraction technique discussed in chapter IV of part A. Two modifications to the original method proved very useful for improving noisy speech. First a dual time constant noise spectrum estimate improved white noise suppression and secondly a spectral notch feature greatly improved narrowband noise quieting. Also, very successful was a new filter method based on adaptive filtering. These filters have been investigated using speech in the military aircraft environment, such as helicopter, AWACS, etc. Informal listening tests indicate good noise suppression.

Chapter V of part A discusses two-dimensional filter approaches. The use of a two-dimensional filter approach to suppress noise in the short-time Fourier transform domain found to have a significant potential. From the informal listening tests, the two dimensional processed speech sounds quieter with added clarity.

Chapters VI and VII of Part A are concerned with fast algorithms for efficient solution of the linear estimation problem and a new recursive linear estimator suitable for rapid estimation of a signal in noise. Efficient methods are developed for optimization of the filter coefficients. Optimal selection of data to be processed is shown to be related to a classic integer programming problem.

iii

B. AUTOREGRESSIVE SPECTRAL ESTIMATION IN NOISE IN THE CONTEXT OF SPEECH ANALYSIS

In this part of the report, an improved method of spectral estimation is described. The method treats the problem of estimating autoregressive (AR) process parameters from sequential discrete time observation corrupted by additive independent noise with known power spectral density. The method has a theoretical foundation relating it to principles of information theory as well as the linear predictive (LP) procedures popularly employed for speech analysis. The estimation procedure enjoys the advantages of noise suppression filtering methods such as those discussed in Part A. Furthermore the method is able to relax the usual LPC error criterion in those spectral regions where the residual speech distortion due to noise suppression is expected to be The method is very general and applies both to widehigh. band and narrowband noise environments. The solution obtained by using this method is shown to be unique.

Computational procedures appropriate for speech analysis applications are developed. The complexity of these algorithms for speech applications is only moderately expensive when compared to the present used methods. The algorithms have been tested using simulated Gaussian and speech signals with known spectral characteristics corrupted by simulated Gaussian noise of known power spectral density. The results obtained by this method are compared with the results obtained by currently popular methods of AR spectral

1v

estimation in noise. By using scatter plots, it is shown that the estimation error is significantly lower when compared to the results obtained by other methods.

TABLE OF CONTENTS

والمعالم والمعالية

Page

Part	A: NOISE	SUPPRE	SSION	AND	SIGNAL	ESTIM	ATION	•		1
	Table of List of F			•••	•••	•••	•••	•••	•	2 4
		troduct: efilter:						••	•	б
	Pr III. Pi	edictor tch Esti	 imatic	n by	 Adapt	ive Fj	 lterii	 ng.	•	10 20
	Ap V. En	plicatio hancemen	ons . nt of	Speed	ch Sig	nals b	y Two	• •	•	31
	VI. Di VII. De	mensiona screte ' sign of	lime H Fast	Stima Recu	ation rsive	 Estima	tors	•••	•	43 52 62
	VIII. Co Cited Ref	nclusion erences								89 95
Part		EGRESSI FOR SPI							•	99
Part		FOR SPE Contents	EECH A	NALY	SIS AP		IONS		•	99 100 101
Part	NOISE Table of List of F Chapter I. In	FOR SPI Contents igures troduct:	EECH A 3 ion .	NALY:	SIS AP	PLICAT	IONS	•••	•	100
Part	NOISE Table of List of F Chapter I. In II. Ge III. Th	FOR SPI Contents igures troduct: neral D: eoretics	EECH A ion . iscuss al For	NALY: 	SIS AP	PLICAT	IONS .	· · ·		100 101 104 110 1 <i>3</i> 4
Part	NOISE Table of List of F Chapter I. In II. Ge III. Th IV. Co V. Re VI. Co	FOR SPE Contents igures troduct: neral D: eoretics mputatics sults nclusion	EECH A ion . iscuss al For onal F	ion mula ormula	SIS AP	PLICAT	IONS .	· · ·	• • • •	100 101 104 110 134 159 197 228
	NOISE Table of List of F Chapter I. In II. Ge III. Th IV. Co V. Re VI. Co Cited Ref	FOR SPE Contents igures troduct: neral D: eoretics mputatics sults nclusion	EECH A	ion mula ormula	SIS AP	PLICAT	IONS .	· · · · · · · · · · · · · · ·	• • • • •	100 101 104 110 134 159 197

PART A

7

NOISE SUPPRESSION AND SIGNAL ESTIMATION

TABLE OF CONTENTS

Page

Chapter

•••

I.		6
	Noise Suppression	6 9
11.	PREFILTERING WITH AN ADAPTIVE DIGITAL PREDICTOR	10
	Filter Configurations	11 12 13 17 17
III.	PITCH ESTIMATION BY ADAPTIVE FILTERING	20
IV.	The General Time Delay Case	21 22 24 25 27 27 27 28
14.	APPLICATIONS	31
	Methods Investigated	31 32 33 34 35 37 39 40 41

Page

۷.	ENHANCEMENT OF SPEECH SIGNALS BY TWO DIMENSIONAL SIGNAL PROCESSING	43
	Two Dimensional Representation	45 47 51
VI.	DISCRETE TIME ESTIMATION	52
	Set Up of Recursion Equations \dots Structure of B_p B_p^T \dots Special Structure of Equation (6.11) \dots Intermediate Steps \dots Special Solutions \dots Structure of Equation (6.11) \dots Structure Steps \dots S	54 55 56 58 61
VII.	DESIGN OF FAST RECURSIVE ESTIMATORS	62
	Problem Statement	64 65 74 79 84 87
VIII.	CONCLUSIONS	89
	Suggestions for Further Study	91
CITED	REFERENCES	93

LIST OF FIGURES

Figure		Page
1.	ADP Configuration	12
2.	Test Signal-time and frequency plot	13
3.	Noise Signal Spectra	14
4.	Pitch Error Sensitivities, LMS algorithm, 1.3 dB SNR input, four and eight weight filters	16
5.	Pitch Error Sensitivity, LMS algorithm, 1.3 dB SNR, eight weight filter	16
6.	ADP Performance, LMS algorithm, eight weight filter	18
7.	ADP Performance, SER algorithm, eight weight filter	18
8.	General Configuration for a time delay TDLADF	23
9.	TDLADF for pitch period estimation	23
10.	TDLADF weight plot for the sentence, "they shook hands for good luck"	26
11.	Expanded weight plot for the start of the word "they"	26
12.	Expanded weight plot using SER adaptive algorithm	26
13.	Pitch frequency error for a TDLADF on the synthetic vowel /a/ corrupted by White Noise	29
14.	Two stage TDLADF for pitch estimation	29
15.	Adaptive predictor for speech filtering	37
16.	Adaptive Transfer Filter	39

17.	SNR performance	41
18.	Log area ratio performance	41
19.	Block diagram for a two-dimensional speech filter	46
20.	Clean speech spectrogram and two-dimensional transform	48
21.	Noisy speech spectrogram and two-dimensional transform	48
22.	Results at each step of a two-dimensional filter process	50

Page

F

ŀ

CHAPTER 1

INTRODUCTION

Digital encoding of speech signals has proven to be an effective means of bandwidth compression. Expanding use of narrowband digital encoding systems has revealed some disappointing limitations in their performance. One limitation is degredation of intelligibility in the presence of background noise [1,2]. This difficulty has drawn more attention as speech encoding systems have attempted to move into military applications. These applications may include the environments such as airborne command posts, cockpits of jet aircraft, helicopters and many others. The goal of this study was to investigate several methods of improving the performance of narrowband coding systems in the presence of acoustically coupled background noise. To complement this, an additional goal was to investigate fast algorithms to implement these algorithms along with descrete-time estimation algorithms.

Noise Suppression

The problem of noise in speech encoders can be addressed as a simple noise filter problem by placing a simple filter in front of the encoder. The problem with

this approach is that most background noise of interest is not stationary and cannot be predicted ahead of time. Noise filters with adaptive structures have shown some promise as prefilters for speech encoding systems [3]. Several of the adaptive filter techniques were investigated in this study. Chapter II describes an adaptive filter which takes advantage of pitch information to reduce background noise. Two adaptive algorithms, the Least Mean Square Algorithm (LMS) [4] and the Sequential Regression Algorithm (SER) [5]. the adaptive evaluated for digital predictor. were Performance was evaluated using signal to noise ratio measurements.

It has been recognized that pitch estimation is not simple if the speech is corrupted by noise [6]. The currently popular pitch extraction methods are generally of two types, autocorrelation analysis and harmonic analysis. The autocorrelation method performs an autocorrelation on the windowed speech data and the pitch period for the voiced speech j S estimated by using the peaks j n the On the other hand, the pitch autocorrelation function. extraction methods are based on performing a discrete Fourier transform on the windowed speech data [7]. However, these methods do not fare well for noisy speech. Chapter III presents a method for determining the pitch period of a voiced speech signal using a tapped delay line adaptive digital filter (TDLADF). These filters have been widely used in sonar applications [8-9]. This approach uses the

TDLADF to estimate the time delay between two parts of the same signal. The time delay then corresponds to the pitch of the signal. This time delay is estimated by processing the weights of the TDLADF.

Several prefiltering methods based upon adaptive techniques are discussed in Chapter IV. One popular method of prefiltering speech involves direct modification of the short-time Fourier transform (STFT) of a speech signal typically called spectral subtraction [10]. This method estimates the STFT of the background noise and then subtracts this from the STFT of the speech plus background noise. The work described in Chapter IV is limited to filtering techniques for suppressing narrowband background noise in speech signals. The methods included are a modified spectral subtraction technique, an inverse transform filter, an adaptive notch placement technique and an adaptive filter These methods are evaluated using signal to technique. nsise ratio measurements and log area ratio performance measurements.

Chapter V presents some interesting image processing techniques to enhance speech. In this approach, the STFT representation of a segment of speech in treated as a twodemensional image data [11]. It has been shown that by the use of image processing operations, such as contrast enhancement and smoothing, background noise can be suppressed and the speech signal can be enhanced.

Fast Recursive Estimators

Real time implementation of the before mentioned filter techniques requires considerable computational power. It is important that all the alogorithms involving in the encoding process be as efficient as possible. During the course of this study, fast algorithms have been developed for linear estimators for rapid estimation of signal in noise. Chapter VI presents a simple and an efficient algorithm for the solution of a generalized least squares estimation problem. Chapter VII presents a recursive linear estimator for rapid estimation of a signal in noise. Efficient methods are developed for optimization of the filter coefficients. Optimal selection of data to be precessed is shown to be related to a classic integer programming problem [12-13].

Finally, Chapter VIII summarizes the results and provides suggestions for further study.

CHAPTER II

PREFILTERING WITH AN ADAPTIVE DIGITAL PREDICTOR

Although linear preditive coding (LPC) techniques used to Encode narrowband speech signals are efficient at data compression [6,14], they degrade significantly if the speech is corrupted with noise [1,2,15]. Frequency domain filtering techniques using pitch period information have been evaluated and found to be of limited usefulness [16]. The objective of this Chapter is to examine the performance of a time domain filtering technique using an estimate of the pitch period and an adaptive digital predictor (ADP), Figure 1, to reduce the noise to speech signals [3].

The noise filtering properties of the ADP with pitch period delay are examined. The primary areas discussed here are the ADP's performance with various noise types. Performance will be evaluated using signal-to-noise ratio (SNR) measurements. This performance measure is used for convenience in the parameter sensitivity investigations although it is recognized that intelligibility is not always a function of objective measures of speech quality.

Two adaptive algorithms, the LMS [4] and the SER [5], were evaluated for the ADP. These algorithms are discussed in the next section. Synthetic speech was used in all evaluations. The section on Data Generation discusses how this speech was generated. The effect of pitch estimate errors on performance is dealt with in the Pitch Estimate Error section. The ADP's performance for various types of noise is the subject of the Narrowband Noise section.

Filter Configurations

Two adaptive algorithms will be considered, the LMS algorithm [4] and the SER algorithm [5]. The LMS algorithm is a suboptimal least squares approach derived using the method of steepest descent. The SER algorithm is optimum in the least squares sense and depends upon the matrix inversion lemma to compute the inverse of the autocorrelation matrix at each iteration. If

$$\mathbf{F}_{\mathbf{k}}^{\mathrm{T}} = \mathbf{f}(\mathbf{k}-\boldsymbol{\Delta}) \ \mathbf{f}(\mathbf{k}-\mathbf{l}-\boldsymbol{\Delta}) \ \dots \ \mathbf{F}(\mathbf{k}-\mathbf{M}-\mathbf{l}-\boldsymbol{\Delta}) \tag{2.1}$$

is the input vector to the adaptive filter of length M, then the LMS algorithm [4] is given by

$$A_{r} = A_{r-1} + vF_{r} e(r),$$
 (2.2)

where A_r is the vector of filter coefficients,

$$\mathbf{e}(\mathbf{r}) = \mathbf{f}(\mathbf{r}) - \mathbf{F}_{\mathbf{r}}^{\mathrm{T}} \mathbf{A}_{\mathbf{r}-1}$$
(2.3)

as shown in Figure 1, and v is a convergence parameter. The SER algorithm [5] is given by

$$\mathbf{A}_{\mathbf{r}} = \mathbf{A}_{\mathbf{r}-1} + \mathbf{Q}_{\mathbf{r}} \mathbf{F}_{\mathbf{r}} \mathbf{e}(\mathbf{r})$$
(2.4)

where e(r) is defined by (2.3) with

$$Q_r = Q_{r-1} - (1/\delta) Q_{r-1} F_r F_r^T Q_{r-1}$$
 (2.5)

and

$$\delta = 1 + \mathbf{F}_{\mathbf{r}}^{\mathrm{T}} \mathbf{Q}_{\mathbf{r}-1} \mathbf{F}_{\mathbf{r}}, \qquad (2.5)$$



Figure 1. ADP Configuration

where A_r is again the coefficient vector, F_r is given by (2.1) and Q_r is the inverse of the autocorrelation matrix of the input to the ADP. Note that (2.5) and (2.6) update Q_r through the matrix inverse lemma.

Data Generation

In these simulations the test data was limited to a single vowel sound /a/. This was done in order that the dynamics of the adaptive filter itself may be observed without the added complexity of interaction of two phonemes. To generate the vowel sound used for the test, natural speech, sampled at 8000 Hz, was first analyzed by linear prediction analysis (LPA). An eighth order model was used. The resulting LPA parameters were then used to synthesize the test waveform. In this way the pitch period and spectral characteristics of the test signal were completely known. Figure 2 shows the waveform and spectrum of the test signal.



Figure 2. Test signal time and frequency plot.

Noise signals for the test were chosen to be simple but representative of the general types encountered. The three noise signals used were wideband, narrowband and a single sinewave. The wideband noise was generated by a Gaussian psuedo-random number generator. For the narrowband noise, the wideband generator output was passed through a digital resonator with center frequency at 2664 Hz and bandwidth of 400 Hz. The sinewave noise consisted of a single sinewave of constant amplitude at a frequency of 2664 Hz. Figure 3 shows the spectra of the noise signals used.

Pitch Estimate Error

لا ما ما ما ما ما

An essential part of the filtering technique is the estimate of the pitch period. It would be useful to know just how accurate the estimate must be to allow the filter



to perform satisfactorily. Figure 4 shows the results of a computer simulation of the ADP in which the pitch estimate given to the ADP is varied from the actual pitch period of the speech signal. The input signal was corrupted to 1.3 dB SNR at the input of the filter with the wideband noise specified in the previous section. The SNR of the output is plotted against the pitch estimate (in data sample periods) used by the ADP. Curves are shown for four and eight weight ADP's. One should note that the filter performs properly only when the estimated pitch falls within a window near the actual pitch period. The width of this window is approximately equal to the number of weights in the filter. The plot in Figure 5 extends from pitch estimates +60 to -60. A minus pitch estimate means that the delay is in the reference input side of the adaptive filter, see Figure 1, instead of the filter side. One should note that there are two more acceptance windows located at zero and -50. Only the windows located at +50 and -50 give improvement in SNR over the unfiltered signal. Note also that total signal improvement is higher if the pitch estimate is slightly lower than the actual signal pitch. These results imply that the pitch estimate need not be perfect but must fall in a window bounded on the high side by the actual pitch and with width approximately that of the filter weights. The results in Figure 4 and 5 are for the LMS algorithm. The SER algorithm shows the same window effect. The window

1



C





SNR, eight weight filter.

width was also found to be relatively insensitive to the type and intensity of the noise used to corrupt the test signal.

Narrowband Noise

Using real speech, Sambur has shown that this ADP configuration provides improvement in SNR in the case of wideband noise [3]. More challenging though is the case of narrowband noise such as is found in many industral and military environments. Figure 6 shows the overall performance of the filter with the LMS algorithm for the single vowel sound /a/ corrupted by varying intensities of wideband, narrowband and single sinewave noise. SNR of the output is plotted against the SNR of the input. Figure 7 shows the same results for the SER algorithm. In each case the filter was allowed to run to convergence before actual SNR calculations were made. One should note that in all but the sinewave noise case, some improvement in SNR was realized. Also note that the sub-optimal LMS algorithm provided more improvement in SNR than the SER algorithm.

Conclusion

A technique using an adaptive digital predictor with pitch period delay to reduce noise in speech signals has been examined. It has been shown for the case of a simple vowel sound that wideband and narrowband noise can be reduced. However, single sinewave interference was not



Figure 6. ADP performance, LMS algorithm, eight weight filter.



Figure 7. ADP performance, SER algorithm, eight weight filter.

reduced. The LMS algorithm provided better performance than the SER algorithm. Sensitivity to pitch period errors was also investigated. Criteria for the accuracy of the pitch period estimate necessary to maintain satisfactory performance were developed.

1

CHAPTER III

PITCH ESTIMATION BY ADAPTIVE FILTERING

Estimation of the fundamental frequency or "pitch" of a voiced speech signal is one of the basic steps in most speech analysis and speech encoding systems [17]. The currently popular pitch extraction methods are generally of two types, autocorrelation analysis and harmonic analysis. The autocorrelation technique performs an autocorrelation on the windowed speech data [6]. If the windowed data contains several pitch periods, the resulting autocorrelation function will have a peak at the delay corresponding to the pitch period. Rather than work with the raw speech data, most schemes will lowpass filter or center clip or both before performing the autocorrelation. This can enhance the result and lower the computation required. In some schemes the input to the autocorrelation process is the prediction residual from an appropriate linear prediction algorithm. In any case, the pitch is estimated by identifying the appropriate peak in the autocorrelation function.

Pitch extraction methods based on harmonic analysis usually start by performing a discrete Fourier transform on the windowed speech data [7]. The transformed data is examined to locate the line spectra features that are characteristic of pitch periodic time signals. Spacing of the line spectra features can then be used to estimate the pitch frequency.

This chapter presents a method for determining the fundamental frequency of a speech signal using a tapped delay line adaptive digital filter (TDLADF). Like the autocorrelation method, this technique tries to measure the time delay between successive pitch period waveforms. The use of TDLADF to estimate the time delay between two signals has been well developed for use in sonar applications [8, 91. Instead of estimating the delay between two different signals, this application uses the TDLADF to estimate the time delay between two parts of the same signal. The weights of the TDLADF are processed to determine the pitch estimate.

The General Time Delay Case

Figure 8 shows the general structure of a TDLADF with its filter input and reference input. If a signal is applied to the filter input and a delayed version of the same signal is applied to the reference input, the adaptive algorithm will minimize the error signal by adjusting the weights of the tapped delay line filter to approximate the unknown delay in the reference input. The resulting converged filter will then have a weight with a value of one at the delay line tap that correponds to the unknown delay and zeros in all the other weights. In most cases it is not necessary to wait until the adaptive algorithm converges to determine the time delay [8]. After only a few iteration steps, a scan of the weight values for the maximum positive value gives a very close estimate to the actual time delay.

The Pitch Estimation Configuration

Figure 9 shows the TDLADF configured for pitch period estimation. The filter input delay is established by the shortest pitch period expected (approximately 3-5 ms). The length of the tapped delay line filter is then determined by the longest pitch period expected (usually 15-20 ms). For example, an expected pitch period range from 5 ms to 15 ms would require a delay of 50 samples and filter length of 100 samples if the sample rate was 10kHz. In the configuration of Figure 9, the TDLADF approximates the delay between two successive pitch periods. Since successive pitch waveforms of natural speech are not exactly alike, the weights will never converge to a single impulse. Still, a clear peak in the weight function will be evident.

Figure 10 gives a plot of the TDLADF weights for the utterance "They shook hands for good luck." The seech data for this example was sampled at 8kHz. The example used a 40 weight adaptive filter with a 40 sample delay on the input. In this plot, the relative darkness of any point shows the value of a particular weight at that point in time. The more positive the value of the weight, the darker it appears on the plot. Since the presence of a pitch periodic signal will be indicated by a positive peak in the weights, only







Figure 9. TDLADF for pitch period estimation.

the positive values of weights are shown. The changes in the pitch period are clearly visible as the peak in the weight values tracks the delay between successive pitch periods. Note that a singular peak is not discernible during the unvoiced "sh" of the word shook.

The Adaptive Algorithm

The adaptive algorithm used for the example of Figure 10 is based on Widrow's LMS adaptive algorithm [4]. In this algorithm the weights of the tapped delay line are updated by the relation

$$W(n+1) = W(n) + u e(n) F(n)$$

where W(n) is the vector of weights, u is the convergence parameter, F(n) is the input signal vector, and e(n) is the error at step n. The LMS algorithm is one of the slower converging adaptive algorithms. This is not a problem in this application since complete convergence is neither required nor wanted for proper pitch estimation. It has been found that the weights in the LMS algorithm can more quickly track a varying pitch period if the algorithm is prevented from converging completely. This is done by adding a relaxation parameter to the weight update equation [19]. The modified update equation is

W(n+1) = vW(n) + ue(n) F(n)
where v is the relaxation parameter. Suitable values of v were found to be in the range of .90 to .99. Other adaptive algorithms may be suitable for this application. This study in concerned with the LMS algorithm.

Visible Detail

Since the TDLADF generates an estimate of the pitch at each incoming data sample, it can reveal subtle variations in the pitch period that would go unnoticed with other pitch estimation schemes. Both the autocorrelation and harmonic analysis techniques must work with data windows that are several pitch periods wide to get good results. Subtle variations in the pitch are averaged out in the process.

Figure 11 shows an expanded plot of the start of the word "they" from the example in Figure 10. The time plot of the actual signal is included with the TDLADF weight plot. Small changes in the pitch period from one pitch period to the next are evident in the weight plot. Note the jump in the pitch period associated with the change in voicing. Even more subtle features of the pitch period variations may be observed if an adaptive algorithm with faster convergence characteristic is used. Figure 12 shows the same example signal of Figure 11 processed by a sequential regression (SER) adaptive algorithm [20]. The SER algorithm is computationally intensive but does converge quickly and, therefore, more detail can be seen in the weight plot.



Figure 10. TDLADF weight plot for the sentence, "They shook hands for good luck."



Figure 11. Expanded weight plot for the start of the word "they".



Figure 12. Expanded weight plot using SER adaptive algorithm.

Pitch Period Resolution

Normally, the time resolution of the pitch period estimate would be limited to the data sampling interval. The estimate can be improved by using a parabolic interpolation over several weight values in the vicinity of the peak weight.

Noise Tolerance

The original motiviation for using an adaptive filter for pitch detection occurred while studying the effect of background noise on low bit rate speech coding systems. The adaptive technique was found to be very robust in the presence of various kinds of noise. Rigorous comparisons with other methods have yet to be made. Figure 13 gives the results of several tests that calculated the average pitch estimate error for various input signal to noise ratios. It has been found that the slow convergence rate of the LMS adaptive algorithm automatically provides much of the smoothing needed by other methods in the presence of noise [21].

Computational Considerations

From a computational standpoint, the TDLADF for pitch detection is very costly. The LMS algorithm requires approximately 200 multiplies per data sample. Since most of the pitch information is contained in the lower 1000 Hz of the speech signal, the data can be lowpass filtered and then down sampled. The number of computations is cut by about the down sampling rate. Down sampling for pitch detection is also common among autocorrelation methods.

One side effect of the down sampling is reduced pitch period resolution. Interpolation will help, but the TDLADF technique offers another alternative. Figure 14 shows a block diagram for a two stage adaptive pitch estimator. The "course" stage uses the lowpass filtered, down sampled version of the signal to obtain an approximate value of the This estimate is then used to adjust the pitch period. delay on the second "fine" stage. The second stage TDLADF, though not down sampled, need only be wide enough to accommodate the expected error in the first estimate. Processing the weights in the "fine" stage gives the same resolution as the original sampled case but at lower computational cost. Initial test with this structure showed that the computational savings were bought at the cost of slightly lower noise tolerance.

Possible Applications

Most interest in pitch estimation techniques centers around the real time encoding of speech data. The TDLADF does not seem to be well suited to this application due to the computational requirements. Recent development of VLSI circuits to implement the adaptive filter algorithms directly in hardware may eventually do away with that limitation.





Figure 14. Two stage TDLADF for pitch estimation.

. . . .

معتوجه ومشاهده ومعارضا ومنا

Other applications for pitch estimation are in disease diagnosis. Detailed study of the speech waveforms of some patients can aid in the detection of certain disorders of the vocal tract and nervous system [22]. The ability of the TDLADF to display subtle features in the pitch period may prove useful for such diagnosis. The computational load should not present a problem in this application since the analysis need not be done in real time.

, ,

Chapter IV NOISE SUPPRESSION METHODS FOR SPEECH APPLICATIONS

Use of linear predictive coding and other speech encoding techniques in military environments has revealed some disappointing limitations in the narrowband digital encoding of speech. Moderate acoustic background noise can severely degrade the overall system performance [2]. This Chapter investigates both frequency and time domain noise suppression techniques that appear to be effective on background noise with non-stationary narrowband characteristics.

Methods Investigated

Subtraction of an estimated noise spectrum in the frequency domain has been shown to be very effective on narrowband interference signals [23,24]. Two modifications to the original technique were investigated as a part of this study. First, placing zeros in suspected large interference bands was found to improve the subjective quietness of the speech. Secondly, it was also found that a dual time constant spectral averager allows for better noise spectrum estimation.

Time domain filters, designed to track and reject suspected narrowband noise signals, have the potential of being

more computationally efficient than frequency transform Four methods for designing time domain filters methods. were investigated. First, a time domain filter may be formed from the proper inverse transform of the inverse of estimated noise spectrum [25]. Another technique the searches the noise spectrum for the largest suspected interference signal and then designs a time domain notch filter with appropriate center frequency and bandwidth. The third technique is an adaptive predictor [18] which adaptively designs a time domain noise suppression filter based on a sample input of the background noise. The fourth is a modification of the adaptive predictor to allow for flatter passbands.

General Spectral Subtraction

The spectral subtraction noise reduction method is the process of subtracting an estimate of the noise power spectral density (PDS) from the corrupted input signal's PSD in an attempt to improve the signal-to-noise ratio (SNR). This subtraction is adaptive in that the estimate of the noise PDS is updated during the absence of speech. The algorithm involves: 1) breaking the input signal into frames and estimating the PSD of each frame by an FFT, 2) updating the estimate of the noise PSD if no speech is present, 3) subtracting the current noise PSD from the signal PSD and 4) transforming the frequency domain result back into the time domain using the input signal's original phase information.

The Noise PSD Estimate

The performance of the spectral subtraction technique depends a great deal on the accuracy of the estimation of the noise PSD. Errors in the noise PSD estimate usually show up as tone burst or spectral artifacts in the output signal. If the noise signal is assumed to be stationary on a short time basis, then some form of smoothing of the frame to frame variations will improve the noise PDS estimation. A first order recursive smoother was used to average each descrete frequency power estimate over several frames. The recursive smoother computes the running average of each discrete frequency by

 $E_{n}(k) = P_{n}(k) + c(E_{n-1}(k) - P_{n}(k))$

where E_n is the recrusive estimate of the PSD based upon P_n , the noise PSD in the nth frame, c is a constant which sets the effective time constant for the average, and k is the decrete frequency index. No smoothing was done between adjacent discrete frequencies. The appropriate time constant for the smoother depends on the frame to frame variability of the noise. A fixed time constant of about 5 frames worked reasonably well. However, a dual time constant scheme based upon adjacent power estimates proved better. If the new power estimate of a particular deicrete frequency is greater than the last estimate then a shorter time constant (1-2 frames) is used. If the new estimate is less than the current estimate then a long time constant (4-5 frames) is applied. This "fast attack -- slow decay" scheme reduced the number of spectral artifacts due to random variations in the noise.

Frequency Domain Notch Filter

The difficulty with very high intensity narrowband noise is that the spectral artifacts due to the noise PSD estimation errors can be very large in the vicinity of the noise concentration. Smoothing is impractical to suppress these large artifacts. It was found, however, that total elimination of the speech PSD in the area of the large narrowband noise reduced the number of large artifacts without greatly distorting the speech signal.

After the noise PSD estimate is subtracted from the input signal PSD, zeros are placed in the resulting PSD at discrete frequency locations that have abnormally large noise power estimates. This eliminates any large noise artifacts that might remain due to error in the noise PSD estimate. Emperical experiments showed that any discrete frequency in the noise estimate that has a power level over four times the average over all frequencies was a good candidate for a frequency domain zero.

Filter Design by Inverse Transform

Once the noise PSD has been estimated, a suppression filter for the noise is easily designed by taking the inverse DFT of the inverse of the PSD [25]. The resulting impulse response forms the weights of a transversal

filter. Linear phase response may also be obtained by proper choice of phase.

To keep the computational load reasonable, the transversal filter must be limited to between 10 and 20 weights. This requires sampling the noise PSD at only 10 to 20 equally spaced frequencies. The resulting poor frquency resolution allows very narrowband noise to be ignored if it falls between frequency samples. For this reason, no performance comparisons were made for this filter technique.

Filter Design by Notch Placement

Since many of the background noises encountered contain one predominant narrowband component, it is unnecessary to use a filter with as many coefficients as is generally required by the inverse transform method. A much more heuristic, yet very practical approach, is to direct the noise filter design process directly toward the major spectral component of the noise.

The method considered here first analizes the estimated power spectrum of the noise to identify the maximum. This maximum is then the primary target of the noise filter. Using the amplitude and bandwidth of the local maximum found in the noise spectrum, a recursive notch filter is designed. The notch filter used in this research was a bilinear transformation of the transfer function,



where ω_0 is the frequency of infinite attenuation and ω_b is the -jdB bandwidth centered about ω_0 . This filter was chosen because its gain at the high and low frequency extremes approach unity resulting in minimal amplitude distortion for portions of the spectrum that are not near the center frequency.

The values of ω_0 and ω_b are obtained by analizing that estimated noise PSD in the following manner. First, the current estimate of the noise PSD is searched for the largest peak that is above a set threshold value. The notch filter center frequency is then set to the frequency of this peak. The bandwidth of the notch filter is set equal to the width of the peak that extends above the threshold value. The threshold value was chosen to be four times the average value of the current noise PSD. The factor four was chosen after many experiments with various noise sources.

The advantages of this recursive notch filter method center around the speed with which such a filter may be implemented. The recursive filter requires very few operations per data sample and the filter parameters need to be calculated only once per frame. Still, the most time consuming operation is the calculations of the noise PSD estimate. Hardware FFT processors should make this practical.

Some background noise sources, such as helicopter noise, contain more than one narrowband noise component. The above method can be extended by cascading several notch

filters. Each filter would adapt and track each narrowband noise.

Filter Design by Adaptive Filter

Design of noise suppression filters by linear prediction has been shown effective for narrowband noise [25]. Figure 15 shows the block diagram of an adaptive predictor for noise suppression in speech. When a noise signal is applied, the adaptive algorithm adjusts the weights of the transversal filter to minimize the error signal. As the algorithm converges, the transfer function of the system approximates the inverse filter that would be required to suppress the input noise signal. When speech is detected, the adaptive algorithm is turned off and the filter weights are held constant at their current values. Adaptation resumes when speech is no longer indicated. For this research, the least mean square (LMS) algorithm [18] was used for the automatic adjustment of the weights.



Figure 15. Adaptive Predictor for Speech Filtering



Figure 16. Adaptive Transfer Filter

In these experiments, white psuedo-random noise is used for the bias noise source. The flattness of the passbands and the depth of the stopbands are controlled by the bias noise to input noise ratio. The power of the noise bias was set based on experimental results to be one-fourth the expected average power of the input speech signal. Flatter passbands and deeper notches could also be achieved by using more weights in the filter.

Comparison of Filter Methods

Construction destruction

A series of tests were conducted to evaluate each of the filter methods previously described except the inverse transform method. The noise used for these experiments was a sample of noise recorded in a RH-53 helicopter. All the tests were run at a simulated sample rate of 8000 Hz. The four techniques evaluated were: a) Spectral subtraction (SSB) with noise PSD smoothing the frequency domain notch placement, b) Adaptive notch placement (ANF) using two second order notch filters, c) Adaptive predictor filter

(APF) using a 15 weight filter, d) Adaptive transfer filter (ATF) using a 15 weight filter. The performance of each filter was compared using a signal-to-noise ratio calculation based on the average spectral error in the corrupted signal before and after filtering. The spectral error was averaged over the speech portions of the input signal only. Also, performance was compared by computing the average log area ratio (LAR) error for a 10th order linear predictor [26].

Objective Comparisons

Figure 17 shows the input and output signal-to-noise ratios for various levels of input noise. Figure 18 shows the LAR error for various levels of input noise. Note that only the SSB and ATF methods consistantly improve SNR and LAR error. Note also that the ATF scored best in SNR improvement but the SSB proved better in LAR error reduction. The deep narrow notches formed by zero placement in the SSB filter cause much spectral error but are easily smoothed over by the relatively low order linear predic-The APF performed poorly due to its tendency to distor. tort the spectrum in the passbands of the filter. The ANF successfully suppressed the large narrowband noise components but did not filter the subtiler narrowband noise that was below its threshold.



Figure 18. Log Area Ratio Performance

Subjective Comparisons

Informal listening tests indicated that all the methods evaluated reduced the perceived level of background noise. Casual listening greatly favored the SSB technique. The SSB method was the only one to remove entirely the large narrowband components of the test helicopter noise. All the other methods had audible residuals of the major noise components. The ATF method had the least distortion, while the APF tended to emphasize the higher frequencies.

Conclusions

Of the four filter methods evaluated, none ofice a complete cure for background noise in speech. Spectral subtraction with zero placement and the adaptive transfer filter were the most effective. The spectral subtraction technique may have an advantage in applications with linear predictive coding since it performed best in the log area ratio error evaluation. Formal subjective tests need to be done to determine the effect these filters have on the intelligibility of noisy speech. The adaptive notch placement scheme and the adaptive predictor were not very effective on the test signals used but might be useful in other applications.

CHAPTER V

ENHANCEMENT OF SPEECH SIGNALS

BY TWO DIMENSIONAL SIGNAL PROCESSING

Recently there has been considerable interest in investigating new avenues for removing high ambient noise in speech [1, 2, 27-31]. Chapter IV presented a discussion and evaluation of several filtering techniques for suppressing background noise in speech signals. Spectral subtraction is an effective method for removing narrowband noise from However, there are difficulties with the speech signals. technique when the noise is wideband or random in nature. For example, consider the white noise case. The average value of the noise spectrum can be easily subtracted from the corrupted signal but the frame to frame variations of the noise will still appear in the output signal as chirps or musical noise. Over estimating the average noise level is helpful in removing this residual noise but only at the cost of removing more speech signal.

The random nature of the chirp noise suggests that some kind of spectral smoothing might be useful to suppress the residual random variations that remain after subtraction of the average noise level. Smoothing of noisy spectral data to minimize the effects of residual artifacts in spectral

subtraction has been shown to be useful for image restoration applications [27]. The power spectrum of each frame of data can be smoothed using any one of several techniques. The linear predictive coding process itself offers some smoothing since only a limited number of poles are available for modeling the signal. In channel vocoder systems, some spectral smoothing can be added by slight overlap of the analysis channels [28]. Frame to frame correlation of speech frequency data has also been used in channel vocoders for noise suppression [29]. In such a system, the output of each frequency channel is lowpass filtered to remove any rapid variations. Logarithmic filtering of the envelope of individual frequency channels has been shown to be useful for enhancing speech corrupted by white noise [30]. Some frame to frame smoothing is introduced in spectral subtraction if the frame size and overlap are made relatively large [31].

The acoustic tube model for speech production implies that the power spectrum of any one frame of data will be fairly continuous as a function of frequency. Also, since the speech parameters do not change rapidly, the frame to frame variations of amplitude of any one frequency will also be continuous. This dual continuity of speech spectral data in both time and frequency suggests that some type of two dimensional filtering might be applicable.

When the speech power spectrum data is displayed as a spectrogram, an image is formed with dimensions of time and

frequency. The spectrogram makes visible both the time and frequency correlation of the speech spectrum. This image can then be processed using two dimensional techniques to smooth, improve contrast, or otherwise enhance the spectral features. The resulting processed spectrum can then be combined with the original phase data and inverse transformed to recover the speech signal. This procedure has been investigated for detecting single tones in white noise [32]. This chapter presents some interesting results on applying two dimensional processing techniques to achieve speech enhancement.

Two Dimensional Representation

The speech signal is first converted to the short-time Fourier transform (STFT) domain by suitable sampling, windowing and discrete Fourier transformation. The transform results in a complex two-dimensional function, which can be represented in the form $M(k,n) \angle P(k,n)$, where M(k,n)corresponds to the magnitude and P(k,n) corresponds to the phase with k and n respectively representing the frequency and the time indices. The plot of M(k,n) is usually called Since M(k, n) is an image, all the image spectrogram. processing techniques are available to 'clean' the image. Figure 19 shows a simplified block diagram of a twodimensional filter approach. The two blocks in the middle simply identify that the magnitude and the phase functions are modified using two separate two-dimensional filters.



Barra and the

Ē

Figure 19. Block diagram for a two-dimensional speech filter.

ويترفق ومراور

The resulting modified magnitude and phase functions are then recombined using an appropriate inverse transform and synthesis method [6] to form the time domain signal.

Filtering in the Double Transformed Domain

To introduce this approach, consider the spectrogram M(k, n) displayed in Figure 20a for the uncorrupted speech

"Don't gift wrap the tall glass. They shook hands for good luck."

Now consider the two-dimensional discrete Fourier transform of M(k, n),

$$\mathcal{F}[M(k, n)] = FM(k, n) / FP(k, n)$$
(5.1)

The function FM(k, n), corresponding to the spectrogram in Figure 20a, is displayed in Figure 20b. It is clear that the two dimensional transform of M(k, n) does not have the usual connotation, as M(k, n) is a function of time and frequency. However, the function in (5.1) exhibits some interesting properties that are amenable for noise filtering. Figure 21a displays the spectrogram for the two sentences given earlier when the speech is corrupted by white noise to a SNR of about 0 dB. Figure 21b displays the function FM(k, n) in (5.1) for the spectrogram in Figure 21a. It is clear from Figures 20a and 21a that there is hardly any resemblance between these spectrograms even though the speech content is the same. However, the story



a. CLEAN SPEECH SPECTROGRAM

is different for FM(k, n). Figures 20b and 21b clearly show high energy concentrations near the origin. This observation is used in the following.

An example of how a speech signal may be filtered by a two-dimensional modification of the spectrogram is shown in Figure 22. The first three parts of the figure, Figures 22a, b, and c, show the original speech spectrogram, the noisy speech spectrogram, and the two dimensional Fourier transform of the noisy speech spectrogram. Figure 22d is obtained from Figure 22c by using a two-dimensional filter. This simply corresponds to passing the noisy spectrogram through a band pass filter. Removing the high frequencies smooths the spectrogram while removing the low frequencies enhances the contrast between the background and the speech Figure 22e corresponds to the filtered speech signal. spectrogram, which is obtained from Figure 22d by inverse transforming. It is clear from Figures 22a and 22e that the filtered spectrogram shows a great deal more features of the original speech signal than the original noisy speech The results presented in Figure 22 are for spectrogram. magnitude filtering only. The original noisy phase can be used for reconstructing the time domain signal.

The results presented in this chapter are still at a preliminary stage. However, the results indicate that there is a significant potential in studying these concepts. In informal listening evaluation, the two-dimensional processed speech sounds quieter with some added clarity. Formal







e. RESULTING FILTERED SPEECH SPECTROGRAM



Figure 22. Results at each step of a two-dimensional filter process.

. - . •

listening tests need to be conducted to justify the results. Other aspects that need to be investigated are in the area of improving the phase estimate and making use of the improved noisy phase in reconstructing the phase. The work done by Oppenheim et al [33] should be helpful in this endeavor. At the same time, the study done by Wang and Lim [34] that the phase is unimportant in speech enhancement should put a different light on the phase in the reconstruction of speech.

Conclusion

This chapter presented some preliminary results on using image processing techniques for speech enhancement. The basic idea is that the two dimensional Fourier transforms of clean and noisy speech spectrograms have most of the speech energy concentrated near the origin and the spectrogram constructed from this high energy area obtained from the noisy spectrogram has more features of the original speech signal than the original noisy speech spectrogram. From the informal listening tests, the two dimensinal processed speech sounds quieter with some added clarity. Further work is necessary in this area.

CHAPTER VI DISCRETE TIME ESTIMATION

The basic problem of prediction is important in many areas, such as speech processing [6], seismic signal processing [35], control theory [36] and many others [37]. The problem is usually reduced to finding the inverse of the data covariance matrix. Considerable deal of work has been done for the case of stationary process wherein the covariance matrix results in a Toeplitz-type matrix. The special structure of the Toeplitz matrix of order M allows for an inversion in $O(M^2)$ operations (multiplications and additions) compared to $O(M^3)$ operations required for the inversion of an arbitrary matrix [38]. In this chapter, a prediction method is presented for the case of non-Toeplitz covariance matrices. A brief review of the problem is presented below. Given the data y_i , $0 \le i \le N-1$, find the coefficients a_i , $1 \leq i \leq M$, in an Mth order predictor

$$\hat{y}_{p} = -\sum_{i=1}^{M} a_{i} y_{p-i}, \quad 0 \le p \le N-1$$
(6.1)

such that the mean squared error

Error =
$$\sum_{p} (y_{p} - \hat{y}_{p})^{2}$$
 (6.2)

is minimized. In matrix form, the least squares problem can be formulated by starting with

$$\begin{bmatrix} d_{0} \\ d_{1} \\ \cdot \\ \cdot \\ \cdot \\ d_{N-1} \end{bmatrix} = \begin{bmatrix} y_{0} \cdot \cdot \cdot \cdot & 0 \\ y_{1} \cdot y_{0} \cdot \cdot \cdot & 0 \\ \cdot \cdot \cdot \cdot \cdot & \cdot & 0 \\ \cdot \cdot \cdot \cdot \cdot & \cdot & \cdot \\ y_{0} \\ \cdot \cdot & \cdot \cdot \cdot & y_{0} \\ \cdot & \cdot & \cdot & \cdot \\ y_{N-1} \cdot y_{N-2} \cdot \cdot & y_{N-M} \end{bmatrix} \begin{bmatrix} a_{1} \\ a_{2} \\ \cdot \\ a_{M} \end{bmatrix}$$
(6.3)

If $d_i = y_{k+i}$, $k \ge 1$, then we identify the prediction as the k step prediction using an mth order predictor. In symbolic matrix form, (6.3) can be written as

$$\underline{\mathbf{d}}_{\mathbf{N}} = \mathbf{Y}_{\mathbf{M}}^{\mathbf{T}} \underline{\mathbf{a}}_{\mathbf{M}} \tag{6.4}$$

where the bars below d and a denote that they are vectors.

It is well known that the least squares solution of (6.4) is given by

$$\underline{\mathbf{a}}_{\mathbf{M}} = (\mathbf{Y}_{\mathbf{M}} \ \mathbf{Y}_{\mathbf{M}}^{\mathbf{T}})^{-1} \ \mathbf{Y}_{\mathbf{M}} \ \underline{\mathbf{d}}_{\mathbf{N}}$$
(6.5)

The recursive solutions for (6.5) have been developed by Lee, Morf, Kailath, Friedlander and others [39-46]. The method presented here is based purely on matrix algebra and can be related to the earlier work. The number of operations required by this method is $O(M^2)c$, where c is a constant.

Set up of Recursion Equations

The solution in (6.5) will be computed in the following manner. Let

$$\underline{\mathbf{E}}_{p}^{T} = [\mathbf{e}_{p+k-1} \ \mathbf{e}_{p+k-2} \ \cdots \ \mathbf{e}_{0}]$$
(6.6)

where

$$\mathbf{e}_{i} = \mathbf{d}_{N-1-i} \tag{6.7}$$

Let

$$\underline{\mathbf{E}}_{\mathbf{p}} = \mathbf{B}_{\mathbf{p}}^{\mathsf{T}} \underline{\mathbf{\mu}}_{\mathbf{p}} \tag{6.8}$$

where

$$\frac{\mu_{p}}{\mu_{p}} = \begin{bmatrix} \mu_{p}(o) & \dots & \mu(p-1) \\ p-1 \end{bmatrix}$$
(6.9)

$$B_{\rho}^{\mathsf{T}} = \begin{bmatrix} 0 & 0 & \cdots & y_{0} \\ 0 & 0 & \cdots & y_{0} & y_{1} \\ \cdots & \cdots & \cdots & \cdots \\ y_{0} & y_{1} & \cdots & \cdots \\ y_{0} & y_{1} & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ y_{K} & y_{K+1} & \cdots & y_{K+p-1} \end{bmatrix}$$
(6.10)

To relate (6.8) to (6.3), we need to define K = N-M in (6.10).

The least squares solution of (6.8) can be obtained from .

$$(\mathbf{B}_{\mathbf{p}} \ \mathbf{B}_{\mathbf{p}}^{\mathsf{T}}) \ \underline{\boldsymbol{\mu}}_{\mathbf{p}} = \mathbf{B}_{\mathbf{p}} \ \underline{\boldsymbol{\Xi}}_{\mathbf{p}}$$
(6.11)

and is

r

$$\underline{\mu}_{p} = (B_{p} B_{p}^{T}) B_{p} \underline{\underline{E}}_{p}$$
(6.12)

and the solution in (6.5) is given by

$$\underline{\mathbf{a}}_{\mathbf{M}}^{\mathrm{T}} = \left[\mu(\mathbf{M}-1) \ \mu_{\mathbf{M}}(\mathbf{m}-2) \ \dots \ \mu_{\mathbf{M}}(\mathbf{0})\right]$$
(6.13)

The vector $\underline{\mu}_p$ in (6.9), derived in (6.12), will be computed recursively and the derivation for this is discussed in the next two sections.

Structure of B_p B_p^T

The matrix B_p in (6.10) can be expresseed in two forms. First,

$$B_{p} = \begin{bmatrix} 0 & B_{p-1} \\ y_{0} & \frac{\chi^{T}}{p-1} \end{bmatrix}$$
(6.14)

where

$$\underline{\mathbf{x}}_{\mathbf{p}-\mathbf{1}}^{\mathsf{T}} = \begin{bmatrix} \mathbf{y}_{\mathbf{1}} & \mathbf{y}_{\mathbf{2}} & \cdots & \mathbf{y}_{\mathsf{K}+\mathbf{p}-\mathbf{1}} \end{bmatrix}.$$
(6.15)

Second,

$$B_{p} = \begin{bmatrix} A_{p-1}^{T} & y_{K} \\ B_{p-1} & C_{p-1} \end{bmatrix}$$
(6.16)

where

$$\mathbf{A}_{p-1}^{\mathsf{T}} = \begin{bmatrix} 0 & 0 & \dots & 0 & y_0 & \dots & y_{\mathsf{K}-1} \end{bmatrix}$$
(6.17)

and

$$\underline{c}_{p-1}^{T} = \begin{bmatrix} y_{K+1} & \cdots & y_{K+p-1} \end{bmatrix}.$$
(6.18)

From (6.14), we can write

$$B_{p} B_{p}^{T} = \begin{bmatrix} B_{p-1} B_{p-1}^{T} & B_{p-1} \frac{X}{p-1} \\ \frac{X}{p-1} B_{p-1}^{T} & y_{0}^{2} + \frac{X}{p-1} \frac{X}{p-1} \end{bmatrix}.$$
 (6.19)

From (6.16), we can write

$$B_{p} B_{p}^{T} = \begin{bmatrix} A_{p-1}^{T} A_{p-1} + y_{K}^{2} & A_{p-1}^{T} B_{p-1}^{T} + y_{K} C_{p-1}^{T} \\ B_{p-1} A_{p-1} + C_{p-1} y_{K} & B_{p-1} B_{p-1}^{T} + C_{p-1} C_{p-1}^{T} \end{bmatrix}.$$
 (6.20)

Equations (6.19) and (6.20) will be used in the following for the proposed recursive algorithm.

Special Structure of Equation (6.11)

For ease of notation, let us define

$${}^{B}_{p-1} \xrightarrow{A}_{p-1} = \underline{T}_{p-1},$$
 (6.21)

$$B_{p-1} = \frac{1}{p-1}$$
 (6.22)

$$\underline{C}_{p-1} y_{K} = \underline{G}_{p-1}$$
 (6.23)

$$B_{p-1} = H_{p-1}, \qquad (6.24)$$

$$B_{p-1} = R_{p-1}$$
 (6.25)

Equation (6.11) can now be expressed for stages (p-1) and p respectively by

$$R_{p-1} = H_{p-1} = (6.26)$$

$$\begin{bmatrix} R_{p-1} & L_{p-1} \\ L_{p-1}^{T} & r_{p} \end{bmatrix} \begin{bmatrix} \underline{\mu}_{p} & (1) \\ \mu_{p} & (2) \end{bmatrix} = \begin{bmatrix} H_{p-1} \\ y_{0} & e_{p+K} & + \underline{\chi}_{p-1}^{T} & \underline{E}_{p-1} \end{bmatrix}$$
(6.27)
where we have used (6.19) for R_{p} and
 $r_{p} = y_{0}^{2} + \underline{\chi}_{p-1}^{T} & \underline{\chi}_{p-1}$. (6.28)

Recursive solutions for (6.27) - (6.28) require the knowledge of the solution of the following equations

$$R_{p-1} \quad \underline{\alpha}_{p-1} = \underline{T}_{p-1}, \tag{6.29}$$

$$R_{p-1} = \frac{\delta}{p-1} = \frac{1}{p-1}$$
 (6.30)

 $R_{p-1} \simeq \frac{1}{2} q_{p-1} \simeq \frac{1}{2} q_{p-1}$ (6.31)

This requires a few intermediate steps, and are discussed below.

Intermediate Steps

Consider the equations

$$\begin{bmatrix} R_{p-1} + C_{p-1} C_{p-1}^{T} \\ \alpha_{p-1} = T_{p-1}$$
(6.32)

$$\begin{bmatrix} R_{p-1} + C_{p-1} C_{p-1}^{T} \end{bmatrix} \stackrel{\delta'}{-p-1} \stackrel{=}{-p-1}$$
(6.33)

$$\begin{bmatrix} R_{p-1} + C_{p-1} C_{p-1}^T \\ Y_{p-1}^T = C_{p-1} \end{bmatrix}$$
It is well known that [47]

$$\left[\begin{array}{c} R_{p-1} + \underline{C}_{p-1} & \underline{C}_{p-1}^{\mathsf{T}} \end{array} \right]^{-1} = \begin{array}{c} -1 & -1 & -1 & -1 & -1 \\ R_{p-1} & -1 & \underline{C}_{p-1} & \underline{C}_{p-1}^{\mathsf{T}} & R_{p-1} & (6.35) \end{array}$$

with

$$\Delta_{p-1} = 1 + \underline{C}_{p-1}^{T} R_{p-1}^{-1} \underline{C}_{p-1} . \qquad (6.36)$$

Using the solutions for (6.29) - (6.31), we can write

$$\Delta_{p-1} = 1 + \left(\frac{1}{y_{K}^{2}}\right) \frac{G_{p-1}^{T}}{y_{p-1}} \qquad (6.37)$$

$$\underline{\alpha}_{p-1} = \underline{\alpha}_{p-1} - \left(\frac{1}{\Delta_{p-1}} y_{K}^{2}\right) \underline{\gamma}_{p-1} \quad (\underline{G}_{p-1}^{T} \ \underline{\alpha}_{p-1})$$
(6.38)

$$\underline{\delta}_{p-1}^{\prime} = \underline{\delta}_{p-1} - \left(\frac{1}{\Delta_{p-1}} y_{K}^{2}\right) \underline{\delta}_{p-1} \left(\underline{G}_{p-1}^{T} \underline{\delta}_{p-1}\right)$$
(6.39)

$$\underline{Y}_{p-1}^{\prime} = \underline{Y}_{p-1} - \left(\frac{1}{\Delta_{p-1}} y_{K}^{2}\right) \underline{Y}_{p-1} \quad (\underline{G}_{p-1}^{T} \underline{Y}_{p-1}) \quad .$$
(6.40)

Now, we can consider the solutions of (6.29) - (6.31) for the p-th stage. For the p-th stage, (6.29) - (6.31) can be expressed in the form

$$\begin{bmatrix} R_{p-1} & L_{p-1} \\ L_{p-1}^{T} & r_{p} \end{bmatrix} \begin{bmatrix} \alpha_{p}(1) \\ \alpha_{p}(2) \end{bmatrix} = \begin{bmatrix} \frac{T}{p-1} \\ \frac{X}{p-1} & A_{p-1} \end{bmatrix}$$
(6.41)

$$\begin{bmatrix} r_{0} & \frac{T}{p-1} + \frac{G}{p-1} \\ -\frac{T}{p-1} + \frac{G}{p-1} & R_{p-1} + \frac{C}{p-1} & \frac{C}{p-1} \end{bmatrix} \begin{bmatrix} \delta_{p}(1) \\ \frac{\delta_{p}(2) \\ -\frac{\delta_{p}(2) \\ -\frac{\delta_{p}}{p-1} & \frac{C}{p-1} & \frac{V_{p+K}}{p-1} \end{bmatrix} = \begin{bmatrix} \frac{A}{p-1} + \frac{V_{p-1}}{p-1} + \frac{V_{p+K}}{p-1} \end{bmatrix} (6.42)$$

$$\begin{bmatrix} R_{p-1} & \frac{L}{p-1} \\ -\frac{T}{p-1} & r_{p} \end{bmatrix} \begin{bmatrix} Y_{p}(1) \\ Y_{p}(2) \end{bmatrix} = \begin{bmatrix} \frac{G}{p-1} \\ -\frac{V_{p}}{p+K} \end{bmatrix}$$

$$(6.43)$$

Note that in (6.42), equation (6.20) is used for $R_{\rm p},$ where

$$r_{0} = \frac{A_{p-1}^{T}}{A_{p-1}} + y_{K}^{2}$$
 (6.44)

The solutions for equations (6.41) - (6.43) are given below. First

$$\alpha_{p}(2) = \frac{(\underline{x}_{p-1}^{T} \underline{A}_{p-1}) - \underline{\delta}_{p-1}^{T} \underline{T}_{p-1}}{r_{p} - \underline{\delta}_{p-1}^{T} \underline{L}_{p-1}}$$
(6.45a)

$$\underline{\alpha}_{p}(1) = \underline{\alpha}_{p-1} - \alpha_{p}(2) \underline{\delta}_{p-1} \qquad (6.45b)$$

Second,

$$\delta_{p}(1) = \frac{\left[\frac{A_{p-1}^{T} \times x_{p-1} + y_{K} \times y_{p+K}\right] - (\underline{\alpha}_{p-1}' + \underline{\gamma}_{p-1}')^{T} (\underline{L}_{p-1} + \underline{C}_{p-1} \times y_{p+K})}{r_{0} - (\underline{\alpha}_{p-1}' + \underline{\gamma}_{p-1}')^{T} (\underline{T}_{p-1} + \underline{G}_{p-1}')} \quad (6.46a)$$

$$\delta_{p}(2) = -\delta_{p}(1) \left(\underline{\alpha}_{p-1}^{\prime} + \underline{Y}_{p-1}^{\prime}\right) + \underline{\delta}_{p-1}^{\prime} + \underline{Y}_{p-1}^{\prime} - \frac{y_{p+K}}{y_{K}}. \qquad (6.46b)$$

Third,

$$Y_{p}(2) = \frac{y_{K} y_{p+K} - \frac{\delta}{p-1} \underline{G}_{p-1}}{r_{p} - \frac{\delta}{p-1} \underline{L}_{p-1}}$$
(6.47a)

$$\chi_{p}(1) = \chi_{p-1} - \gamma_{p}(2) \chi_{p-1}$$
 (6.47b)

The proof for these is rather straight forward, and is illustrated below for (6.41). Premultiplying both sides of (6.41) by

$$\begin{bmatrix} - L_{p-1}^{T} R_{p-1}^{-1} & 1 \end{bmatrix}$$
(6.48)

and using (6.30) and simplifying, we have (6.45a). Considering the first set of equations in (6.41), we have

$$R_{p-1} \underline{\alpha}_{p-1} + \underline{L}_{p-1} \alpha_p(2) = \underline{T}_{p-1}$$
(6.49)

By using (6.29) and (6.30) in (6.48), we have

$$R_{p-1}\left[\underline{\alpha}_{p-1} + \alpha_{p}(2) \underline{\delta}_{p-1} - \underline{\alpha}_{p-1}\right] = 0$$

and (6.45b) follows. In a similar manner, the others can be shown.

Final Solution

Using the analysis discussed above, we can obtain the solution of (6.27), and is

$$\mu_{p}(2) = \frac{\left(y_{o} \ e_{p+K} + \underline{x}_{p-1}^{T} \ \underline{E}_{p-1}\right) - \underline{\delta}_{p-1}^{T} \ \underline{H}_{p-1}}{r_{p} - \underline{\delta}_{p-1}^{T} \ \underline{L}_{p-1}}$$
(6.50a)

 $\mu_{p}(1) = \mu_{p-1} - \mu_{p}(2) \delta_{p-1}$ (6.50b)

It is clear that the solutions for equations (6.26), (6.29) - (6.31) are assumed to be known for stage (p-1). The solutions for stage p are given in (6.45) - (6.47) and (6.50). It is interesting to point out that we have used four sets of equations to solve the generalized prediction problem as compared to one set of Teoplitz-type normal equations for the stationary process. Parallels can be seen between the above solution for one set of equations and the classical solutions of Toeplitz-type normal equations [48].
CHAPTER VII

DESIGN OF FAST RECURSIVE ESTIMATORS

In many signal processing applications, computational speed is of considerable importance. It is often desired to have an excellent estimate available, and generated rapidly from a large amount of data. This means that only a limited number of multiplications are allowed in obtaining the estimate. It is therefore important that the multiplications be selected so that they are maximally effective in generating a good estimate. The coefficient involved in the multiplication should be optimal, and the data multiplied should be optimally selected with regard to some performance measure. It is reasonable, if much data is to be processed, to spend a great deal of design effort in solving the optimization problem. Of course this design effort, if it is very involved, must be done before the data becomes available, i.e. must not require the data but only a statistical i t knowledge of the data.

In this Chapter we propose a recursive digital filter with a fixed number of multiplications as our estimating structure. In certain cases [49], primarily when state models are available, recursive estimators have proved to be a computationally efficient means of solving the normal

equations and generating the best linear estimate. 0fcourse the Kalman filter [50] is the best known of such results. The filter structure we propose is not in general optimal. It is in part dictated by the allowable number of multiplications. The coefficients are optimized, and the structure is optimized to the extent that the best measurements are selected. Optimal measurement strategies have previously been considered in control and estimation applications with state models and white noise [51-54]. In this Chapter however, we only assume the availability of a statistical knowledge of the observation set and its relation to the signal to be estimated. At each stage, a subset of the data is to be summed and multiplied by the best coefficient, and combined with a linear combination of previous estimates to provide the new estimate. The parameter optimization component of the design is relatively simplified, with no difficult matrix inversions required to obtain the best set of coefficients. The choosing of the best data selection vector is shown to be related to a classical family of integer programming problems which have received much attention [12, 13, 55], and include the famous "Traveling Salesman Problem." For a good review of the integer programming problem imbedded in this study the reader is referred to [12]. The most computationally efficient means of solving the problem available at this time may be found in [13].

Problem Statement

Consider the situation in which a large volume of data is available. This data is collected in one large vector, y, having elements y_i . A recursive estimator is to be designed, of the form

$$\hat{x}_{j} = \alpha_{j} \hat{e}_{j}^{T} y + \hat{\Sigma} \hat{\gamma}_{j} \hat{x}_{j-1} j = M+1, M+2, ...$$
(7.1)

where α_j and γ_{ji} are scalars to be selected in order to minimize the performance measure

$$J_{j} = E\{(x-\hat{x_{j}})^{2}\}$$

(7.2)

The parameter x is the unknown signal to be estimated by processing the data according to (7.1). The vector, ϵ_j , is a selection vector whose elements are only 0 or 1. Thus the estimator has its complexity restricted to M+1 multiplications per iteration. The problem is to select the coefficients, α_j and γ_{ji} , and the selection vector, ϵ_j , in order to obtain the best possible performance. The only assumption required for the design is that

$$\mathbf{P}_{\mathbf{y}\mathbf{y}} = \mathbf{E}\{\mathbf{y}\mathbf{y}^{\mathrm{T}}\}$$
(7.3)

 $\mathbf{P}_{\mathbf{yx}} = \mathbf{E}\{\mathbf{yx}\} \tag{7.4}$

are known quantities. The design is to be carried off line, so that the only significant calculations which must be done in real time are the M+1 multiplications. For $2 \le j \le M$, the structure of the filter is

$$\hat{\mathbf{x}}_{j} = \alpha_{j} \varepsilon_{j}^{T} \mathbf{y} + \sum_{\substack{i=1\\i=1}}^{j-1} \hat{\mathbf{x}}_{j-i}$$

(7.5)

while x_1 is of the form

 $\hat{\mathbf{x}}_1 = \alpha_1 \varepsilon_1^T \mathbf{y}$

(7.6)

Since these initial estimates require fewer multiplications, they could be generated more quickly than the later estinates provided by (7.1). Thus a different and variable period between estimates could be used during start up.

The problem will be solved by using the fact that

$$\min_{\mathbf{a}_{j},\{\gamma_{ji}\},\varepsilon_{j}} \min_{\mathbf{b}_{j}} \min_{\mathbf{b}_{j}} \min_{\mathbf{b}_{j}} \min_{\mathbf{b}_{j},\{\gamma_{ji}\}} \min_{\mathbf{b}_{j},\{\gamma_{ji}\}} \max_{\mathbf{b}_{j},\{\gamma_{ji}\}} \max_{\mathbf{b}_{j},\{\gamma_{ji}\},\{\gamma_{j$$

(7.7)

Thus the problem is solved by considering a standard L-Q parameter optimization problem, and then optimizing with respect to the selection vector, ϵ_i .

Parameter Optimization

The estimate obtained from (7.5) may be written as

$$\hat{\mathbf{x}}_{\mathbf{j}} = \mathbf{a}_{\mathbf{j}} \boldsymbol{\varepsilon}_{\mathbf{j}}^{\mathsf{T}} \mathbf{y} + \frac{\mathbf{x}}{\mathbf{z}} \mathbf{y}_{\mathbf{j}} \boldsymbol{\beta}_{\mathbf{j}-\mathbf{i}}^{\mathsf{T}} \mathbf{y}$$

$$\mathbf{j} = \mathbb{M} + 1, \mathbb{M} + 2, \dots \qquad (7.8)$$

where β is calculated recursively according to the -lgo-rithm

$$\beta_{\underline{\ell}}^{T} = \alpha_{\underline{\ell}} \varepsilon_{\underline{\ell}}^{T} + \sum_{\substack{i=1\\i=1}}^{M} \gamma_{\underline{\ell}i} \beta_{\underline{\ell}-i}^{T}$$

$$\underline{\ell} = M+1, M+2, \dots$$

(7.9)

For $2 \le j \le M$, the estimate is calculated as

$$\hat{\mathbf{x}}_{j} = \alpha_{j} \varepsilon_{j}^{T} \mathbf{y} + \sum_{i=1}^{j-1} \gamma_{ji} \beta_{j-i}^{y}$$
(7.10)

and the vectors, $\boldsymbol{\beta}$, are calculated as

 $\beta_{\underline{l}}^{T} = \alpha_{\underline{l}} \varepsilon_{\underline{l}}^{T} + \sum_{\substack{i=1\\i=1}}^{\underline{l}-1} \dot{\gamma}_{\underline{l} \underline{i}} \beta_{\underline{l}-\underline{i}}^{T}$

(7.11)

The initial condition for (7.10) is

$$\mathbf{x}_{1} = \mathbf{a}_{1} \boldsymbol{\epsilon}_{1}^{T} \mathbf{y}$$
(7.12)

and for (7.11),

$$\boldsymbol{\beta_1}^{\mathrm{T}} = \boldsymbol{\alpha_1} \boldsymbol{\varepsilon_1}^{\mathrm{T}} \tag{7.13}$$

For notational convenience, we designate

$$\boldsymbol{\beta_{ji}^{T} = \beta_{j-i}^{T}}$$
(7.14)

The performance measure may be written as

$$J_{j} = E\{(x-\alpha_{j}\epsilon_{j}^{T}y - \sum_{i=1}^{M} \gamma_{ji}\beta_{ji}^{T}y)^{2}\}$$

$$j=M+1, M+2, \dots \qquad (7.15)$$

Optimization at each stage requires that

$$\frac{\partial J_{1}}{\partial \alpha_{j}} = 0; \quad \frac{\partial J_{1}}{\partial \gamma_{ji}} = 0; \quad i=1,...,M$$
(7.16)

as necessary conditions for an optimum. This leads to the set of linear equations

$$\mathbf{P}_{\mathbf{j}\mathbf{j}} = \mathbf{d}_{\mathbf{j}}$$
(7.17)

where

$$\mathbf{v}_{\mathbf{j}} \equiv \begin{bmatrix} \mathbf{a}_{\mathbf{j}} \\ \mathbf{v}_{\mathbf{j}\mathbf{l}} \\ \vdots \\ \vdots \\ \mathbf{v}_{\mathbf{j}\mathbf{k}} \end{bmatrix}$$
(7.18)

 P_{j} is an (M+1) x (M+1) symmetric matrix partitioned as

$$\mathbf{P}_{\mathbf{j}} = \begin{bmatrix} \mathbf{P}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} & \mathbf{P}_{\boldsymbol{\varepsilon}\boldsymbol{\beta}} \\ \mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\varepsilon}} & \mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}} \end{bmatrix}_{\mathbf{j}}$$
(7.19)

and d_j is partitioned as

$$\mathbf{d_j} = \begin{bmatrix} \mathbf{d_e} \\ \mathbf{d_\beta} \end{bmatrix}_{\mathbf{j}}$$
(7.20)

The submatrix, $P_{\beta\beta},$ has as its $ik^{\mbox{th}}$ element

$$\mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}_{\mathbf{j}\mathbf{k}}} = \boldsymbol{\beta}_{\mathbf{j}\mathbf{i}}^{\mathbf{T}} \mathbf{P}_{\mathbf{y}\mathbf{y}} \boldsymbol{\beta}_{\mathbf{j}\mathbf{k}}$$
(7.21)

The term $P_{\varepsilon\varepsilon}$ is a scalar defined as

$$\mathbf{P}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}_{\mathbf{j}} \quad \mathbf{y}_{\mathbf{y}} \boldsymbol{\varepsilon}_{\mathbf{j}}$$
(7.22)

while $P_{\varepsilon\beta}$ is an M dimensional row vector whose $k^{\mbox{th}}$ element is

$${}^{\mathbf{P}} \boldsymbol{\varepsilon} \boldsymbol{\beta}_{\mathbf{k}} {}^{\mathbf{T}} {}^{\mathbf{F}} \boldsymbol{\beta}_{\mathbf{j}} {}^{\mathbf{F}} \mathbf{j} \mathbf{k} {}^{\mathbf{Y}} \mathbf{y} {}^{\mathbf{F}} \mathbf{j}$$
(7.23)

It is assumed that $\epsilon_j \neq 0$. The term d_{ϵ} is a scaler

$$\mathbf{t} = \mathbf{c}_{\mathbf{j}}^{\mathbf{T}} \mathbf{p}_{\mathbf{x}}$$
(7.24)

and $d_{\boldsymbol{\beta}}$ is an M dimensioned column vector with $\boldsymbol{k}^{\texttt{th}}$ element

$$\mathbf{d}_{\boldsymbol{\beta}_{\mathbf{k}}} = \boldsymbol{\beta}_{\mathbf{j}\mathbf{k}}^{\mathbf{T}_{\mathbf{p}}} \mathbf{y} \mathbf{x}$$
(7.25)

Equations such as (7.17) occuring in linear estimation theory are generally referred to as normal equations. When P_j is nonsingular, the optimal set of coefficients is obtained by solving for v_j as

$$v_j = P_j^{-1} d_j \qquad (7.26)$$

During the start up period $(2 \le j \le M)$, the preceding equations (7.15 - 7.26) are applicable with M replaced by j-1. When the set of coefficients obtained from (7.26) is used we get

$$\min \{J_j | given \varepsilon_j\} = P_{xx} - d_j^T P_j^{-1} d_j$$

$$a_j, \{\gamma_j\}$$
(7.27)

where $P_{XX} = E\{x^2\}$.

According to (7.7), we want to minimize (7.27) with respect to ϵ_j where the elements of ϵ_j can be only 1 or 0. Clearly this is equivalent to maximizing the expression.

 $\mathbf{J}_{\mathbf{j}}^{*} \equiv \mathbf{d}_{\mathbf{j}}^{\mathsf{T}} \mathbf{P}_{\mathbf{j}}^{-1} \mathbf{d}_{\mathbf{j}}$ (7.28)

Optimizing the selection Vector

It will be shown that the maximization of J^* is related to a classic problem [12, 13, 55] generally referred to as the quadratic assignment problem. Maximization of J^* is equivalent to maximizing a ratio of quadratic forms in ϵ_j . Although this is an easy problem, with an elegant solution when ϵ_j may be freely chosen [56], it is a difficult problem when each element of ϵ_j is either a zero or a one. Therefore in most applications one would probably have to restrict the selection of ϵ_j to a search for the best out of a reasonable number of candidate vectors. The candidate vectors would be hueristically selected, with some guiding principles which will be discussed.

It is known [57] that $P_{\overline{J}}$ can be written as



(7.29)

where

 $C = \left[P_{\beta\beta} - P_{\beta\epsilon} P_{\epsilon\beta} P_{\epsilon\epsilon}\right]$ (7.30)

$$\mathbf{B} = -\mathbf{P}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1} \mathbf{P}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}^{-1} \mathbf{C}$$
(7.31)

$$\mathbf{A} = \mathbf{P} - \mathbf{P} -$$

The performance measure may be written as

$$J^* = d_{\varepsilon}^{T} A d_{\varepsilon}^{+2} d_{\varepsilon}^{T} B d_{\beta}^{+} d_{\beta}^{T} C d_{\beta}$$
(7.33)

where we have left off the subscript j for convenience. Using the matrix inversion lemma [58], we see that (7.30) can be expressed as

$$\mathbf{c} = \mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} - \left[\mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} \mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\varepsilon}}^{\mathbf{p}} \mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1}\right] / \left[\mathbf{P}_{\boldsymbol{\varepsilon}\boldsymbol{\beta}}^{\mathbf{p}} \mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{-1} \mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\varepsilon}}^{-1} \mathbf{P}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}}\right]$$
(7.34)

- We may write
- $\mathbf{d}_{\boldsymbol{\beta}} = \bar{\boldsymbol{\beta}}^{\mathrm{T}} \mathbf{P}_{\mathbf{y}\mathbf{x}}$ (7.35)

$$P_{\beta \varepsilon} = \overline{\beta}^{T} P_{yx} \varepsilon$$
(7.36)

if $\tilde{\beta}_j$ is defined according to

$$\bar{\boldsymbol{\beta}}_{j}^{T} \equiv \begin{bmatrix} \frac{\boldsymbol{\beta}_{j1}}{\cdot} \\ \vdots \\ \vdots \\ \vdots \\ \boldsymbol{\beta}_{jM}} \end{bmatrix}$$
(7.37)

Substitution from (7.22) and (7.34 - 7.37) in the last term of (7.33) gives

$$\mathbf{d_{\beta}^{T}Cd_{\beta}} = \frac{\boldsymbol{\varepsilon}^{T}\boldsymbol{Q}_{22}\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}^{T}R\boldsymbol{\varepsilon}}$$
(7.38)

where R is defined as

$$\mathbf{R} = \mathbf{P}_{\mathbf{y}\mathbf{y}} \mathbf{\beta} \mathbf{P}_{\mathbf{\beta}\mathbf{\beta}} - \mathbf{1}_{\mathbf{\beta}} \mathbf{T}_{\mathbf{y}\mathbf{y}} - \mathbf{P}_{\mathbf{y}\mathbf{y}}$$
(7.39)

and Q_{22} as

$$Q_{22} \equiv (\lambda^{T} P_{yx}) R - P_{yy} \lambda \lambda^{T} P_{yy}$$
(7.40)

where

$$\lambda \equiv \beta P_{\beta\beta} {}^{-1} {}^{\beta} {}^{T} P_{yx}$$
(7.41)

Similarly, the second term in (7.33) is of the form

$$2d_{\varepsilon}^{T}Bd_{\beta} = \frac{2\varepsilon^{T}Q_{12}\varepsilon}{\varepsilon^{T}R\varepsilon}$$
(7.42)

where

$$Q_{12} \equiv P_{yx} \lambda^{T} P_{yy}$$
(7.45)

The first term is of the form

$$d_{\varepsilon}^{T}Ad_{\varepsilon} = \frac{\varepsilon^{T}Q_{11}\varepsilon}{\varepsilon^{T}R\varepsilon}$$
(7.44)

with Q_{11} defined as

$$\mathbf{Q}_{11} \equiv -\mathbf{P}_{\mathbf{y}\mathbf{x}} \mathbf{P}_{\mathbf{y}\mathbf{x}}$$
(7.45)

Thus the expression for J^* may be written as

$$\mathbf{J}^{\star} = \frac{\boldsymbol{\varepsilon}^{\mathrm{T}} [\boldsymbol{\varphi}_{11} + 2\boldsymbol{\varphi}_{12} + \boldsymbol{\varphi}_{22}] \boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}^{\mathrm{T}} \boldsymbol{\kappa} \boldsymbol{\varepsilon}}$$
(7.46)

This is more conveniently written as

$$\mathbf{J}^{\star} = \frac{\boldsymbol{\varepsilon}^{\mathsf{T}} \left[\mathbf{Q}_{11}^{\dagger} + \mathbf{Q}_{12}^{\dagger} + \mathbf{Q}_{22}^{\dagger} \right] \boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}^{\mathsf{T}} \mathbf{R} \boldsymbol{\varepsilon}}$$
(7.47)

After substituting in (7.47) for the terms Q_{ij} , we get



where

$$\bar{\mathbf{M}} \sim \beta \mathbf{P}_{\beta\beta}^{\left(-1\right)\beta^{\mathrm{T}}}$$
(7.49)

The first term is a scalar, unaffected by the choice of ϵ . The second term is a ratio of two quadratic forms in ϵ . Thus (7.48) may be written as

 $J^* = P \frac{T_{\overline{MP}}}{y_{\overline{X}}} + J^{**}$ (7.50)

where

$$J^{\star\star} = \frac{\varepsilon^{T} G^{T} P P^{T} G \varepsilon}{\varepsilon^{T} G P \varepsilon}, \quad \varepsilon \neq 0$$
(7.51)

and

$$G^{T} = [I - P_{yy}M].$$
 (7.52)

The goal of selecting ϵ_j is to maximize the ratio expressed in (7.51). The symmetric weighting matrices in the quadratic forms in the numerator and denominator are positive semi-definite, and positive definite respectively. We remark that the maximization of (7.51) must be done at each stage, and so it is a significant task. (We have left off the index J indicating the jth stage.)

Although problems related to the performance indicator (7.51) have been treated in the literature [53-55], the difficulty of the quadratic assignment problem should not be underestimated. As an example, for even a modest amount of data, say y is of length 26, it would take over a minute to calculate (7.51) for each possibility, assuming the calculation could be done in a microsecond. If the length of the data vector were 100, the time required to check all possibilities would be measured in centuries.

Instead of checking all possibilities, we can restrict ourselves to the most likely candidates, and we can eliminate those selection vectors which have previously been used, or which can be formed by summing those selection vectors already used. Knowledge of the truly optimal linear estimate,

 $\hat{\mathbf{x}}_{\mathbf{o}} = \mathbf{P}_{\mathbf{x}\mathbf{y}}\mathbf{P}_{\mathbf{y}\mathbf{y}}^{-1}\mathbf{y} = \sum_{i=1}^{N} \mathbf{a}_{i}\mathbf{y}_{i}$

(7.53)

can be used to find the likely candidates. As a simple example, if

$$\hat{x}_{0} = 1.1y_{1} + .9y_{2} + 1.0y_{3} + .11y_{4} + .09y_{5}$$

(7.54)

then one would probably guess that $\epsilon_1^{\rm T} = [11100], \epsilon_2^{\rm T} = [00011]$ would be excellent choices for selection vectors. Furthermore one would expect performance to be good with only two multiplications allowed. We believe, however, that even with problems as simple as this, intuition is subject to error when choosing selection vectors. Naturally when y is a very long vector some computational assistance will be required to select the candidate vectors. It should be remembered that this is a part of the design, and that the purpose of going to this design effort is to restrict ourseleves to an algorithm with few multiplications required so that the filtering algorithm generates estimates quickly. We are willing to spend considerable extra design effort to get a fast processor.

The Design Algorithm

To start the algorithm, we must select α_1 , and ϵ_1 to minimize J_1 where

 $J_1 = E\{(x-x_1)^2\}$

(7.55)

This gives for α_1 ,

 $a_1 = \epsilon_1^{T_P} / \epsilon_1^{T_P} / y_y \epsilon_1$

(7.56)

Since when this α_1 is used,

$$J_{1} = P_{xx} - \varepsilon_{1}^{T_{p}} P_{yx}^{T} \varepsilon_{1} / \varepsilon_{1}^{T_{p}} P_{yy} \varepsilon_{1}$$
(7.57)

we select ϵ_1 to maximize the quality



(7.58)

Thus α_1 and ϵ_1 are obtained, and β_1 is found as

 $\beta_1^T = \alpha_1 \varepsilon_1^T$

(7.59)

The next step is to form the term

 $\bar{M}_{2} = \beta_{1}\beta_{1}^{T}/(\beta_{1}^{T}P_{yy}\beta_{1})$ (7.60)

and evaluate

 $G_2^T \equiv [I - P_{yy} \overline{M}_2]$

(7.61)

The terms $G_2^T P_{yy}$ and $G_2^T P_{yx}$ are calculated. The expression indicated by (7.51)

$$J_{2}^{\star\star} \approx \frac{\varepsilon_{2}^{T} G_{2}^{T} P_{yx} P_{yx}^{T} G_{2} \varepsilon_{2}}{\varepsilon_{2}^{T} G_{2}^{T} P_{yy} \varepsilon_{2}}$$

(7.62)

is maximized with respect to ϵ_2 .

The coefficients, α_2 and Y_{21} are found according to (7.26), i.e.

$$\begin{bmatrix} \alpha_{2} \\ \gamma_{21} \end{bmatrix} \begin{bmatrix} \frac{P_{\varepsilon_{2}\varepsilon_{2}}}{P_{\varepsilon_{2}}\beta_{1}} \\ \frac{P_{\beta_{1}\varepsilon_{2}}}{P_{\beta_{1}\beta_{1}}} \end{bmatrix} \begin{bmatrix} d_{\varepsilon_{2}} \\ d_{\beta_{1}} \end{bmatrix}$$
(7.63)

where

$$P_{\epsilon_{2}\epsilon_{2}} = \epsilon_{2}^{T_{p}} yy^{\epsilon_{2}} P_{\beta_{1}\beta_{1}} = \beta_{1}^{T_{p}} yy^{\beta_{1}}$$

$$P_{\epsilon_{2}\beta_{1}} = \epsilon_{2}^{T_{p}} yy^{\beta_{1}} d_{\epsilon_{2}} = \epsilon_{2}^{T_{p}} yx d_{\beta_{1}} = \beta_{1}^{T_{p}} yx$$

$$(7.64)$$

Then β_2 is found as

$$\beta_2^{T} = \alpha_2 \varepsilon_2^{T} + \gamma_{21} \beta_1^{T}$$
(7.65)

Using the notation indicated by (7.14)

$$\beta_{31}^{T} = \beta_{3-1}^{T} = \beta_{2}^{T}$$

$$\beta_{32}^{T} = \beta_{3-2}^{T} = \beta_{1}^{T}$$
(7.66)

and according to (7.14)

$$\overline{\boldsymbol{\beta}}_{3}^{T} = \begin{bmatrix} \boldsymbol{\beta}_{31}^{T} \\ \boldsymbol{\beta}_{32}^{T} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_{2}^{T} \\ \boldsymbol{\beta}_{1}^{T} \end{bmatrix}$$
(7.67)

The matrix M_3 is formed as

$$\overline{\mathbf{M}}_{3} = \overline{\mathbf{\beta}}_{3} [\overline{\mathbf{\beta}}_{3}^{T} \mathbf{P}_{yy} \overline{\mathbf{\beta}}_{3}]^{-1} \mathbf{\beta}_{3}^{T}$$
(7.68)

Equations (7.61) and (7.62) are applicable with the subscript "2" replaced by "3" and ϵ_3 is selected accordingly.

The new coefficients are obtained by solving (7.26)

$$\begin{bmatrix} \alpha_{3} \\ \gamma_{3} \end{bmatrix} - \begin{bmatrix} \frac{P}{\epsilon_{3}\epsilon_{3}} & \frac{P}{\epsilon_{3}\beta_{3}} \\ \frac{P}{\beta_{3}\epsilon_{3}} & \frac{P}{\beta_{3}\beta_{3}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{d}{\epsilon_{3}} \\ \frac{d}{\beta_{3}} \end{bmatrix}$$
where $\gamma_{3} - \begin{bmatrix} \gamma_{31} \\ \gamma_{32} \end{bmatrix}$ and (7.69)

$$P_{\varepsilon_{3}\varepsilon_{3}} = \varepsilon_{3}^{T}P_{yy}\varepsilon_{3} \qquad P_{\beta_{3}\beta_{3}} = \overline{\beta}_{3}^{T}P_{yy}\overline{\beta}_{3} \qquad (7.70)$$

$$P_{\varepsilon_{3}\beta_{3}} = \varepsilon_{3}^{T}P_{yy}\overline{\beta}_{3} \qquad d_{\varepsilon_{3}} = \varepsilon_{3}^{T}P_{yx} \qquad d_{\beta_{3}} = \overline{\beta}_{3}^{T}P_{yx} \qquad (7.71)$$

These steps can be generalized. Let us assume that we are in the start up period (j<M+1), and have $\beta_j,\;\alpha_j,\;$ and γ_j where

$$\overline{B}_{j} = \begin{bmatrix} B_{j-1} \\ B_{1} \end{bmatrix} \qquad r_{j} = \begin{bmatrix} Y_{j1} \\ Y_{j}, j-1 \end{bmatrix}$$
(7.72)

We then solve for $\beta_{\rm j}$ according to (7.11)

$$\beta_{j}^{T} = \alpha_{j} \epsilon_{j}^{T} + \frac{j-1}{\sum_{i=1}^{T} \gamma_{ji} \beta_{j-1}^{T}}$$
(7.73)
and form
$$\overline{\beta}_{j+1}^{T} = \left[\frac{\beta_{j}}{\overline{\beta}_{j}^{T}} \right]$$
(7.74)

(7.74)

Next, the matrix M_{j+1} is calculated:

$$\overline{\mathbf{M}}_{j+1} = \overline{\mathbf{B}}_{j+1} [\overline{\mathbf{B}}_{j+1}^{\dagger} \mathbf{P}_{yy} \overline{\mathbf{B}}_{j+1}]^{-1} \overline{\mathbf{B}}_{j+1}$$
(7.75)
and \mathbf{G}_{j+1} is evaluated as

The vector ϵ_{j+1} which maximizes

$$\mathbf{J}_{j+1}^{**} = \varepsilon_{j+1}^{T} \mathbf{G}_{j+1}^{T} \mathbf{y}_{\mathbf{y}}^{\mathbf{y}} \mathbf{y}_{\mathbf{x}}^{\mathbf{y}} \mathbf{y}_{\mathbf{y}}^{\mathbf{f}} \mathbf{j} + 1 \mathbf{G}_{j+1}^{\mathbf{f}} \mathbf{g}_{j+1}^{\mathbf{f}} \mathbf{y}_{\mathbf{y}}^{\mathbf{f}} \mathbf{j} + 1$$
(7.77)

is selected.

كم فرد و و ما

We then solve for the coefficients

$$\begin{bmatrix} \frac{\alpha}{j+1} \\ \gamma_{j+1} \end{bmatrix} = \begin{bmatrix} \frac{P_{\varepsilon_{j+1}\varepsilon_{j+1}}}{P_{\beta_{j+1}\varepsilon_{j+1}}} & \frac{P_{\varepsilon_{j+1}\beta_{j+1}}}{P_{\beta_{j+1}\beta_{j+1}}} \\ \frac{P_{\varepsilon_{j+1}\beta_{j+1}}}{P_{\beta_{j+1}\beta_{j+1}}} & \frac{-1 \begin{bmatrix} d_{\varepsilon_{j+1}} \\ d_{\beta_{j+1}} \\ d_{\beta_{j+1}} \end{bmatrix} }{(7.78)}$$

 ${}^{P} {}^{e} {}^{j+1} {}^{e} {}^{j+1} {}^{T} {}^{P} {}^{yy} {}^{e} {}^{j+1} 1$ ${}^{P} {}^{e} {}^{j+1} {}^{\beta} {}^{j+1} {}^{r} {}^{e} {}^{j+1} {}^{T} {}^{p} {}^{j} {}^{j} {}^{j+1}$ ${}^{P} {}^{e} {}^{j+1} {}^{\beta} {}^{j+1} {}^{e} {}^{j+1} {}^{T} {}^{p} {}^{yy} {}^{\beta} {}^{j+1}$ ${}^{P} {}^{e} {}^{j+1} {}^{\beta} {}^{j+1} {}^{e} {}^{j+1} {}^{T} {}^{p} {}^{yx} {}^{d} {}^{\beta} {}^{j+1} {}^{e} {}^{j+1} {}^{P} {}^{yx} {}^{d} {}^{\beta} {}^{j+1} {}^{e} {}^{j+1} {}^{P} {}^{yx} (7.79)$ $The algorithm is then repeated as given until it is desired to limit the number of multiplications to M+1 as we have indicated in this paper. Assuming that we have <math>\beta_{j}, o_{j}, o_{j}$

$$\mathbf{\overline{\beta}}_{j}^{T} = \begin{bmatrix} \boldsymbol{\beta}_{j-1}^{T} \\ \vdots \\ \boldsymbol{\beta}_{j-M}^{T} \end{bmatrix} \qquad \mathbf{\gamma}_{j} = \begin{bmatrix} \mathbf{\gamma}_{j,1} \\ \vdots \\ \mathbf{\gamma}_{j,M} \end{bmatrix}$$

where

(7.80)

we solve for β_j according to (7.9). Then (7.75) through (7.78) are applied and (7.77) is maximized. Equation (7.78) is used to select the new coefficients. The algorithm is thus established for all j.

The reader may be troubled at this point because it appears that an ever larger matrix needs to be inverted at each stage, during the start up period and that a large matrix (if M is large) needs to be inverted thereafter. We shall show that there are computationally efficient recursive means of carrying out the required matrix inversion.

Efficient Matrix Inversion

The Start Up Period:

It is necessary at each stage to calculate two matrix inverses $P_{BB_i}^{-1}$ and P_j^{-1} . In this section we indicate how

to calculate these terms at stage j+1 given these terms are available at stage j. Thus a recursive procedure is established. During the start up period we note that

$$\mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}_{j+1}} = \bar{\boldsymbol{\beta}}_{j+1}^{T} \mathbf{P}_{yy} \bar{\boldsymbol{\beta}}_{j+1} = \begin{bmatrix} \boldsymbol{\beta}_{j}^{T} \mathbf{P}_{yy} \boldsymbol{\beta}_{j} & \boldsymbol{\beta}_{j}^{T} \mathbf{P}_{yy} \bar{\boldsymbol{\beta}}_{j} \\ \bar{\boldsymbol{\beta}}_{j}^{T} \mathbf{P}_{yy} \boldsymbol{\beta}_{j} & \boldsymbol{P}_{\boldsymbol{\beta}\boldsymbol{\beta}_{j}} \end{bmatrix}$$
(7.81)

The Matrix inverse [57], can be written as

$$\begin{bmatrix} \mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}} \\ \mathbf{j+1} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{\alpha}_{11} & \boldsymbol{\alpha}_{12} \\ \hline \boldsymbol{\alpha}_{12} & \boldsymbol{\alpha}_{22} \end{bmatrix}$$
(7.82)

where

$$\mathbf{a}_{22} = \left[\mathbf{P}_{\beta\betaj} - \frac{\mathbf{\tilde{\beta}}_{j}^{\mathbf{T}} \mathbf{P}_{yy} \mathbf{\beta}_{j} \mathbf{\beta}_{j}^{\mathbf{T}} \mathbf{P}_{yy} \mathbf{\bar{\beta}}_{j}}{\mathbf{\beta}_{j}^{\mathbf{T}} \mathbf{P}_{yy} \mathbf{\beta}_{j}}\right]^{-1}, \qquad (7.83)$$

$$\alpha_{12} = \frac{-\beta_{j}^{T} P_{yy} \overline{\beta}_{j} \alpha_{22}}{\beta_{j}^{T} P_{yy} \beta_{j}}, \qquad (7.84)$$

and 11 is

$$a_{11} = \frac{1}{\beta_{j}^{T} p_{yy} \beta_{j}} + \frac{\beta_{j}^{T} p_{yy} \overline{\beta}_{j} \alpha_{22} \overline{\beta}_{j}^{T} p_{yy} \beta_{j}}{(\beta_{j}^{T} p_{yy} \beta_{j})^{2}}$$
(7.85)

Using the matrix inversion lemma [58], we see that (7.83) may be written as

$$a_{22} = P_{\beta\beta}^{-1} - P_{\beta\beta}^{-1} r^{T} [rP_{\beta\beta}^{-1} r^{T} - \beta_{j}^{T} yy^{\beta} j]^{-1} rP_{\beta\betaj}$$
(7.86)

where

$$\mathbf{r} = \mathbf{\beta}_{\mathbf{j}}^{\mathbf{T}} \mathbf{p}_{\mathbf{j}} \mathbf{\bar{\beta}}_{\mathbf{j}}$$
(7.87)

Equation (7.86) requires only the inversion of a scalar, and since $P_{\beta\beta}^{-1}$ j is known, we see that we have an efficient way of finding $P_{\beta\beta j+1}^{-1}$ given $P_{\beta\beta j}^{-1}$. Once $P_{\beta\beta j+1}^{-1}$ is known it is not a difficult matter to calculate P_{j+1}^{-1} using equations (7.29 - 7.32) and (7.34). In this case also, only a scalar needs to be inverted. During the start up period then, matrix inversion does not present a problem. This is also the case during the remaining period with $j \ge M+1$.

After Start Up:

In the previous section a method was developed for inverting a matrix of increasing dimension, using the result from the previous stage. In this section we are inverting a different matrix of the same dimension at each stage. Because the matrix at one stage is very closely related to the matrix at the previous stage, it is possible to obtain the new inverse from the old, with surprisingly small amount of calculation.

It is known that the lower right hand portion of ${}^{P}_{\beta\beta j+1}$ is the same as the upper left hand portion of ${}^{P}_{\beta\beta j}$. Specifically,

$$\mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}_{j+1}} = \begin{bmatrix} \frac{\boldsymbol{\beta}_{j}^{T} \mathbf{P}_{yy} \boldsymbol{\beta}_{j}}{\boldsymbol{\beta}_{j}^{*} \mathbf{P}_{yy} \boldsymbol{\beta}_{j}} & \boldsymbol{\beta}_{j}^{T} \mathbf{P}_{yy} \boldsymbol{\beta}_{j}^{*} \\ \boldsymbol{\beta}_{j}^{*} \mathbf{P}_{yy} \boldsymbol{\beta}_{j}^{T} & \boldsymbol{\beta}_{j}^{*T} \mathbf{P}_{yy} \boldsymbol{\beta}_{j}^{*} \end{bmatrix} \equiv \begin{bmatrix} \frac{d_{11}}{d_{12}} & \frac{d_{12}}{d_{12}} \\ \frac{d_{12}}{d_{12}} & \mathbf{C}_{11} \end{bmatrix}$$
(7.88)

and

$$P_{\beta\beta} = \begin{bmatrix} \beta_{j}^{*T} P_{yy} \beta_{j}^{*} & \beta_{j}^{*T} P_{j} \beta_{j-M} \\ \beta_{j}^{T} P_{yy} \beta_{j}^{*} & \beta_{j-M}^{T} P_{yy} \beta_{j-M} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{12}^{T} & C_{22} \end{bmatrix}$$
(7.89)

where

$$\boldsymbol{\beta_{j}^{*T}} = \begin{bmatrix} \boldsymbol{\beta_{j-1}^{T}} \\ \vdots \\ \boldsymbol{\beta_{j-M+1}^{T}} \end{bmatrix}$$
(7.90)

Suppose that $P_{\beta\beta j}^{-1}$ is known and particled as

$$P_{\beta\beta}^{-1} = \begin{bmatrix} \frac{b_{11} + b_{12}}{T} \\ \frac{b_{12} + b_{22}}{D} \end{bmatrix}$$
(7.91)

where b_{11} is an (M-1) x (M-1) matrix and b_{22} is a scalar. We know from [57], that this matrix inverse can also be written as

$$\mathbf{P_{\beta\betaj}^{-1}} = \begin{bmatrix} \frac{c_{11}^{-1} + c_{11}^{-1}c_{12}Dc_{12}^{-1}c_{11}^{-1} & -c_{11}^{-1}c_{12}^{-1}}{-Dc_{12}^{-1}c_{11}^{-1} & D} \end{bmatrix}$$

(7.92)

where

$$\mathbf{b} = [\mathbf{c}_{22} - \mathbf{c}_{12}^{\mathsf{T}} \mathbf{c}_{11}^{-1} \mathbf{c}_{12}]^{-1}$$
(7.93)

Comparing (7.91) and (7.92) it is clear that we can solve for C_{11}^{-1} in terms of known quantities.

$$c_{11}^{-1} - b_{11} [I - c_{12} b_{12}^{T}]^{-1}$$
 (7.94)

From the matrix inversion lemma, the above may be evaluated as

$$c_{11}^{-1} = b_{11} \{ I + c_{12} [1 - b_{12}^{T} c_{12}]^{-1} b_{12}^{T} \}$$
 (7.95)

Therefore when $P_{\beta\beta j}^{-1}$ is known we may evaluate C_{11}^{-1} with only a scalar inversion. As will be seen, knowledge of C_{11}^{-1} will allow us to easily evaluate $P_{\beta\beta j+1}^{-1}$. As in the obtaining of (7.92), we can derive an expression for $P_{\beta\beta j+1}^{-1}$.

$$\mathbf{P}_{\boldsymbol{\beta}\boldsymbol{\beta}_{j+1}}^{-1} = \begin{bmatrix} \frac{\cdot 1}{d_{11}} + \frac{1}{d_{12}^2} d_{12}^{\mathbf{F}d_{12}} & \frac{-d_{12}\mathbf{F}}{d_{11}} \\ \frac{-\mathbf{F}d_{12}}{d_{11}} & \mathbf{F} \\ \frac{-\mathbf{F}d_{12}}{d_{11}} & \mathbf{F} \end{bmatrix}$$
(7.96)

where F is defined as,

$$\mathbf{r} = \left[\mathbf{c_{11}} - \frac{\mathbf{d_{12}}^{\mathbf{1}} \mathbf{d_{12}}}{\mathbf{d_{11}}}\right]^{-1}$$
(7.97)

but may be more conveniently obtained, again using the matrix inversion lemma:

$$\mathbf{F} = \mathbf{C}_{11}^{-1} - \mathbf{C}_{11}^{-1} \mathbf{d}_{12}^{T} [\mathbf{d}_{12}^{T} \mathbf{C}_{11}^{-1} \mathbf{d}_{12}^{T} - \mathbf{d}_{11}^{T}]^{-1} \mathbf{d}_{12}^{T} \mathbf{C}_{11}^{-1}$$
(7.98)

Only a scalar need be inverted.

Therefore we have established a convenient mechanism of generating $P_{\beta\beta j+1}$. First obtain C_{11}^{-1} using (7.95). Then obtain F using (7.98) and substitute the results in (7.96). After obtaining $P_{\beta\beta j+1}^{-1}$, it is easy to calculate P_{j+1}^{-1} using equations (7.29 - 7.32) and (7.34). For both start up and afterwords, we have thus established a methodology for recursively calculating matrix inverses. This, in effect, frees up more time for the quadratic assignment problem which is now clearly seen to be the only real difficulty in the design procedure. We will not consider matrix inversion during the transition between start up and fixed length operation here.

An Example

In this section, we shall illustrate the design of the filtering algorithm with an academic example. It is assumed that 4 measurements are available

 $y_k = x + v_k$, k = 1, ..., 4 (7.99) and that x and v_k are uncorrelated. The noise, v_k , is zero mean with known statistics

 $E\{v_{1}^{2}\}=1 \qquad E\{v_{3}^{2}\}=11 \qquad E\{v_{1}v_{k}\}=0; i\neq k$ $E\{v_{2}^{2}\}=2 \qquad E\{v_{4}^{2}\}=12 \qquad (7.100)$

and the desired signal, x, is known to have zero mean and variance of unity. The optimal linear estimate is easily found to be

$$\hat{x}_{0} = P_{xy} P_{yy}^{-1} y = .374y_{1} + .187y_{2} + .034y_{3} + .031y_{4}$$
, (7.101)
and the minimum mean square error is

$$J_{o} = E\{(x-x_{o})^{2}\} = .374$$
(7.102)

In applying our algorithm we first want the best estimate which involves a single multiplication

 $\hat{\mathbf{x}}_{1} = \boldsymbol{\alpha}_{1} \boldsymbol{\varepsilon}_{1}^{T} \mathbf{y}$ (7.103)

We evaluate the choice of ϵ_1 which maximizes (7.58), limiting ourselves to three candidates which appear to be possibilities,

 $\boldsymbol{\varepsilon}_{1} = \begin{bmatrix} 1000 \\ 1100 \\ 1111 \end{bmatrix}$ (7.104)

It turns out that the middle choice is the one which maximizes (7.58), so evaluating α_1 gives for the first estimate

$$x_1 = .286(y_1 + y_2)$$

85

(7.105)

and results in a performance,

 $J_1 = E\{(x-x_1)^2\} = .429$ (7.106)

For the next stage one could argue that the best thing to do is to make a distinction between the two terms in (7.101)with the larger coefficients, or to give some weighting to the smaller terms. Thus we select ϵ_2 from among the candidates

 $2 = \begin{cases} [1000] \\ [0100] \\ [0011] \end{cases}$ (7.107)

As it turns out, the last choice is the best, resulting in the largest value of (7.62). Thus the best estimate at stage 2 is

 $\hat{x}_2 = .035(y_3 + y_4) + .930\hat{x}_1$ (7.108)

Substitution from (7.105) gives

$$x_2 = .035(y_3 + y_4) + .266(y_1 + y_2)$$
(7.109)

and this seems reasonable in view of (7.105). The performance measure is

$$J_2 = E\{(x-x_2)^2\} = .399$$
 (7.110)

As an observation, the other two candidates would result in equivalent performance with each other, a fact which could easily be reasoned to. The reader is encouraged to work through this example to notice that there is not enough difference in the choice of ϵ_2 to make one confident that the "intuited" solution is always going to be the correct solution vector even in an academic problem. This points out the need for a mathematical approach such as developed here.

Discussion

We have presented a method for designing a recursive linear filter with a fixed number of multiplications allowed for each interaction. This is ideal for the situation when estimates are going to be required rapidly, and there will be a large amount of data available all at once. The problem has had two aspects, parameter optimization, and selection vector optimization. The parameter optimization has been shown to have a solution which may be obtained in a computationally efficient manner. The selection vector optimization problem is difficult. We have related it to a classic problem in operations research, referred to as the quadratic assignment problem. The difficulty may be limited by limiting the candidate vectors to a few reasonable choices.

In this Chapter we have considered only stagewise optimization rather than optimization over the entire sequence of estimates generated. Since even the simpler problem considered here has an open research area imbedded in it (the quadratic assignment problem), it may be premature to consider solving the more general dynamic optimization





5

,

ちんてう アス

MICROCOPY RESOLUTION TEST CHART NATIONAL BUREAU OF STANDARDS-1963+A problem. The solution of such a problem is ultimately desirable and could be pursued within the context of modern control theory. It is also desirable to solve a problem similar to that posed here, but in a more general vector format. The formal mathematics of parameter optimization is not a difficulty in pursuing the vector problem. Indeed it is again the integer programming that presents a computational limitation. While we acknowledge that one may be reasonable speed hours of computational effort in the design of a rapid algorithm, it is clearly not reasonable to spend centuries at such design.

It is the belief of the authors that the design presented herein, coupled with some heuristic appraoches toward limiting the number of selection vector candidates, represents a reasonable approach to the design of a rapid recursive algorithm for a certain class of important estimation problems.

CHAPTER VIII

CONCLUSIONS

The goal of this research was to investigate the possibility of combining an adaptive filter with a linear predictive coding algorithm to form a robust and efficient system for narrowband encoding of speech signals corrupted by noise. Various prefiltering techniques for improving linear coding systems were evaluated. Primary techniques of interest were the pitch tracking adaptive filter and the spectral subtraction filter. The pitch tracking adaptive filter proved successful in suppressing white noise in voiced speech sounds but did not work well when the noise was narrowband such as a single sine wave. It was found that interaction between the pitch period and the narrowband noise produces a bias error in the adaptation of the filter.

Failure of the pitch tracking adaptive filters to suppress narrowband noise prompted the investigation of several other prefiltering methods. The most successful of the filters evaluated was the spectral subtraction technique. Two modifications to the original method proved very useful for improving noisy speech. First a dual time constant noise spectrum estimate improved white noise suppression and

secondly a spectral notch feature greatly improved narrowband noise quieting. Also, very successful was a new filter method based on adaptive filtering.

Work with the spectral subtraction filter method suggested a more general approach to speech filtering. Short time Fourier analysis of speech produces a two dimensional representation of a speech signal which may be processed much like image data. The investigation found that some types of speech and noise signals may be separated using two dimensional filtering on the short time Fourier transform representation of a noisy speech signal.

The performance of the pitch tracking adaptive filter depends on the quality of the pitch period estimate used to set the input delay. Early attempts to implement the filter were frustrated by the degradation of currently available pitch algorithms in the presence of noise. It was found that the adaptive filter itself could be modified to provide a robust pitch estimate. This technique was used extensively throughout the research to provide pitch estimates for various processing algorithms.

To complement the filtering algorithms, fast algorithms have been derived for efficient solution of the linear estimation problem. These include a fast algorithm for the solution of the general discrete time linear estimation problem and a new recursive linear estimator suitable for rapid estimation of a signal in noise. The approach is related to the classical integer programming problem.

Suggestions for Further Study

The results presented here suggest several avenues one could take in the areas discussed. These are presented below.

- Considering that the adaptive algorithms are computationally complex, fast algorithm development in this area is important.
- The results presented on the enhancement of speech signals by two dimensional signal processing are still at a preliminary stage. This approach may be considered as a special case of the problem of estimating time varying process parameters in the presence of stationary noise. This area is wide open as all the image processing techniques are available for noise suppression, coding, data compression, etc.
 - Most of the estimation algorithms are based upon the least squares analysis (l_2 analysis). It is worthwhile to investigate the possibilities of using l_1 (or in general l_p , $1 \le p \le 2$) analysis for signal estimation when the signal is burried noise. Again, this area has wide implication in speech processing.
- In the results presented on the design of fast recursive estimators, we have considered only stage wise optimization rather than optimization over the entire sequence of estimates. The more general dynamic optimization problem may be a difficult problem to tackle.

However, it is desirable to solve this problem and could be pursued within the context of modern control theory.

CITED REFERENCES

- 1. Teacher, C. F., and Watkins, H., "ANDVT Microphone and Audio System Study", Ketron Final Report, 1978.
- Teacher, C. F., "Performance of LPC Vocoders in a Noisy Environment," Proc. IEEE Int. Conf. on Acoustrics, Speech and Signal Processing, p. 216-219, 1979.
- Sambur, M. R., "Adaptive Noise Cancelling for Speech Signals," <u>IEEE Trans. Acoustics, Speech and</u> Signal Processing, ASSP-26, pp. 419-429, 1978.
- 4. Widrow, B., et al, "Adaptive Noise Cancelling: Principles and Applications," Proc. IEEE, Vol. 65 No. 12, p. 1692, 1975.
- 5. Soldan, D. L., and Ahmed, N., "On a Sequential Regression Predictor," <u>Electronics Letters</u>, Vol. 14, pp. 118-119, 1978.
- 6. Rabiner, L. R., and Schafer, R. W., "<u>Digital</u> <u>Processing of Speech Signals</u>," Prentice-Hall, 1978.
- 7. Seneft, S., "Real-Time Harmonic Pitch Detector", <u>IEEE</u> <u>Trans. on Acoustics, Speech, and Signal</u> <u>Processing</u>, Vol. ASSP-26, No. 4, pp. 358-365, 1978.
- 8. Reed, F. A., Feintuch, P. L. and Bershad, N. J., "Time Delay Estimation Using the LMS Adaptive Filter-Static Behavior," <u>IEEE Trans. on Acoustics,</u> <u>Speech and Signal Processing</u>, Vol. ASSP-29, No. 3, pp. 561-571, 1981.
- 9. Feintuch, P. L., Bershad N. J., and Reed, F. A., "Time Delay Estimation Using the LMS Adaptive Filter-Dynamic Behavior," <u>IEEE Trans. on Acoustics</u>, <u>Speech and Signal Processing</u>, Vol. ASSP-29, No. <u>3</u>, pp. 571-577, 1981.
- Boll, S. F., "Adaptive Noise Cancelling in Speech Using the Short Time Transform," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 692-695, 1980.

 Woods, J. W. and Ingle, V. K., Two-Dimensional Processing of Spectrogram Data," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 39-42, 1978.

- 12. Lawler, E., "The Quadratic Anignment Problem: A Brief Review", in Combinatorial Programming: Methods and Applications, B. Roy edition, D. Reidel Publishing Company, Dordrecht, Holland, p. 351, 1975.
- 13. Yarlagadda, R., "Unit Integer Quadratic Binary Programming," Journal of Mathematical Analysis and Applications, Vol. 94, pp. 246-267, 1983.
- 14. Atal B. S. and Hanauer, S. L., "Speech Analysis and Synthesis by Linear Prediction of the Speech Wave," J. Acoust. Soc. Am., Vol. 50, pp. 637-655, 1971.
- 15. Kemp D. P. and Russell, J. W., "Evaluation of Two Narrowband Speech Algorithms," Proc. IEEE Int. Conf. on Acoustics Speech and Signal Processing, p. 990, 1979.
- 16. Curtis R. A. and Niederjohn, R. J., "An Investigation of Several Frequency-domain Processing Methods for Enhancing the Intelligibility of Speech in Wideband Random Noise," Proc. Int. Conf. on Acoustics Speech and Signal Processing, pp. 602-605, 1978.
- 17. Markel, J. D. and Gray, A. H., "Linear Prediction of Speech," Springer-Verlag, New York, p. 190, 1976.
- 18. Widrow, B., "Adaptive Filters," in <u>Aspects of Network</u> and <u>System Theory</u>, R. Kalman and N. DeClaris, Eds., Holt, Rinehart, and Winston, New York, pp. 563-87, 1971.
- 19. Ahmed, N., Elliott, G. R. and Sterns, S. D., "Long-Term Stability Considerations of Adaptive Predictors," Conference Record of the Twelfth Asilomar Conference on Circuits, Systems and Computer, Asilomar, Calif, November 1978.
- 20. Soldan, D. L., "A Study of the Sequential Regression (SER) Algorithm and its Applications," Ph.D. dissertation, Kansas State University, Manhattan, Kansas, 1979.
- 21. Secrest, B. G. and Doddington, G. R., "Postprocessing Techniques for Voice Pitch Trackers,"Proc. IEEE Int. Conf. on Acoustics Speech and Signal Processing, pp. 172-175, 1982.

22. Laver, J., Hiller, S. and Hanson, R., "Comparative Performance of Pitch Detection Algorithms on Dysphonic Voices," Proc. IEEE Int. Conf. on Acoustics Speech and Signal Processing, pp. 192-195, 1982.

- 23. Boll, S. F., "A Spectral Subtraction Algorithm for Suppression of Acoustic Noise in Speech," IEEE Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 200-203, 1979.
- 24. Lim, J. S., "A Spectral Subtraction Algorithm for Suppression of Acoustic Noise in Speech," IEEE Trans. Acoustics, Speech, Signal Processing, Vol. ASSP-26, pp. 471-472, 1978.
- 25. Ketchum, J. W. and Proakis, J. G., "Adaptive Algorithms for Estimating the Suppressing Narrowband Interference in PN Spread-Spectrum Systems," <u>IEEE Trans. on Comm</u>., Vol. Com-30, pp. 913-924, 1982.
- 26. Viswanathan, R., Russel W. and Makhoul, J., "Objective Speech Quality Evaluations of Narrowband LPC Vocoders," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing.
- 27. Nawab, H., Oppenheim, A. V. and Lim, J. S., "Improved Spectral Subtraction for Signal Restoration," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 1105-1108, 1981.
- 28. McAulay, R. J. and Malpass, M. L., "Speech Enhancement Using a Soft-Decision Noise Suppression Filter," <u>IEEE Trans. Acoustics, Speech, and Signal</u> <u>Processing</u>, Vol. ASSP-28, No. 2, p. 137, April 1980.
- 29. McAulay, F. J. and Malpass, M. D., "A Real-Time Noise Suppression Filter for Speech Enhancement and Robust Channel Vocoding," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 699-702, 1980.
- 30. Langhans T. and Strube, H. W., "Speech Enhancement by Nonlinear Multiband Envelope Filter," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 156-159, 1982.
- 31. Berouti, M., Schwartz, R. and Makhoul, J., "Enhancement of Speech Corrupted by Acoustic Noise," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 208-211, 1979.
- 52. Woods, J. and Ingle, V. K., "Two-Dimensional Processing of Spectrogram Data," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, p. 39, 1978.
- 33. Oppenheim, A. V., Lim, J. S., Kopec, G. and Pohlig, S. C., "Phase in Speech Pictures," Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, pp. 632-637, 1979.
- 34. Wang, D. L. and Lim, J. S., "The Unimportance of Phase in Speech Enhancement," <u>IEEE Trans. Acoustics,</u> <u>Speech, and Signal Processing</u>, Vol. ASSP-30, No. <u>4</u>, August 1982.
- 35. Robinson, E. A. and Treitel, S., <u>Geophysical Signal</u> Analysis, Prentice-Hall, Inc., pp. 163-169, 1980.
- 36. Akaike, H., A New Look at the Statistical Model Identification, <u>IEEE Trans. Autom. Contr.</u>, Vol. AC-19, 1974, pp. 716-723.
- 37. Kay, S. M. and Marple, S. L., Spectrum Analysis-A Modern Perspective, <u>Proc. IEEE</u>, Vol. 69, pp. 1380-1419, 1981.
- j8. Trench, W., An Algorithm for the Inversion of Finite Hankel Matrices, SIAM J. Appl. Math., Vol. 13, pp. 1102-1107, 1965.
- 59. Morf, M., Lee, D. T., Nickolls, J. R. and Vieira, A., "A Classification of Algorithms for ARMA Models and Ladder Realizations," Proc. IEEE Conf. on Acoustics, Speech, and Signal Processing, Hartford, CT, pp. 13-19, 1977.
- 40. Morf, M., Vieira, A. and Lee, D., "Ladder Forms for Identification and Speech Processing," Proc. IEEE Conference on Decision and Control, New Orleans, LA, pp. 1074-1078, Dec. 1977.
- 41. Morf, M. and Lee, D., "Recursive Least Squares Ladder Forms for Fast Parameter Tracking," Special Session on Adaptive Noise Cancelling, Proc. 1978 IEEE Conf. Decision and Control, San Diego, CA, pp. 1362-1367, 1979.
- 42. Morf, M., "Ladder Forms in Estimation and System Identification," Proc. 11th Asilomar Conf. on Circuits, Systems and Computers, Nov. 7-9, Monterey, CA, pp. 424-429, 1977.

- 43. Lee, D. T. L., Morf, M. and Friedlander, B., "Recursive Square-Root Ladder Estimation Algorithms," IEEE Joint Special Issue on Adaptive Signal Processing, Vol. CAS-28, No. 6, pp. 627-641, June 1981.
- 44. Morf, M., Dickinson, B., Kailath, T. and Vieira. A., "Efficient Solutions of Covariance Equations for Linear Prediction," IEEE Trans. Acoustics, Speech and Signal Processing, Vol. ASSP-25, No. 5, pp. 429-435, October 1977.
- 45. Morf, M., Vieira, A., Lee, D. and Kailath, T., "Recursive Multichannel Maximum Entropy Spectral Estimation," IEEE Trans. Geoscience Electr., Vol. GE-16, No. 2, pp. 85-94, April 1978.
- 46. Friedlander, B., Morf, M., Kailath, T. and Ljung, L., New Inversion Formulas for Matrices Classified in terms of their distance from Toeplitz Matrices, Lin. Alg. and Applns., Vol. 27, pp. 31-60, 1979.
- 47. Graybill, F. A., <u>Introduction to Matrices with</u> <u>Applications in Statistics</u>, Wadsworth Publishing Company, Inc., p. 211, 1969.
- 48. Zohar, S., The Solution of Toeplitz Set of Normal Equations, J. Assoc. Comput. Mach., Vol. 21, pp. 272-276, 1974.
- 49. Friedlander, B., Kailath, T., Morf, M. and Ljung, L., "Extended Levinson and Chandrasekhar Equations for General Discrete-time Linear Estimation Problems," <u>IEEE Transactions on Automatic</u> Control, Vol. AC-23, pp. 653-659, August, 1978.
- 50. Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," Trans. of ASME, J. of Basic Engr., pp. 35-45, March, 1960.
- 51. Meier, L., Peschon, J. and Dressler, R., "Optimal Control of Measurement Subsystems," IEEE <u>Transactions on Automatic Control</u>, Vol. AC-12, pp. 528-535, October, 1967.
- 52. Sauer, G. and Melsa, J., "Stochastic Control with Continuously Variable Observation Costs for a Class of Discrete Nonlinear Systems," <u>IEEE</u> <u>Transactions on Automatic Control</u>, Vol. AC-19, pp. 234-239, June, 1974.
- 53. Herring, K. and Melsa, J., "Optimum Measurements for Estimation," IEEE Transactions on Automatic Control, Vol. AC-19, pp. 264-266, June, 1974.

54. Athans, M., "On the Determination of Optimal Costly Measurement Strategies for Linear Stochastic Systems," <u>Automatica</u>, Vol. 18, pp. 397-412, July, 1972.

- 55. Laughhunn, D., "Quadratic Binary Programming with Application to Budgeting Problems," <u>Operations</u> <u>Research</u>, Vol. 18, pp. 456-461, 1970.
- 56. Gantmacher, F., The Theory of Matrices, Vol. 1, Chelsea Publishing Co., N.Y., pp. 317-326, 1959.
- 57. Meditch, J., <u>Stochastic Optimal Linear Estimation and</u> Control, <u>McGraw Hill</u>, N.Y., p. 102, 1969.
- 58. Sage, A., and Melsa, J., <u>Estimation Theory with</u> <u>Application to Communications and Control</u>, McGraw Hill, N.Y., p. 500, 1971.

PART B

AUTOREGRESSIVE SPECTRAL ESTIMATION IN NOISE FOR SPEECH ANALYSIS APPLICATIONS

TABLE OF CONTENTS

Chapter	•	Page
I.	INTRODUCTION	104
	Motivation	105
	Motivation	108
II.	GENERAL DISCUSSION	110
	Model Selection	112
	Model Selection	116
	Noise Corruption	121
	Noise Filtering	123
	Effect on Resolution	
		126
	Quantization and Computation	129
	Remarks	131
111.	THEORETICAL FORMULATION	134
	Application to Speech Analysis	135
	Ideal Formulation	137
	Noise Filtering Formulation	144
	Weighted Information Formulation	145
	Properties of the Weighted Information	147
	Remarks	155
IV.	COMPUTATIONAL FORMULATION	159
	Poduction to Algobraic Por-	160
	Reduction to Algebraic Form	
	Evaluation of Coefficients	169
	Iterative Techniques	174
	Formulae for Vector Quantization	185
	Remarks	193
۷.	RESULTS	197
	Gaussian Signals	197
	Speech and Speech-Like Signals	221
VI.	CONCLUSION	228
	Suggestions for Future Research	230
CITE	D REFERENCES	234

÷.,

LIST OF FIGURES

ĥ

Figu	re	Page
1.	Noise Filter Characteristics	127
2.	Time Domain Coefficient Evaluation	171
3.	Frequency Domain Coefficient Evaluation	173
4.	Mixed Time-Frequency Domain Coefficient Evaluation	175
5.	The contour in the complex z-plane	188
6.	Test Signal Spectra	199
7.	Scatter Plot; Estimates of Noiseless Signal Process	201
8.	Gaussian Signal & Colored Noise (10 dB)	202
9.	Gaussian Signal & Colored noise (10 dB)	202
10.	Noise Filtering; Gaussian Signal & Colored Noise (10 dB)	204
11.	Weighted Information; Gaussian Signal & Colored Noise (10 dB)	204
12.	Noise Filtering; Gaussian Signal & Colored Noise (10 dB)	206
13.	Weighted Information; Gaussian Signal & Colored Noise (10 dB)	206
14.	Noise Filtering; Gaussian Signal & White Noise (10 dB)	208
15.	Weighted Information; Gaussian Signal & White Noise (10 dB)	208
16.	Noise Filtering; Gaussian Signal & White Noise (10 dB)	209
17.	Weighted Information; Gaussian Signal & White Noise (10 dB)	209 Page

18.	Noise Filtering; Gaussian Signal & White Noise (O dB)	210
19.	Weighted Information; Gaussian Signal & White Noise (O dB)	210
20.	Noise Filtering; Gaussian Signal & White Noise (O dB)	211
21.	Weighted Information; Gaussian Signal & White Noise (O dB)	211
22.	Periodic Signal & Colored Noise (10 dB)	214
23.	Periodic Signal & Colored Noise (10 dB)	214
24.	Noise Filtering; Periodic Signal & Colored Noise (10 dB)	215
25.	Weighted Information; Periodic Signal & Colored Noise (10 dB)	215
26.	Noise Filtering; Periodic Signal & Colored Noise (10 dB)	216
27.	Weighted Information; Periodic Signal & Colored Noise (10 dB)	216
28.	Noise Filtering; Periodic Signal & White Noise (10 dB)	217
29.	Weighted Information; Periodic Signal & White Noise (10 dB)	217
30.	Noise Filtering; Periodic Signal & White Noise (10 dB)	218
31.	Weighted Information; Periodic Signal & White Noise (10 dB)	218
32.	Noise Filtering; Periodic Signal & White Noise (O dB)	219
33.	Weighted Information; Periodic Signal & White Noise (O dB)	219
34.	Noise Filtering; Periodic Signal & White Noise (O dB)	229
35.	Weighted Information; Periodic Signal & White Noise (O dB)	220
		Page

6

A a a a a a a a a a

Ĩ

36.	Vowel Spectrum in Quiet Environment 2	24
37.	Vowel Spectrum in White Noise 2	25
38.	Vowel Spectrum; Power Subtraction 2	27
39.	Vowel Spectrum; Magnitude Subtraction 2	27

.

CHAPTER 1

INTRODUCTION

Nearly all sciences are concerned with the analysis of measurement data. The following chapters will present a new tool for the analysis of time series measurements; in particular, a new method of spectral estimation is presented. Many spectral estimation methods already exist and, increasingly, new methods continue to be developed; therefore, it is appropriate to reflect, briefly, upon the reasons for such continued activity in an area already so well researched.

A synergism exists between advances in computer technology and advances in practical methods of time series analysis. As more effective (and complex) methods of time series analysis are developed, the demands for smaller, cheaper, and faster digital circuitry (capable of implementing these methods within the size/cost/power constraints of various applications) are increased. As smaller, cheaper, faster and more reliable digital circuitry becomes available, more complex (and effective) methods of time series analysis become practical. Fundamentally, however, it is the demand for improved solutions to engineering problems that motivates the desire for more effective methods of time series analysis.

Motivation

Most information we have about the world around us is received indirectly through time series measurements. In the case of vision, one determines the shape (and other characteristics) of an object by reception (measurement) of light waves scattered by the object. In the case of speech, one determines the intended message of the speaker by reception (measurement) of acoustic pressure waves. Prospecting, manufacturing, astronomy, medicine, and economics are but a few of the areas that can benefit from improved methods of time series analysis.

Spectral estimation is one of the most important areas of time series analysis. In many cases, knowledge of the time series spectrum is adequate to answer all important questions regarding the system producing the time series; in the case of a stable time-invariant linear input-output system, knowledge of the output process spectrum (together with the statistics of the stationary input process) will completely characterize the system.

Noise corruption is among the fundamental problems of time series analysis. All useful analysis techniques for measurement data are at least mildly tolerant of noise since there always exists a small probability of measurement error; some techniques are specifically designed to account for knowledge of the noise statistics in the analysis of noise-corrupted measurement data. Regardless of the analysis technique, the fundamental performance limits are always

reduced by the presence of noise.¹ Consequently, it is always advisable to minimize noise corruption as much as is practical; still, practical constraints imposed by some situations do not permit the reduction of noise corruption to insignificant levels so that sophisticated analysis techniques are required to achieve the best possible performance.

Spectral estimation is of fundamental importance to the various applications of speech analysis and practical constraints imposed by many of these applications do not permit the reduction of noise corruption to insignificant levels. such applications include low Examples of data rate digital voice communications systems and speech recognition/understanding systems among others; often the cost and/or inconvenience of shielding from environmental noise makes significant acoustic noise corruption inevitable.

Autoregressive (AR) spectral models have been successful for various systems involving speech analysis; moreover, numerous speech synthesis systems based upon the AR model have become commercially available in recent years. Because the currently available practical methods for AR parameter estimation yield poor results in common noise

¹In some specialized circumstances the performance limits are unchanged by the presence of noise. Even when this is the case, the complexity of the analysis methods required to achieve these limits is usually increased by the noise presence.

environments but are effective in sufficiently quiet environments, it is reasonable to retain the AR model for the speech process while attempting to develop improved methods for estimating the AR parameters.

The fundamental limit to the performance of any estimation procedure depends upon the available information. In theory, even the most obscure (but not unrelated) additional information may be used to improve a parameter estimate; of course, one should rely first upon information that is both easily available and expected to provide substantial improvement.

Most recent efforts to overcome the poor performance of classical AR estimators in noise, including the present one, have attempted to employ information regarding the noise statistics in addition to the noise corrupted time series observations. This information is often provided simply by deploying additional sensors intended to measure the noise directly; other speech analysis systems employ prior segments of the primary observation signal that are thought to be free from speech activity to predict the current relevant noise statistics.

The present work does not address the problem of obtaining accurate noise statistics. Assuming appropriate noise statistics to be available, the following chapters develop a new and improved method of estimating the AR signal parameters from noise corrupted time series observations. As might be expected, the method entails increased computational cost over less effective techniques; it is expected that performance requirements of speech analysis (and other) applications - as well as cost reductions that are continually provided by advances in computer technologyshall, in many cases, make the advantages of this method appear relatively inexpensive.

Overview

Chapter II provides a general discussion of the various issues and techniques of spectral estimation; particular attention is given to the problems of AR spectral estimation. In addition, this discussion introduces basic formulae and provides an historical perspective for the subsequent chapters.

Chapter III presents the theoretical foundations of the new (weighted information) estimation procedure. After some additional motivational discussion, the method is formulated as an approximation to an ideal (but intractable) formulation and a generalization of a commonly employed (noise filtering) estimation procedure. In addition to the general formulation, significant contributions of this chapter include the analogy leading to equation (3.20) and the properties developed in the fifth section.

Chapter IV discusses a variety of computational methods relevant to AR estimation based upon the weighted information formulation. It is considered that the area of

computational procedures as requiring the greatest attention for further extension and refinement of this work. Only the formulae for vector quantization, in particular Equations (4.58a) and (4.81), appear ready for detailed cost/performance analyses.

Chapter V demonstrates clearly that the weighted information formulation leads to reduced estimation error as compared to the more common noise filtering formulation. Examples from both simulated and real speech are provided. The demonstration relies upon the reader's visual assessment of scatter plots; thus it is somewhat qualitative. A more quantitative assessment (e.g. a comperison of empirical variance to theoretical performance bounds) would be interesting; however, one would still have difficulty evaluating the significance of a reduction in empirical variance to the performance of a particular system. Without a full implementation one must rely upon experience and judgement as well as the available experimental evidence.

Finally, Chapter VI summarizes the results of this effort and provides suggestions as to how this work may be effectively extended and refined.

CHAPTER II

GENERAL DISCUSSION

Spectral estimation is a problem of statistical inference with a long history due to its pervasive importance in scientific applications [1]. Modern empirical spectral analysis began to take shape as an organized discipline with the introduction in 1893 of the periodogram by Schuster [2].

Given N observations $\{x_n; n=0,1,\ldots,N-1\}$ of a time series at unit time intervals the periodogram, $f(\theta)$, is defined as

$$f(\theta) = X_{N}(e^{i\theta}) X_{N}(e^{-i\theta})/N$$
(2.1)

where

$$X_{N}(z) = \sum_{n=0}^{N-1} x_{n} z^{-n}; z = e^{i\theta}$$
 (2.2)

Still in use today, the periodogram was practically the sole computational tool of empirical spectral analysis until Yule introduced in 1927 his method of autoregressive (AR) spectral analysis [3].

An AR(P), or Pth order autoregressive, model spectrum, $g(\theta)$, is characterized by a model gain, σ , and a monic Pth order polynomial, $z^{P}A_{P}(z)$, and is defined by them as

$$g(\theta) = \sigma^2 / |A_p(e^{i\theta})|^2$$
(2.3)

The polynomial may be characterized by a variety of parameter sets. One parameter set, known as predictor coefficients $\{a_n; n=1,2,\ldots,P\}$, defines the polynomial according to

$$A_{P}(z) = \sum_{n=0}^{P} a_{n} z^{-n} ; a_{0} = 1$$
 (2.4)

In contrast to Schuster's nonparametric method of spectral analysis, Yule's parametric method first introduces the above mathematical model, justified by physical arguments, and then uses the available data to estimate the model parameters. These estimates are provided by the solution to the Yule-Walker [4] equations

$$\sum_{m=0}^{P} r_{jn-mj} a_{m} = \sigma^{2} \delta_{n} ; n=0,1,\ldots,P \qquad (2.5)$$

where

$$r_n = \sum_{m=0}^{N-n-1} x_m x_{m+n} / N ; n=0,1,...,P \qquad (2.6)$$

are the biased sample autocorrelation lag estimates.

Model Selection

A variety of other parametric spectral models have been introduced and studied during the past half century; several of them are worth noting. The moving-average (MA) model, like the AR model, is characterized by a polynomial but differs in that the polynomial appears in the numerator; the Schuster periodogram may be viewed as an MA model spectrum.¹ Similarly, ARMA models are described by both numerator and denominator polynomials; these spectra are of particular importance in engineering applications since they characterize all stable linear systems with a finite dimensional state vector. The Blackman-Tukey [5] model spectrum consists of a finite sum of cosine terms; it is obtained by Fourier [6] transformation of the product of the autocorrelation sequence and a finite support window. The Pisarenko [7] model consists of a constant plus a finite number of delta functions. Various combinations of these models are also occasionally employed.

Most often a new model is introduced (together with a procedure for estimating its parameters) simply because it seems reasonable relative to the phenomenon being studied and due to deficiencies in the currently popular

¹Facts such as these tend to blur the distinction between parametric and nonparametric methods. Since any estimate can be described as a member of some parametric family once it has been derived, the distinction may be seen as one of spirit rather than substance.

models.^{2, 3} More recently the various results of this "unscientific" approach have been "justified" theoretically; this justification usually takes the form of a principle that should be employed as a guide when the requirement of consistency with the available information leaves several alternatives. The principle is usually embodied in the form of a functional whose extreme value is to be found while the information is provided in the form of constraint equations (or inequalities) for this variational problem.

Much of the current literature is devoted to the "principle of maximum entropy" which was enunciated by Jaynes [8, 9]. If the process is zero-mean stationary and Gaussian⁴, it is completely characterized by its power spectral density function, $g(\theta)$, (or "spectrum" for short) and the process entropy is expressed in terms of it by

$$Q = \int_{-\pi}^{\pi} \ln g(\theta) d\theta / 2\pi$$

(2.7)

⁴The Gaussian assumption may be avoided in the case of correlation constraints. Working directly with probability densities the Gaussian form may be derived as that which maximizes the entropy [10, p. 944].

²We shall adopt this pragmatic view later when modeling speech in an acoustically noisy environment.

³Sometimes a model is used in spite of its less reasonable form simply because the available parameter estimation methods yield more successful overall results. Thus AR models are employed (instead of the Pisarenko model) to estimate the frequencies of pure sinusoids in white noise from short data records.

As demonstrated by Burg [11], if the entropy is subsequently maximized subject to correlation constraints⁵

$$\mathbf{r}_{n} = \int_{-\pi}^{\pi} \mathbf{g}(\boldsymbol{\theta}) \, e^{in\boldsymbol{\theta}} \, d\boldsymbol{\theta}/2\pi \; ; \; n=0,1,\ldots,P \qquad (2.8)$$

one may derive the AR(P) form for $g(\theta)$ as given by Equation (2.3). The AR(P) form together with the constraint Equations (2.8) are then sufficient to yield the Yule-Walker Equations (2.5) from which the model parameters may be determined. If cepstral constraints⁶ are employed in place of correlation constraints the spectrum maximizing Equation (2.7) has an MA form while both correlation and cepstral constraints lead to an ARMA model. The Pisarenko model is "justified" by deriving it as the minimum energy solution under correlation constraints⁷, excepting the energy ($\alpha = 0$) constraint [12].

Another principle discussed in the recent literature is the "principle of minimum cross-entropy" [13]. Introduced by Kullback (under the name "directed divergence") as an

⁷It may also be related to the maximum entropy principle by noting that the AR(P) model approaches the Pisarenko model as r_0 is decreased to the point where the correlation matrix becomes singular [7, p. 355].

⁵The values on the left-hand side are given in terms of the data; for example, by Equation (2.6).

⁶These place constraints directly on the "cepstrum" (or log power spectrum) and are expressed by Equations (2.8) if $g(\theta)$ is replaced by its logarithm while the left-hand side values are expressed in terms of the data.

information measure [14], it has a number of interesting properties neatly collected in [15]. In terms of probability densities the cross-entropy is given by

$$S(q,p) = \int q(\bar{x}) \ln[q(\bar{x})/p(\bar{x})] d\bar{x} \qquad (2.9)$$

and measures the expected information for discrimina ion⁸ per observation from $q(\bar{x})$ [14]. A symmetric version of this measure, S(q,p) + S(p,q), was introduced earlier by Jeffreys [16] who emphasized the invariance of this measure with respect to coordinate transformations; unlike entropy, cross-entropy shares this important property.

As an inference procedure, minimum cross-entropy analysis requires a prior estimate of the density, $\underline{p}(\bar{x})$, as well as new information in the form of constraints and derives a new posterior estimate of the density, $q(\bar{x})$, by minimizing S(q,p) subject to the constraints [17]. In the case that the prior density is uniform the procedure is equivalent to maximum entropy; with correlation constraints the posterior density is found to be Gaussian AR(P) with parameters satisfying the Yule-Walker Equations (2.5).

⁸Fully, S(q,p) is said to measure the expected information for discrimination in favor of the (correct) hypothesis that the density is $q(\bar{x})$ and against the (competing) hypothesis that the density is $p(\bar{x})$ per observation from $q(\bar{x})$.

Parameter Estimation

The foregoing discussion leaves the impression that the correct path to formation of a spectral estimate is clear: simply select a guiding principle (undoubtedly related to the notion of entropy), gather the available information, and solve the well defined mathematical problem that results. Seldom is the practical situation so simple.

Typically the numerical constraints are not given conveniently, say, in terms of exact knowledge of the autocorrelation function at equally spaced lags. More often, only a few irregularly spaced noise corrupted samples of the time series are available; from this data the numerical constraints must be estimated. Even when permitted the luxury of bountiful regularly spaced and noise-free data, numerous difficulties remain. Assuming a maximum entropy principle, should estimates of the autocorrelation, cepstral, or some other numerical constraints be formed? How should these estimates be formed and how many⁹ of them should be formed?

The Yule AR(P) estimation procedure outlined at the beginning of this chapter provides one solution: having selected the model as AR and its order as P, form the biased autocorrelation lag estimates, Equation (2.6), and use these

⁹This is the problem of order determination. Various estimators of the order parameter, based upon notions of information theory, have been proposed and discussed by Akaike [18] and Parzen [19, 20] among others. Often the order parameter is selected simply upon the basis of experience with the phenomenon under study.

as if they were the true values. These autocorrelation lags then uniquely determine the AR(P) model parameters (and vice versa) via the Yule-Walker Equations (2.5). This description is explicit but fails to provide significant insight as to why it might be good. The formulation may be derived from a variety of viewpoints, each with its own merit and yielding greater understanding of the procedure.

Linear Prediction (LP) theory leads to one derivation of this formulation [21]. In this derivation the AR model is viewed as a predictor and the model parameters are determined to minimize the prediction error

$$e_n = x_n - \hat{x}_n = x_n + \sum_{m=1}^{P} a_m x_{n-m}$$
 (2.10)

in a mean-square sense. Depending upon the details of treatment of the ends of the data record one may derive the Yule-Walker procedure (also known as the "autocorrelation LP method") or a variant known as the "covariance LP method". Both of these methods have their proponents. The Linear Prediction theory is very similar to Yule's original considerations in which the e_n are viewed as random driving disturbances to the Pth order inhomogeneous difference Equation (2.10).

Other variants of the autocorrelation LP method are based upon a recursive lattice structure for the prediction filter [22]. In addition to the "forward" predictor $A_p(z)$, these variants consider a "backward" predictor, $B_p(z)$; both predictors are characterized by the set of reflection coefficients $\{k_n; n=1,2,\ldots,P\}$ according to

$$A_n(z) = A_{n-1}(z) + k_n z^{-1} B_{n-1}(z) ; A_0(z) = 1$$
 (2.11a)

$$B_n(z) = z^{-1}B_{n-1}(z) + k_n A_{n-1}(z)$$
; $B_0(z) = 1$ (2.11b)

The z-transform of the forward prediction error process after n filtering stages is simply $A_n(z) X(z)$; similarly the z-transform of the backward prediction error process is $B_n(z) X(z)$. Mean-square criteria are applied to the forward and backward error processes to obtain a variety of estimators for the reflection coefficients: one of particular importance, due to Burg [23], determines k_n to minimize the sum of the variances of the forward and backward error processes after n filtering stages. For truely ergotic processes, all these AR estimation procedures are asymptotically equivalent to the autocorrelation LP method for large values of N; as parameter estimation procedures these methods are most important for problems involving mildly nonstationary data of limited quantity.

In addition to these various "minimum mean square prediction error" formulations, another important derivation of the Yule procedure is due to Itakura and Saito [24]. Assuming an AR(P) model for the zero-mean stationary Gaussian process, they employ the maximum likelihood method and show

that the solution is obtained, asymptotically for large N, by minimizing a "spectral matching criterion"

$$I(f,g) = \int_{-\pi}^{\pi} \{ [f(\theta)/g(\theta)] - \ln[f(\theta)/g(\theta)] - 1 \} d\theta/2\pi \qquad (2.12)$$

where $f(\theta)$ is the Schuster periodogram given by Equation (2.1).

It is readily verified, by differentiating I(f,g) with respect to the parameters of $g(\theta)$, that the minimum is obtained when the correlation matching property

$$\int_{-\pi}^{\pi} f(\theta) e^{in\theta} d\theta/2\pi = \int_{-\pi}^{\pi} g(\theta) e^{in\theta} d\theta/2\pi \qquad (2.13)$$

is satisfied for n=0,1,...P. By recognizing the left-hand side as the lag product autocorrelation estimates

$$\mathbf{r}_{n} = \int_{-\pi}^{\pi} \mathbf{f}(\theta) \ e^{in\theta} \ d\theta/2\pi \qquad (2.14)$$

the correlation matching property leads easily to the Yule-Walker Equations (2.5); see [25, pp. 445-6]. Recently Kay [26] has developed another variant by similarly applying the maximum likelihood method to zero-mean stationary Gaussian AR(P) processes but eliminating the large N approximation; again this variant treats the problem of limited data.

The functional (2.12), although it is usually attributed to Itakura and Saito in the current speech literature, was apparently first developed by Pinsker [27]. Assuming

only that the two processes are zero-mean and Gaussian, Pinsker showed¹⁰

$$\lim_{N \to \infty} S(p,q)/N = I(p,q)/2$$
 (2.15)

This theorem provides an information theoretic interpretation of the Itakura-Saito spectral matching criterion. Moreover, from a functional inference point of view, one might derive the Yule-Walker procedure by replacing q by an assumed AR(P) spectral model, $g(\theta)$, replacing p by a rough spectral estimate provided by $f(\theta)$, and then minimizing I(f,g).

The last derivation should be contrasted with the minimum cross-entropy development discussed earlier. In that formulation the AR(P) form was derived from given correlation constraints while this formulation derives the correlation constraints from the given AR(P) form. Both developments employ (different) prior estimates and minimize a measure of information divergence between the prior and posterior estimates; however, the information divergence is not a symmetric measure and the unknown (posterior) estimate appears as the second argument in the current formulation

 $^{^{10}}$ The notation is somewhat abused here. On the left p and q represent the joint probability densities of N consecutive random variables; on the right p and q are power spectral density functions.

while it appears as the first argument in the minimum crossentropy development. Nonetheless, the resultant procedures are both the same as the Yule procedure. In the next chapter a variant of this last derivation will be considered.

Noise Corruption

The problem of noise corruption to the observations pervades estimation problems. Generally all useful estimators are at least mildly tolerant of noise corruption while their performance degrades if the corruption becomes particularly severe. The most common problem considered is that of an additive independent noise process; this problem is of considerable importance in practical applications.

Upon initial reflection, the problem of estimating the parameters of both the noise and signal processes from time series observations alone may seem impossible. Indeed, the problem of determining the individual variances of two independent additive zero-mean stationary white Gaussian processes is completely confounded regardless of the quantity of data available. However, if one process is non-Gaussian, estimates of third and higher order statistics can be useful in estimating these lower order statistics. Parzen discusses the use of the "bispectrum" to estimate the spectrum of a non-Gaussian process in additive independent white Gaussian noise [28].

When both processes are Gaussian the problem is not always confounded. Since the sum of two additive

independent ARMA processes is also an ARMA process one might hope to find estimators for the parameters of the two additive processes when the number of parameters for the combined process is not exceeded by the total number of parameters of the two processes. For example, Pagano [29] discusses the problem of estimating the P + 2 parameters of additive AR(P) and white processes by first estimating the 2P + 1 parameters of a single equivalent ARMA(P,P) process and then using these 2P + 1 estimates to initialize a procedure for estimating the originally sought P + 2 parameters; it seems critical however that the order of the AR process does not degenerate (i.e. is actually nonzero).

This latter problem is fairly close in spirit to the problem considered in the following chapters. There the signal and noise processes are additive, independent, and zero-mean Gaussian; moreover, the signal process is AR(P). The problem may seem more complex because the noise process need not be white; however, a considerable simplification is achieved because the noise process spectral density (hence, all its statistics) is assumed to be known in addition to the time series observations. In practice the noise statistics are estimates provided by other observations but the large amount of data available for these estimates makes them quite reliable.

Noise Filtering

Wiener [30] considered the intimately related problem of extrapolating a time series from noise corrupted observations. When the zero-mean signal and noise processes are additive and independent with known power spectral density functions $(g(\theta) \text{ and } \mu(\theta) \text{ respectively})$ then the minimum variance linear extrapolating filter is the Wiener filter whose frequency response characteristic is

$$H(\theta) = g(\theta) / [g(\theta) + \mu(\theta)]$$
(2.16)

This is sometimes referred to as the unrealizable Wiener filter since it is noncausal; the corresponding impulse response function extends both backward and forward in time to infinity. It is easy to show that the variance of the extrapolation can only be reduced to zero if the support of the signal spectrum has a null (or zero-measure) intersection with the support of the noise spectrum; in this case the frequency response, $H(\theta)$, will be unity on the support of $g(\theta)$ and zero elsewhere. Others, most notably Kalman [31], have since extended and refined Wiener's pioneering work.

A common procedure for dealing with additive noise is to first form a realizable estimate of the Wiener filter (or some other "optimal" filter), $\hat{H}(\Theta)$, and apply it to the noise corrupted observations. The resulting data are then treated as noise-free observations of the signal process and standard estimation procedures are employed to obtain an estimate of the signal spectrum. When the noise spectrum, $\mu(\theta)$, is known this procedure involves some mildly circular reasoning since Equation (2.16) indicates that knowledge of $H(\theta)$ is equivalent to knowledge of $g(\theta)$.¹¹ Nonetheless, this process has been demonstrated to be advantageous in speech analysis and other applications; a survey of these methods may be found in [32].

Ţ

Much recent effort [33-39] has concentrated upon implementation structures and estimation procedures for $\hat{H}(\theta)$; typically these procedures employ side information in addition to the noise corrupted time series observations. Often the methods are nonlinear and time-varying with both theoretical and heuristic foundations. Regardless of the technique, one may always subsequently define a short-timeinvariant linear equivalent frequency response characteristic in terms of the short-time input and output signal z-transforms, X(z) and Y(z), by

 $\hat{H}(\theta) = Y(e^{i\theta})/X(e^{i\theta})$ (2.17)

¹¹Hence we would have $\hat{g} = \mu \hat{H}/(1-\hat{H})$. The conceptual difficulties may be circumvented by considering the overall noise cancelling filter/spectral estimation scheme as a single estimation procedure; especially since the procedure usually does not employ (2.16) to form the final estimate of the signal spectrum.

One convenient categorization distinguishes frequency domain methods [33-36] from time domain methods [37-39]. Among the frequency domain methods, the noise cancelling filter frequency response characteristic usually appears explicitly; the simpler (and less heuristic) methods present $\hat{H}(\theta)$ as a function of the short-time signal to noise spectral density ratio estimate¹²

$$SNR(\boldsymbol{\theta};\boldsymbol{\alpha}) = \{ [\mathbf{f}(\boldsymbol{\theta})]^{\boldsymbol{\alpha}} \neq [\boldsymbol{\mu}(\boldsymbol{\theta})]^{\boldsymbol{\alpha}} \} / [\boldsymbol{\mu}(\boldsymbol{\theta})]^{\boldsymbol{\alpha}}$$
(2.18)

Two important classes of filter response characteristics are the subtraction class given by^{13}

$$\hat{H}_{1}(\boldsymbol{\theta};\boldsymbol{\alpha},\boldsymbol{\beta}) = \{SNR(\boldsymbol{\theta};\boldsymbol{\alpha})/[1 + SNR(\boldsymbol{\theta};\boldsymbol{\alpha})]\}^{\boldsymbol{\beta}}$$
(2.19)

and the soft suppression class given by

$$\hat{H}_{2}(\theta;\alpha,\beta) = \{ [1 + \hat{H}_{1}(\theta;\alpha,1/2)]/2 \} \{ \Phi(\theta;\alpha,\beta) / [1 + \Phi(\theta;\alpha,\beta)] \}$$
(2.20a)

¹²Equation (2.18) employs the monus function, defined by x+y = (x-y + |x-y|)/2, to insure a nonnegative result.

¹³Various special frequency response characteristics are worth separate mention here. The Wiener filter [30] frequency response is $\hat{H}_1(\theta;1,1)$. The power subtraction filter and the magnitude subtraction filter [35] have frequency response characteristics $\hat{H}_1(\theta;1,1/2)$ and $\hat{H}_1(\theta;1/2,1)$ respectively. Finally, the soft suppression class due to McAulay and Malpass [36] has the frequency response $\hat{H}_2(\theta;1,\beta)$.

where

$$\Phi(\theta;\alpha,\beta) = \exp[-\beta] I_0[2\sqrt{\beta[1 + SNR(\theta;\alpha)]}] \qquad (2.20b)$$

and $I_0[\cdot]$ denotes the zeroth order modified Bessel function of the first kind. These "suppression rules" are plotted for selected values of α and β as a function of SNR(θ) in Figure 1.

Effect on Resolution

In speech applications, vocal tract resonances are not extremely sharp and are moderately well separated in frequency; consequently one is generally concerned with accurate estimation of the spectral shape and high resolution estimation is not a priority.¹⁴ In other applications (such as sonar, radar, and medicine) accurate frequency estimation and resolution of discrete ("line") and narrowband spectra are issues of fundamental importance. Periodogram and Blackman-Tukey spectral estimates have a fundamental frequency resolution limit determined by the length of the observation interval; AR estimators have become quite popular due, in part, to their greatly improved resolving power.

 $^{^{14}\}mbox{Hence}$, even very low resolution methods that divide the (4 kHz) voice bandwidth into fewer than two dozen "channels" can be quite effective.



Still, the resolution (as well as other performance indicators) varies among the different AR estimators and, for each, is influenced by a variety of factors.

Noise corruption is one of the important factors limiting the resolving power of AR estimators. Several authors have considered the problem of estimating the parameters of a fixed number of sinusoids from discrete-time observations corrupted by zero-mean additive white Gaussian noise of unknown variance. For this specialized problem the Cramer-Rao performance bounds¹⁵ may be computed [40]. As is well known, the complicated nonlinear maximum likelihood estimation procedure will achieve these bounds; Tufts and Kumaresan [41], using AR estimation procedures, have developed computationally simpler high resolution frequency estimators that nearly achieve these bounds while Cadzow, et. al. [42] claim still better performance using a singular value decomposition (SVD) approach. In many practical circumstances additional information may be available so that

¹⁵In general, the Cramer-Rao bounds indicate the minimum variance a parameter estimate can achieve [43]. An estimate achieving the minimum variance is an "efficient" estimate. In [40] the bounds upon an unbiased frequency estimate are considered (they depend upon the assumed distribution as well as the number of data points) and are presented as a function of the signal to noise ratio. In [44], the efficiency loss of any method based upon the use of correlation estimates instead of the original data is studied.

these bounds may be exceeded;¹⁶ for example, Quirk and Liu [45] describe a simple filtering and decimation scheme (which employs knowledge of the frequency bands in which the sinusoids are located) that improves the resolution of (any) subsequent AR estimator. In a similar vein, adaptive prefilters (that employ a reference process correlated with either the signal or noise portion of the objective process, but not both) have been devised to "enhance" narrowband signals in noise [46].

Quantization and Computation

While spectral estimation, per se, is not concerned with the problems of quantization and computation, the ultimate utility of an estimation procedure can depend strongly upon these (and other) issues. If the procedure explicitly recognizes that only one of a finite predefined set of conclusions can be reached, the situation is sometimes distinguished by referring to the "detection" (instead of the "estimation") problem.

In many digital speech recognition and communication systems the goal of spectral analysis is to solve a detection problem; in addition, the system designer must solve the problem of selecting the best finite set of models to

¹⁶More precisely, the true bounds are reduced by the availability of additional information. Consequently new estimators that account for this additional information can be devised that outperform (in terms of variance) any estimator that does not account for the additional information.

employ. Until recently, these systems would find the solution to an estimation problem and then employ a (somewhat ad hoc) quantization procedure to select a model from among the finite set. If the number of models in the finite set was sufficiently large, this procedure could be quite effective; however, one measure of goodness for the finite set of models is often how few models are in the set.

In the past decade technological advances have permitted the use of increasingly complex computational procedures while still meeting size/cost/power constraints imposed by the application. Consequently more sophisticated and effective (but previously unmanagable) techniques for estimation/detection and quantization of spectral models have been studied in earnest. The numerous variants of a class of techniques generally referred to as "vector quantization" [47-53] have recently achieved considerable success by reducing the finite number of models by about 9 orders of magnitude with only slight degradation in other measures of system performance.

Many of these vector quantization techniques are founded upon minimization of the asymptotic information divergence I(f,g). Of considerable interest in the use of this measure is the triangle equality property; if $g(\theta)$ minimizes I(f,g) over the set of all stable AR(P) models and $h(\theta)$ is any other model in a (possibly finite) subset then

$$I(f,h) = I(f,g) + I(g,h)$$
 (2.21)

As a consequence of this property one may solve the detection problem, which minimizes I(f,h), by first solving the estimation problem, which minimizes I(f,g), and then solving the quantization problem which minimizes I(g,h).

Remarks

The general problem of spectral estimation has been discussed; this discussion has emphasized issues and methods associated with autoregressive estimation. Autoregressive spectral models are important in numerous practical applications; consequently they have received considerable attention in the literature. The AR form may be derived from either the maximum entropy or the minimum cross-entropy principle when correlation constraints are considered; alternatively the AR form may be assumed and correlation constraints derived using a linear prediction formulation. The correlation constraints, together with the AR form, are sufficient to derive the Yule-Walker equations which relate the model parameters to the prescribed correlation values.

The asymptotic maximum likelihood formulation of Itakura and Saito assumes an AR form and derives the correlation constraints; in the course of this development a "spectral matching criterion" is minimized. The earlier derivation by Pinsker of this spectral matching criterion from an asymptotic information divergence formulation makes clear that, while the AR form is necessary to derive the
Yule-Walker equations, the spectral matching criterion is applicable independant of the spectral model form.

Noise corruption pervades estimation problems and useful estimators are generally at least mildly tolerant of additive noise. Often additional data is available to help characterize or distinguish the noise and signal processes; many estimation problems are concerned with the development of effective and computationally feasible methods for incorporating this additional information. A common procedure, employed when an accurate noise spectrum estimate is known, first applies an estimated noise cancelling filter to the corrupted data and then uses the output as "noise-free" data from which to estimate the signal spectrum. Ultimately the effect of noise corruption will be to decrease the best performance possible with any spectral estimator.

In the following chapters a new spectral estimator is developed. As is common, the fundamental observations are assumed to be equally spaced samples of a zero-mean stationary Gaussian time series corrupted by additive independent zero-mean stationary Gaussian noise of known power spectral density, $\mu(\theta)$. This problem occurs in many applications involving speech analysis (as well as others) wherein the noise spectrum is estimated from data taken during speech inactivity.

The amount of data available to estimate the signal spectrum is usually limited by the nonstationary character of speech; the speech statistics are usually stationary only

over very short time intervals varying in duration. One study [54] has observed speech waveforms and subjectively judged that the duration for which a segment may be considered stationary varies from about 4 ms. to over 360 ms. with most of the distribution contained in the range of 12 ms. to 174 ms.; most speech analysis systems employ a fixed analysis interval approximately 20 to 25 ms. in duration. The use of a fixed analysis interval (with no particular attempt at optimum time alignment of end points) is simply a practical method of limiting the computational burden; while suboptimal spectral estimates are thereby achieved for long acoustic events, perhaps the most severe deleterious effect is the slurring of very short events and transitions.

In order to employ at a later time a noise estimate obtained during speech inactivity, the noise statistics are assumed to remain stationary over much longer time intervals; since one of the primary noise sources is ambient environmental noise acoustically coupled to the speech, the validity of this assumption must be checked in each situation. In many practical circumstances the noise is stationary over long intervals; for example, in aircraft, the noise statistics typically vary only with the flight condition. On the other hand, if the corrupting noise is another speech signal the assumption of long term noise stationarity is certainly invalid.

CHAPTER III

THEORETICAL FORMULATION

In this chapter several related procedures for estimating AR(P) process parameters from noise corrupted time series observations are developed. In the first section the problem is motivated as one arising in speech applications. In the next section an ideal formulation is discussed; unfortunately the resulting nonlinear system of equations is sufficiently complicated to make analytical solution intractable.¹ In the third section a first approximation to the ideal formulation is developed and shown to be essentially equivalent to the noise filtering procedures discussed in Chapter II. In the fourth section a second, improved, approximate formulation employing a weighted information measure is developed;² some important

¹Numerical solution may be feasible in some cases but this is not investigated in the present work.

²This weighted information formulation assumes a central role in this work. In fact, this was the original foundation and was developed heuristically following the work of Chu and Messerschmitt [55, 56]. The theoretical foundation (as an approximation to the "ideal" formulation) was subsequently developed because the heuristic development could only specify the weight function qualitatively and a more quantitative characterization was required.

properties of the weighted information measure are derived in the fifth section. Finally, the last section reflects upon these formulations, their relationship to other estimation procedures, and problems of spectral estimation and speech analysis to which they may be applied.

Application to Speech Analysis

Acoustic events in speech are often modeled as a white zero-mean Gaussian stationary excitation of a linear system. The linear system response is usually identified with the vocal cavity response which depends upon the position of speech articulators (tongue, lips, teeth, etc.); the excitation is usually assumed to be physically localized although its position may vary with different speech events.

The linear system model may be criticized in various ways; still it has had considerable success in practical situations. The particular case of an AR (or all-pole linear) system model can be justified on the basis of a lossless acoustic tube of varying cross-sectional area. The analogy of an acoustic tube with the oral or nasal cavity alone is clear; however, some speech sounds reflect the combined response characteristics of the oral and nasal cavities indicating that a full ARMA model would be more appropriate. A more complete discussion of acoustic tube modeling of the vocal tract may be found in [21].

Based upon the considerable success of AR models in speech applications, as well as the physical analogies that

may be drawn between AR models and the vocal tract via acoustic tube modeling, the AR speech model is adopted here. In most applications the deleterious effects of the pressure transducer, analog amplifier, anti-aliasing prefilter, and the digitizer have been carefully minimized and may be ignored. Some applications permit the system designer to ensure that the pressure transducer response reflect only the speech of the intended speaker; more often, conflicting goals deny the designer this flexibility so that the microphone transduces other ambient environmental acoustic events that appear as unwanted "noise" in the observed signal. Consequently, while the AR model is adopted for the speech spectrum, it is inadequate as a model for the observed sig-

Some ambient noise is a direct environmental response to the speech itself (e.g. echoes) or is short, transient, and generally unpredictable by nature (e.g. a gunshot, dropped book, engine backfire, cough, etc). Other ambient noise is repetitive (e.g. machine-gun fire) or steady by nature (e.g. drone of engines, rushing air, running water, whine of a turbine). This last (steady) type of noise is the primary focus of many speech analysis systems; typically these systems exploit the steady nature of the noise to determine noise statistics during speech activity from signal observations made during speech inactivity. With multiple transducers (or other clever system design techniques) the statistics of a much broader class of noises may be

known during speech activity. In the following it is only assumed that, during each analysis interval, the noise in the primary (objective) observation signal be zero-mean Gaussian stationary additive and independent of the speech; the noise is, therefore, completely characterized by a spectral density function, $\mu(\theta)$, which is assumed to be known.

The goals of speech analysis are many and varied. In communications the goal is often to achieve a minimal data rate subject to a quality or communicability constraint. In artificial intelligence the goal is usually to "understand" the speech with phonetic or written transcription often arising as an intermediate step. Some other goals include the identification of the speaker, the identification of the language, translation of the voice of one speaker to that of another in the same or a different language, and the screening/diagnosis of disease (e.g. laryngeal cancer). Spectral estimation is at the foundation of speech analysis for all these goals and accurate AR model estimation in noise is fundamental to the estimation of speech spectra in practical environments.

Ideal Formulation

Let

 $h(\theta) = g(\theta) + \mu(\theta)$ (3.1)

be the observed process power spectral density model where $\mu(\theta)$ is the known additive noise process spectrum and $g(\theta)$ is the unknown AR(P) power spectral density model characterizing the signal process; see Equation (2.3). Let $f(\theta)$ be the Schuster periodogram defined for the N time series observations by Equation (2.1). If the signal and noise processes are independent zero-mean real stationary Gaussian processes then the maximum likelihood method is asymptotically equivalent, for large N, to minimizing I(f,h) with respect to the AR(P) process parameters. Any parameter set minimizing I(f,h) and corresponding to a stable AR(P) process shall be considered here to be an ideal solution to the estimation problem.

This formulation of the estimation problem as a minimization problem may also be derived from an information theoretic viewpoint. Let $\tilde{f}(\theta)$ be the true observed process power spectral density so that $I(\tilde{f},h)$ represents the asymptotic information divergence between the true spectrum and an arbitrary model spectrum. Clearly it is desirable to find the model $h(\theta)$ minimizing $I(\tilde{f},h)$; if the minimum value is zero then $h(\theta) = \tilde{f}(\theta)$ almost everywhere. Since $\tilde{f}(\theta)$ is unavailable, replace it by a rough estimate, $f(\theta)$, and find $h(\theta)$ to minimize I(f,h).

Minimization of I(f,h) is subject to several interesting interpretations; the maximum likelihood and minimum information divergence interpretations have been given

above, a third noise filtering interpretation is now provided. Notice that I(f,h) = I(Hf,g) where $H(\theta)$ is the frequency response of the Wiener filter given in Equations (2.16). The quantity $H(\theta)f(\theta)$ may be interpreted as a rough estimate of the spectrum of a process obtained by passing the observed process through a filter whose power³ spectral response characteristic is $H(\theta)$; minimization of I(f,h) =I(Hf,g) may then be understood as a standard LP (or maximum entropy, etc.) fit to the noise filtered process. Of course, $H(\theta)$ is not known but is a function of the unknown parameters of $g(\theta)$; one must simply imagine finding a parameter set defining $H(\theta)$ that also corresponds to the best LP fit, $g(\theta)$, to the output process.

The functional I(f,h) is minimized by computing its derivative with respect to each parameter of $g(\theta)$ and setting the result to zero. For an arbitrary parameter, ξ , this is

 $\int_{-\pi}^{\pi} [[H(\theta) g(\theta) - H^{2}(\theta) f(\theta)]/g^{2}(\theta)] (\partial g(\theta)/\partial \xi) d\theta/2\pi = 0$ (3.2)

³This is not to say that the observed process is passed through a Wiener filter whose <u>frequency</u> response is $H(\theta)$. Recall that the Wiener filter is designed to minimize the mean-square prediction error; the output process doing this does not have the signal process spectrum, $g(\theta)$, but instead the spectrum $H(\theta)g(\theta)$. Alternatively, $H(\theta)f(\theta)$ may be interpreted as a rough estimate of the cross-spectrum between the input and output processes of the Wiener filter.

Using Equations (2.3) and (2.4), the partial derivatives of $g(\theta)$ are

$$\partial g(\theta) / \partial \sigma^2 = g(\theta) / \sigma^2$$
 (3.3a)

and, for l = 1, 2, ..., P

$$\partial g(\theta) / \partial a_{\ell} = -g^2(\theta) \sum_{m=0}^{P} 2a_m \cos[(\ell-m)\theta]/\sigma^2$$
 (3.3b)

Defining⁴

$$W_n = \int_{-\pi}^{\pi} \{H^2(\theta) f(\theta) - H(\theta) g(\theta)\} e^{in\theta} d\theta/2\pi \qquad (3.4)$$

and substituting Equation (3.3) in Equation (3.2) yields

$$\sum_{m=0}^{P} (a_{m}/\sigma^{2}) \sum_{\ell=0}^{P} (a_{\ell}/\sigma^{2}) V_{\ell-m} = 0$$
 (3.5a)

and, for l=1, 2, ..., P

$$\sum_{m=0}^{P} (a_m / \sigma^2) V_{l-m} = 0$$
 (3.5b)

while a little further manipulation of Equations (3.5)

 $^{^{4}}$ It is worth noting that the quantities, V_n, defined by Equations (3.4) are the components of the gradient vector of I(Hf,g) where differentiation is defined with respect to the inverse correlation parametrization of g(θ); see Equation (3.22).

yields, for $l=0,1,\ldots,P$

$$\sum_{\mathbf{m}=0}^{\mathbf{P}} \mathbf{a}_{\mathbf{m}} \mathbf{v}_{\boldsymbol{\ell}-\mathbf{m}} = 0 \tag{3.6}$$

The symmetry of the functions $f(\theta)$, $g(\theta)$, and $H(\theta)$ may be used to demonstrate that $V_{-n} = V_n$ while it is easy to see that Equations (3.6) are satisfied if

$$V_n = 0$$
; n=0,1,...,P (3.7)

To show that Equations (3.7) must be satisfied if a stable filter is to be obtained, rewrite the system of Equations (3.6) in matrix form as

$$\left\{ \begin{bmatrix} 1 & a_{1} & \cdots & a_{P-1} & a_{P} \\ a_{1} & a_{2} & \cdots & a_{P} & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{P-1} & a_{P} & \cdots & 0 & 0 \\ a_{P} & 0 & \cdots & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{P-2} & \cdots & 1 & 0 \\ 0 & a_{P-1} & \cdots & a_{1} & 1 \end{bmatrix} \right\} \begin{bmatrix} v_{0} \\ v_{1} \\ \vdots \\ v_{P-1} \\ v_{P} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$
(3.8)

The coefficients of a stable P-1st order predictor $\{\hat{a}_n; n = 1, 2, ..., P-1\}$ are given recursively in terms of a stable Pth order predictor according to

$$\begin{bmatrix} I + k_{p} J \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ a_{1} \\ \vdots \\ a_{p-1} \\ a_{p} \end{bmatrix} = \begin{bmatrix} 1 \\ \hat{a}_{1} \\ \vdots \\ \hat{a}_{P-1} \\ 0 \end{bmatrix}$$
(3.9)

where 1 is the identity matrix, J is the reversal matrix

$$J = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix}$$
(3.10)

 $k_p = a_p$ is a reflection coefficient⁵ and

$$[I + k_P J]^{-1} = [I - k_P J]/(1 - k_P^2)$$
(3.11)

Applying the nonsingular transformation⁶ $[I + k_P J]^{-1}$ to Equation (3.8) does not change the solution and yields

$$\left\{ \begin{bmatrix} 1 & \hat{a}_{1} & \cdots & \hat{a}_{P-1} & 0 \\ \hat{a}_{1} & \hat{a}_{2} & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ \hat{a}_{P-1} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & \hat{a}_{P-2} & \cdots & 1 & 0 \\ 0 & \hat{a}_{P-1} & \cdots & \hat{a}_{1} & 1 \end{bmatrix} \right\} \begin{bmatrix} V_{0} \\ V_{1} \\ \vdots \\ V_{P-1} \\ V_{P} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$
(3.12)

⁵These are the same reflection coefficients used in the forward-backward recursion; see Equation (2.11).

⁶Bounded input, bounded output (BIBO) stability requires and is guaranteed by the condition $|k_n| < 1$ for $n = 1, 2, \ldots, P$ which also guarantees that the indicated transformation is nonsingular.

The last equation shows V_P to be a linear combination of $V_0, V_1, \ldots, V_{P-1}$ and the reduced system

$$\left\{ \begin{bmatrix} 1 & \hat{a}_{1} & \cdots & \hat{a}_{P-1} \\ \hat{a}_{1} & \hat{a}_{2} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \hat{a}_{P-1} & 0 & \cdots & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & \hat{a}_{P-1} & \cdots & 1 \end{bmatrix} \right\} \begin{bmatrix} V_{0} \\ V_{1} \\ \vdots \\ V_{P-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$
(3.13)

is of the same form as Equation (3.8). Consequently, stability requires that each V_n be a linear combination of the previous V_l , $l=0,1,\ldots,n-1$, while the final reduced system is simply $V_0 = 0$. Hence, if only stable minima of I(f,h) are sought these minima must satisfy Equations (3.7) which may be rewritten, for n=0,1,...,P, as

$$\int_{-\pi}^{\pi} H(\theta) H(\theta) f(\theta) e^{in\theta} d\theta/2\pi = \int_{-\pi}^{\pi} H(\theta) g(\theta) e^{in\theta} d\theta/2\pi \quad (3.14)$$

This is a highly complicated nonlinear system of equations that appears to be very difficult to solve analytically. Note that, in the absence of noise, $\mu(\theta) = 0$ and $H(\theta) = 1$ so that the system reduces to Equations (2.13) as expected; in this case it is well known that the system always possesses a unique stable solution.

In general no admissable solution exists; the following example will serve to illustrate. Consider an AR(0) process corrupted by white noise of known variance μ . The system of equations reduces to

$$\mathbf{r}_{0} = \int_{-\pi}^{\pi} \mathbf{f}(\boldsymbol{\theta}) \, d\boldsymbol{\theta}/2\pi = \sigma^{2} + \mu \qquad (3.15)$$

If $r_0 \ge \mu$ the system is solved by $\sigma^2 = r_0 - \mu$ which yields the minimum value $I(f,h) = I(f,r_0) = 0$. If $r_0 < \mu$ the system does not possess a real solution; however, I(f,h) is always minimized by selecting $\sigma^2 = r_0 \div \mu$.

Noise Filtering Formulation

Since Equations (3.14) appear so difficult to solve, it is natural to consider alternate formulations. From the observation that I(f,h) = I(Hf,g) and the interpretation of $H(\theta)$ as the power spectral response of a noise filter a simple and reasonable procedure is to replace $H(\theta)$, which depends upon unknowns, by an estimate $\hat{H}(\theta)$. Several classes of estimates have been presented in Equations (2.19) and (2.20).

Once the data has been processed by the filter with power response $\hat{H}(\theta)$ a "noise-free" rough estimate is available

$$\hat{\mathbf{f}}(\boldsymbol{\theta}) = \hat{\mathbf{H}}(\boldsymbol{\theta}) \, \mathbf{f}(\boldsymbol{\theta}) \tag{3.16}$$

Then, minimization of $I(\hat{f},g) = I(\hat{H}f,g)$ is achieved by the

solution to the equations

$$\int_{-\pi}^{\pi} \hat{H}(\theta) f(\theta) e^{in\theta} d\theta/2\pi = \int_{-\pi}^{\pi} g(\theta) e^{in\theta} d\theta/2\pi$$
for n=0,1,...,P (3.17)

This, of course, leads easily to the Yule-Walker equations with the difference that the estimated correlation values are now given by the left-hand side of Equation (3.17); the reader is urged to compare this equation with Equations (3.14) and (2.13).

Weighted Information Formulation

The previous approximate formulation encompasses a wide variety of estimation procedures that have been studied in recent years. If $\hat{f}(\theta)$, given by Equation (3.16), is a good rough estimate of the noise-free power spectral density the resultant model parameters can be expected to be accurate. Consequently, considerable effort has been expended trying to find the best form of $\hat{H}(\theta)$ and, ultimately, the best means of computing the correlation values on the left hand side of Equation (3.17).

Generally speaking, any estimate can be expected to be more accurate if there is less corrupting noise; in particular, $\hat{f}(\theta)$ can be expected to be more accurate in those spectral regions where the signal to noise density ratio is large. Since the reliability of the rough estimate $\hat{f}(\theta)$ varies with frequency, the criteria for fitting a model to $\hat{f}(\theta)$ should reflect this variation in reliability. The frequency weighted spectral distance measure introduced by Chu and Messerschmitt [55, 56] provides precisely the required flexibility for such a criteria. The criteria is derived from the asymptotic information divergence, I(f,g), by noting that the integrand in Equation (2.17) is a nonnegative error measure; the frequency weighted variant is obtained by introducing a multiplicative nonnegative weight function to the integrand of I(f,g) to yield

$$I_{W}(f,g) = \int_{-\pi}^{\pi} W(\theta) \{ [f(\theta)/g(\theta)] - \ln[f(\theta)/g(\theta)] - 1 \} d\theta/2\pi$$
(3.18)

If $W(\theta)$ is constant, minimization of $I_W(\hat{f},g) = I_W(\hat{H}f,g)$ is equivalent to minimization of $I(\hat{f},g) = I(\hat{H}f,g)$. To reflect the greater reliability of $\hat{f}(\theta)$ in some spectral regions, $W(\theta)$ should be selected to be large where the signal to noise density ratio is large. To remain consistent with AR estimation procedures that work well in the absence of noise, $\hat{H}(\theta)$ should approach unity and $W(\theta)$ should approach a constant as $\mu(\theta)$ approaches zero. Specific procedures for selecting $\hat{H}(\theta)$ have been studied in the past [32-39] and important examples are given in Equations (2.19) and (2.20); the above considerations provide a qualitative understanding of an appropriate selection for $W(\theta)$ but a more specific, quantitative understanding is required. To minimize $I_W(\hat{H}f,g)$ Equation (3.18) is differentiated with respect to the parameters of $g(\theta)$ and the results are set to zero. The procedure is the same, mutatis mutandis, as that followed for minimizing I(Hf,g) and yields the system of equations

$$\int_{-\pi}^{\pi} \mathbb{W}(\theta) \ \hat{H}(\theta) \ f(\theta) \ e^{in\theta} \ d\theta/2\pi = \int_{-\pi}^{\pi} \mathbb{W}(\theta) \ g(\theta) \ e^{in\theta} \ d\theta/2\pi \quad (3.19)$$

Comparison of Equations (3.19) to Equations (3.14), which result from the ideal formulation, immediately suggests the required quantitative criteria for selecting $W(\theta)$. Specifically, $W(\theta)$ should be selected so that, at least approximately,

$$W(\theta) = \hat{H}(\theta) \tag{3.20}$$

and $\hat{H}(\theta)$ should estimate $H(\theta)$. This selection is supported by the previous heuristic considerations which indicated that $W(\theta)$ should be large where the signal to noise density ratio is large.

Properties of the Weighted Information

In this section three important results concerning the weighted information measure, $I_W(f,g)$, are developed. These results also apply to the asymptotic information divergence, I(f,g), as a special case where $W(\theta) = 1$. The first result generalizes the triangle equality property for I(f,g), see

Equation (2.21); that this property generalizes appropriately is of interest to the use of the weighted information measure in place of the (unweighted) asymptotic information divergence for vector quantization.

The Kullback information number and the asymptotic information divergence are well known to be convex with respect to general classes of probability and spectral densities. With the appropriate definition for convex superposition of AR(P) spectra, the second important result is that the class of stable AR(P) spectra is convex and the weighted information measure is strictly convex with respect to this class.⁷ As a consequence, $I_W(\hat{H}f,g)$ can have at most one local minimum with respect to this class; moreover, if such a minimum exists it is also a global minimum.

Finally, the third result shows that the second mixed partial derivative of $I_W(\hat{H}f,g)$ defines a positive definite quadratic form. This shows that any stable solution to Equation (3.19) is a local minimum of $I_W(\hat{H}f,g)$; this could also have been demonstrated using the strict convexity. Combined with the previous result this shows that Equation

A set, \mathscr{P} , is convex if it always contains the convex superposition of two elements in the set. A convex superposition is a map $x_3 = CS(x_1, x_2; Y)$ defined for $0 \le Y \le 1$ and all $x_1, x_2 \in \mathscr{P}$ such that $x_3 = x_1$ if Y = 1 and $x_3 = x_2$ if Y = 0; if $x_1 = x_2$ then $x_3 = x_2 = x_1$ for all Y. A function f(x) defined on a convex set \mathscr{P} is said to be convex if $Yf(x_1) + (1-Y) f(x_2) \ge f(x_3)$ and strictly convex if equality implies $x_1 = x_2$.

(3.19) can have at most one stable solution (although unstable solutions can, and often do, exist); moreover, if such a solution exists, it is the global minimum among stable AR(P) spectral models.

The question of existence is not addressed in this set of results. The existence of a stable solution to Equations (3.19) is assumed but remains an open question in general; existence can be demonstrated in special cases, e.g. $W(\theta) =$ 1, while experimental results are discussed in Chapter V. Because the proofs are nonconstructive, they do not assist with the question of existence nor do they provide algorithms for computation of a solution; computational procedures are discussed in Chapter IV. It is worth noting that if no solution to Equations (3.19) exists then, since $I_W(\hat{H}f,g)$ must possess a minimum in the closure of the set of stable AR(P) spectra, the minimum occurs as a limit point of the set.

To simplify the following discussion the set of stable AR(P) spectra shall be denoted \mathscr{R}_P . Each element of the set may be characterized by a P+1-tuple of real parameter values satisfying appropriate (stability) criteria. Four characterizations of \mathscr{R}_P are presented below:

<u>Predictor Coefficients</u>. Let $A_p(z)$ be given by Equation (2.4) with all roots of $A_p(z)$ inside the unit circle. Then $(\sigma, a_1, a_2, \ldots, a_p)$ denotes an arbitrary element of \mathscr{R}_p if $\sigma > 0$. <u>Reflection Coefficients</u>. Let $A_p(z)$ be given by Equation (2.11) with $|k_n| < 1$ for $n=1,2,\ldots,P$. Then $(\sigma,k_1,k_2,\ldots,k_p)$ denotes an arbitrary element of \mathscr{R}_p if $\sigma > 0$.

<u>Autocorrelation Coefficients</u>. Let the real symmetric Töeplitz quadratic form given by

$$T(\bar{x}) = \sum_{m,n=0}^{P} r_{m-n} x_m x_n$$
 (3.21)

be positive definite. Then (r_0, r_1, \ldots, r_P) denotes an arbitrary element of \mathscr{R}_P .

Inverse Correlation Coefficients. Let

$$1/g(\theta) = \sum_{n=-P}^{P} u_{|n|} e^{in\theta}$$
(3.22)

be a positive function of 6 in $[-\pi,\pi)$. Then (u_0,u_1,\ldots,u_P) denotes an arbitrary element of \mathscr{R}_P .

These represent only a few of the infinitely many ways of characterizing \mathscr{R}_{p} . The first three parametrizations are well known with the corresponding terminology well established in the literature. Each set of predictor coefficients is related to a unique set of reflection coefficients by a continuous bijection defined by the Levinson-Durbin recursion. Each set of autocorrelation coefficients defines a unique set of predictor coefficients according to the Yule-Walker equations while the autocorrelation coefficients may be retrieved from the predictor coefficients using Equations (2.3) and (2.8).

The last parametrization is less common than the other three; these parameters have been denoted "inverse correlation coefficients" since they are the autocorrelation coefficients of a moving-average process whose spectral density function is inverse to that of the defined AR(P) spectrum. Each set of predictor coefficients uniquely defines the inverse correlation coefficients according to

$$u_n = \sum_{m=0}^{P-n} a_m a_{m+n} / \sigma^2$$
; $a_0 = 1$; $n=0,1,\ldots,P$ (3.23)

That the predictor coefficients may be retrieved in a unique fashion from the inverse correlation coefficients is more difficult to establish. Positivity of Equation (3.22) generally establishes only the possibility of several appropriate predictor coefficient sequences; closer inspection reveals that only one of these sequences satisfies the stability requirements. The question is taken up in somewhat greater detail by Blackman and Tukey [5, pp. 126-7].

The first result follows easily using the inverse correlation coefficient parametrization of the AR(P) spectral density, Equation (3.22), together with Equations (3.19) and (3.18).

<u>Theorem 3-1</u>. (Triangle Equality). Let $g_1(\theta)$ be an AR(P) spectral density satisfying Equation (3.19) and let $g_2(\theta)$ be any other AR(P) spectral density. Then

$$I_{W}(\hat{H}f,g_{2}) = I_{W}(\hat{H}f,g_{1}) + I_{W}(g_{1},g_{2})$$
 (3.24)

The inverse correlation coefficient parametrization of AR(P) models in \Re_P is used here to define the convex superposition of two models according to

$$\overline{u}_3 = CS (\overline{u}_1, \overline{u}_2; Y) = Y \overline{u}_1 + (1 - Y) \overline{u}_2$$
(3.25)

for $0 \le Y \le 1$. Since (3.22) remains a strictly positive function for \overline{u}_3 when \overline{u}_1 and \overline{u}_2 define strictly positive functions, this shows \mathscr{R}_p to be a convex set and leads to the second result.

Lemma 3-1. (Strict Convexity). Let $g_3(\theta)$ be a stable AR(P) spectrum defined by the convex superposition of the two stable AR(P) spectra $g_1(\theta)$ and $g_2(\theta)$. Then

$$I_{W}(f,g_{3}) \leq YI_{W}(f,g_{1}) + (1-Y)I_{W}(f,g_{2})$$
 (3.26)

for $0 \le Y \le 1$ with equality only if $g_1(\theta) = g_2(\theta)$.

<u>Proof.</u> Using the inverse correlation coefficient parametrization and the definition of convex superposition for AR(P) spectra it is easy to show that

$$g_{3}(\theta) = 1/\{[Y/g_{1}(\theta)] + [(1-Y)/g_{2}(\theta)]\}$$
(3.27)

Together with Equations (3.18) this yields

$$I_{W}(f,g_{1}) + (1-Y) I_{W}(f,g_{2}) - I_{W}(f,g_{3})$$

=
$$\int_{-\pi}^{\pi} W(\theta) \ln\{[g_{1}(\theta)]^{Y} [g_{2}(\theta)]^{1-Y}/g_{3}(\theta)\} d\theta/2\pi \quad (3.28)$$

From the theorem on geometric and harmonic means the argument of the logarithm in Equation (3.28) is not less than one and equals one only if $g_1(\theta) = g_2(\theta)$. The lemma follows easily.

<u>Theorem 3-2</u>. (Uniqueness). $I_W(f,g)$ can have at most one local minimum in \mathscr{R}_P ; if such a minimum exists it is also a global minimum.

<u>Proof</u>. Let $g_1(\theta)$ and $g_2(\theta)$ be two distinct local minima and form their convex superposition $g_3(\theta)$. Without loss of generality assume $I_W(f,g_1) \ge I_W(f,g_2)$. With $Y \ne 1$ the previous lemma gives

 $I_{W}(f,g_{3}) < \gamma I_{W}(f,g_{1}) + (1-\gamma) I_{W}(f,g_{2}) < I_{W}(f,g_{1})$ (3.29)

But $g_3(\theta)$ is arbitrarily close⁸ to $g_1(\theta)$ for Y arbitrarily

⁸The Euclidean metric applied to the inverse correlation coefficients shall suffice to define closeness here.

close to one, so that this inequality contradicts the assumption that $g_1(\theta)$ is a local minimum. The second part of the theorem follows by assuming $g_1(\theta)$ is a local minimum while $g_2(\theta)$ is any distinct element of \mathscr{R}_P such that $I_W(f,g_2) \leq I_W(f,g_1)$ and then repeating the above argument.

In order to establish the final theorem of this section the second mixed partial derivative of $I_W(\hat{H}f,g)$ is shown to define a positive definite quadratic form. The variables

$$\mathbf{v}_{n} = \begin{cases} \mathbf{u}_{0} & \text{for } n=0 \\ 2\mathbf{u}_{n} & \text{for } n\neq0 \end{cases}$$
(3.30)

are defined for n=0,1,...,P so that the first partial derivatives are

$$\hat{f}_{n} = \partial I_{W}(\hat{H}f,g) / \partial v_{n}$$

$$= \int_{-\pi}^{\pi} W(\theta) \{\hat{H}(\theta) \ f(\theta) - g(\theta)\} \cos(n\theta) \ d\theta / 2\pi \qquad (3.31)$$

and the second mixed partial derivatives are

$$L_{nm} = \partial \hat{V}_{n} / \partial v_{m}$$

=
$$\int_{-\pi}^{\pi} W(\theta) [g(\theta)]^{2} \cos(n\theta) \cos(m\theta) d\theta / 2\pi \qquad (3.32)$$

Clearly, the quadratic form

$$\sum_{m,n=0}^{P} x_n x_m L_{nm} = \int_{-\pi}^{\pi} W(\theta) [g(\theta)]^2 \left[\sum_{n=0}^{L} x_n \cos(n\theta) \right]^2 d\theta/2\pi$$
(3.33)

is positive definite. This proves the following

<u>Theorem 3-3</u>. (Absence of False Solutions). Any stable AR(P) solution to Equations (3.19) is a local minimum.

Note that this does not eliminate the possibility of unstable solutions to Equations (3.19), nor does it establish the existence of a stable solution. Since the previous theorem has established the uniqueness of a minimum this theorem establishes the

<u>Corrollary 3-1</u>. Equations (3.19) can have at most one stable AR(P) solution. If such a solution exists it is the unique absolute minimum of $I_W(\hat{H}f,g)$ over \mathscr{R}_P .

Remarks

Three general formulations for estimating the parameters of an AR(P) process in noise have been discussed. The first "ideal" formulation has theoretical foundations resting upon principles of information theory as well as the maximum likelihood method. The second two formulations are developed as approximations to the first.

The need for approximate formulations arises due to the difficulty posed by the nonlinear equations resulting from the ideal formulation. The first approximate formulation leads to the Yule-Walker equations but with modified correlation values; algorithms for solving the Yule-Walker equations are computationally simple and well understood while methods for evaluating the modified correlation values have been carefully studied in recent years.

While this first, noise filtering, approach has led to demonstrable performance improvements in noise environments over the standard noise free formulation (and reduces to the noise free formulation in noise free environments), still better performance is desired. Rather than attempt direct solution of the ideal formulation the second approximate formulation is developed. Evidence that this weighted information formulation leads to improved performance over the noise filtering formulation is presented in Chapter V; neither approximate formulation is expected to perform as well as the "ideal" formulation.

The weighted information formulation is related to other techniques that have appeared in the literature. Consider the situation wherein the desired signal spectrum is essentially zero outside the region $\theta \in [-\pi/\mathfrak{D}, \pi/\mathfrak{D})$ while the noise spectrum is essentially zero inside this region. The foregoing theory indicates that an appropriate selection for the weight function is

 $W(\Theta) = \hat{H}(\Theta) = \begin{cases} 1 & \Theta \in [-\pi/\mathfrak{D}, \pi/\mathfrak{D}) \\ 0 & \text{otherwise} \end{cases}$ (3.34)

so that the weighted information is

$$I_{W}(\hat{H}f,g) = \int_{-\pi/\mathfrak{D}}^{\pi/\mathfrak{D}} \{ [f(\theta)/g(\theta)] - \ln[f(\theta)/g(\theta)] - 1 \} d\theta/2\pi \quad (3.35)$$

With the change of variable $\tilde{\Theta}/\mathfrak{D} = \Theta$ this may be rewritten

$$I_{W}(\hat{H}f,g) = (1/\mathfrak{D}) \int_{-\pi}^{\pi} \{ [f(\tilde{\theta}/\mathfrak{D})/g(\tilde{\theta}/\mathfrak{D})] - \frac{1}{d\tilde{\theta}/2\pi}$$
(3.36)

Clearly the indication here is to low pass filter and decimate the observed signal before fitting the AR(P) model to the resulting data. This is precisely the technique employed by Quirk and Liu [45] to improve the resolution of AR(P) estimation in noise; they considered the use of AR(P)estimators to determine the frequencies of sinusoids in noise and demonstrated that the filtering/decimation scheme is clearly advantageous when the sinusoids are a priori known to lie in some fixed frequency range.

The problem which motivates the present work concerns signal and noise spectra that are both generally nonzero throughout the entire frequency range, $[-\pi,\pi)$; hence the luxury of simple filtering/decimation schemes is not permitted. On the other hand, the difficulties associated with very limited quantities of data are not the primary focus of this work so that the asymptotic formulation is considered adequate.

Computational issues for the weighted information formulation are discussed in Chapter IV. Equations (3.19) are cast in algebraic form and their (exact) analytical solution is discussed. Approximate (numerical) solution methods might be developed based upon the resulting analytical system of equations or directly upon minimization of $I_{W}(\hat{H}f,g)$; the latter approach is adopted to develop a simple iterative procedure based upon the notion of a contraction mapping. In addition, computational procedures appropriate to the use of the weighted information for vector quantization are discussed. Since in many applications the "vector quantization codebook" may be designed "off-line" using noise free speech data, questions associated with the codebook design problem are not discussed; instead, computational procedures for the "on-line" minimization of $I_{W}(\hat{H}f,g)$ over the finite codebook are developed.

CHAPTER IV

COMPUTATIONAL FORMULATION

In this chapter computational procedures for the solution of Equations (3.19) are discussed. In the first section the system is reduced to an algebraic form by assuming the weight function to take the form of an AR(M) power spectral density; once cast as a nonlinear algebraic system of equations, analytic procedures for solving the system are discussed. In the second section, techniques for evaluating the coefficients of the system are discussed.

Analytic solution of the nonlinear algebraic system becomes increasingly difficult as the order of the weight function, M, is increased. While numerical polynomial root solving procedures could be systematically applied, the third section develops instead an iterative procedure based upon the idea of a contraction mapping. Together with sampled frequency domain processing techniques, these iterative procedures do not restrict the weight function to an all-pole form. The fourth section develops computational formulae required for the use of the weighted information in vector quantization; an extension of Jensen's theorem is developed to permit closed form evaluation of the appropriate integrals when the weight function assumes an AR(M) form. Finally, the last section concludes this chapter with some final remarks concerning these computational methods.

Reduction to Algebraic Form

Let

$$P_{n} = \int_{-\pi}^{\pi} W(\theta) \hat{H}(\theta) f(\theta) e^{in\theta} d\theta / 2\pi$$

$$n = 0.1...P$$
(4.1)

denote the coefficients appearing on the left hand side of Equations (3.19). Let

$$\hat{\rho}_{n} = \int_{-\pi}^{\pi} \mathbb{W}(\theta) \ g(\theta) \ e^{in\theta} \ d\theta/2\pi \qquad (4.2)$$

n = 0, 1, ..., P+M

denote the quantities appearing on the right hand side of Equations (3.19). Observe that the index of $\hat{\rho}_n$ is permitted to range beyond P to P+M. If W(0) is an AR(M) spectrum given by

$$W(\theta) = \sigma_{W}^{2} / [B_{M}(e^{i\theta}) B_{M}(e^{-i\theta})]$$
(4.3a)

where

$$B_{M}(z) = \sum_{m=0}^{M} b_{m} z^{-m}; b_{0} = 1$$
 (4.3b)

and if $g(\theta)$ is an AR(P) spectrum given by Equations (2.3) and (2.4) then their product is an AR(P+M) spectrum given by

$$W(\theta) g(\theta) = \sigma_{W}^{2} \sigma^{2} / [C_{P+M}(e^{i\theta}) C_{P+M}(e^{-i\theta})]$$
(4.4a)

where

$$C_{P+M}(z) = A_{P}(z) B_{M}(z) = \sum_{m=0}^{P+M} c_{m} z^{-m}; c_{0} = 1$$
 (4.4b)

The quantities defined by Equation (4.2) are related to the polynomial coefficients in Equation (4.4b) by the Yule-Walker equations

Equations (3.19) assign numerical values to some of the entries in the coefficient matrix according to

 $\hat{\rho}_n = \rho_n ; n=0,1,...P$ (4.6)

while the remaining entries are to be considered as unknowns. The elements of the column vector are defined as a linear combination of the coefficients of the unknown polynomial, $A_p(z)$, by Equation (4.4b) which may be rewritten in matrix form as

$$\begin{bmatrix} 1 \\ c_1 \\ \vdots \\ \vdots \\ c_{P+M} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ b_1 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & b_M \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ \vdots \\ a_P \end{bmatrix}$$
(4.7)

Equations (4.5), (4.6), and (4.7) define a nonlinear system of P+M+1 multivariate polynomials in the P+M+1 unknowns σ , a_1 , a_2 , ..., a_p , $\hat{\rho}_{p+1}$, $\hat{\rho}_{p+2}$, ..., $\hat{\rho}_{p+M}$. Each polynomial is a first order function of each unknown while each term in these polynomials may involve up to two distinct unknowns. The properties of the weighted information developed in Chapter III indicate that this system of equations can have at most one stable solution; if a stable solution exists it is the solution sought.

Assuming the AR(M) weight function to be stable the product polynomial, $C_{P+M}(z)$, also has all its roots inside the unit circle and may be expressed recursively in terms of a set of reflection coefficients according to

$$C_n(z) = C_{n-1}(z) + k_n z^{-n} C_{n-1}(z^{-1}); C_0(z) = 1$$
 (4.8)

for n = 1, 2, ..., P+M. If the coefficient matrix in Equation (4.5) were entirely known then the Levinson-Durbin recursion¹ could be applied to yield $C_{P+M}(z)$. Since some of the entries in the coefficient matrix are unknown, the Levinson-Durbin recursion cannot proceed beyond the determination of $C_{p}(z)$; the remaining reflection coefficients $\{k_{P+1}, k_{P+2}, ..., k_{P+M}\}$ are unspecified (beyond the stability requirement that $|k_n| < 1$) by Equations (4.5) and may be considered as new unknowns replacing $\{\hat{P}_{P+1}, \hat{P}_{P+2}, ..., \hat{P}_{P+M}\}$.

These remaining reflection coefficients should be selected so that $C_{P+M}(z) = 0$ modulo $B_M(z)$. Once these have been determined the solution may be obtained by simple polynomial division from

$$A_{p}(z) = C_{p+M}(z)/B_{M}(z)$$
 (4.9)

together with

$$\sigma^{2} = (\rho_{0} / \sigma_{W}^{2}) \prod_{n=1}^{P+M} (1 - k_{n}^{2})$$
(4.10)

To determine the remaining reflection coefficients it is generally simpler to consider the polynomials

¹This well-known algorithm may be found in many fairly recent publications; for example, see [21, p. 55ff]. An exposition by the authors is contained in [57] and [58].

$$\widetilde{C}_{P+M}(z) = z^{-(P+M)} C_{P+M}(z^{-1})$$
(4.11)

$$\tilde{B}_{M}(z) = z^{-M} B_{M}(z^{-1})$$
 (4.12)

so that the condition to be satisfied is

$$\widetilde{C}_{P+M}(z) = 0 \mod \overline{B}_{M}(z)$$
 (4.13)

Modulo reduction is then accomplished more simply by repeated use of the substitution

$$z^{-M} = -\sum_{\ell=0}^{M-1} b_{M-\ell} z^{-\ell}$$
(4.14)

in $\widetilde{C}_{P+M}(z)$ until all powers of z^{-1} larger than M-1 have been eliminated. The reduction process is facilitated by using the recursion (4.8) to express $\widetilde{C}_{P+M}(z)$ as

$$\widetilde{C}_{P+M}(z) = \widetilde{C}_{P}(z) \ \widetilde{E}_{M}(z) + z^{-P} \ \widetilde{C}_{P}(z^{-1}) \ \widetilde{F}_{M}(z)$$
(4.15)

where

$$\widetilde{E}_{n}(z) = z^{-1} \widetilde{E}_{n-1}(z) + k_{p+n} z^{-(n-1)} \widetilde{F}_{n-1}(z^{-1}); \widetilde{E}_{0}(z) = 1$$
 (4.16a)

$$\widetilde{F}_{n}(z) = z^{-1} \widetilde{F}_{n-1}(z) + k_{P+n} z^{-(n-1)} \widetilde{E}_{n-1}(z^{-1}); \widetilde{F}_{0}(z) = 0$$
 (4.16b)

$$\widetilde{C}_{p}(z) = z^{-P} C_{p}(z^{-1})$$
(4.17)

With these formulae the reduction is accomplished in part by determining

$$\widetilde{D}_{M-1}(z) = \widetilde{C}_{P}(z) \mod \widetilde{B}_{M}(z)$$
 (4.18a)

and

$$D_{M-1}(z) = C_P(z) \mod \widetilde{B}_M(z) = z^{-P} \widetilde{C}_P(z^{-1}) \mod \widetilde{B}_M(z)$$
 (4.18b)

The condition to be satisfied is then

$$\widetilde{D}_{M-1}(z) \widetilde{E}_{M}(z) + D_{M-1}(z) \widetilde{F}_{M}(z) = 0 \mod \widetilde{B}_{M}(z)$$
 (4.19)

Modulo reduction of the left-hand side of Equation (4.19) leads to an M-1st order polynomial whose M coefficients must be equated to zero; this yields a system of M nonlinear polynomial equations in the M unknowns $\{k_{P+1}, k_{P+2}, \ldots, k_{P+M}\}$. While these equations are nonlinear some reflection will reveal that each polynomial equation is linear (i.e., of first degree) in each of the unknowns; the nonlinearity enters by way of terms involving products of different unknowns.

and

Because of this structure, systematic algebraic elimination² will yield an Mth order polynomial in a single unknown; each acceptable root of this polynomial will yield an M-1st order polynomial in a second unknown. Continuing in this fashion one successively solves Mth, M-1st, ... order polynomial equations possibly generating M factorial potential solutions of which at most one satisfies the stability criteria. This method is feasible for small values of M (e.g. $M \leq 4$) but for larger values of M one must generally resort to numerical polynomial root solving procedures.³

For the case M=2, let

 $\tilde{D}_1(z) = \tilde{d}_0 + \tilde{d}_1 z^{-1}$ (4.20a)

 $D_1(z) = d_0 + d_1 z^{-1}$ (4.20b)

²Several methods (such as those due to Euler, Bezout, or Sylvester) are available; one should take care not to introduce extraneous roots. For a general discussion see [59, Vol. II, p. 70ff] or [60, p. 277ff].

⁹The recommendation that M not exceed four is made based upon the fact that general polynomial equations of degree five and higher cannot be solved algebraically [59, Vol. II, p. 286]. Of course this does not eliminate the possibility of transcendental solutions [59, Vol. I, p. 274] or the possibility that some special structure, may be discovered (or imposed) to aid in the solution.

and let

Ľ

1

$$G_{j}(z) = \widetilde{D}_{1}(z) \widetilde{E}_{2}(z) + D_{1}(z) \widetilde{F}_{2}(z) = \sum_{m=0}^{3} g_{m} z^{-m}$$
 (4.21)

denote the left hand side of Equation (4.19). Using Equations (4.16) these coefficients are

$$g_0 = d_0 k_{P+2}$$
 (4.22a)

$$g_1 = \tilde{d}_0 k_{P+1} k_{P+2} + d_0 k_{P+1} + d_1 k_{P+2}$$
 (4.22b)

$$\mathbf{g}_2 = \tilde{\mathbf{d}}_0 + \tilde{\mathbf{d}}_1 \mathbf{k}_{P+1} \mathbf{k}_{P+2} + \mathbf{d}_1 \mathbf{k}_{P+1}$$
(4.22c)

$$g_3 = \widetilde{d}_1 \tag{4.22d}$$

while modulo reduction yields

$$g_0 - b_2 g_2 + b_1 b_2 g_3 = 0$$
 (4.23a)

$$g_1 - b_1 g_2 + (b_1^2 - b_2) g_3 = 0$$
 (4.23b)

Expanding Equations (4.23) yields

$$p_0 k_{P+2} + p_1 = k_{P+1}(q_0 k_{P+2} + q_1)$$
 (4.24a)

$$\hat{p}_{0} k_{P+2} + \hat{p}_{1} = k_{P+1} (\hat{q}_{0} k_{P+2} + \hat{q}_{1})$$
 (4.24b)
where

4

1

$$\mathbf{p}_{0} = \mathbf{d}_{0} \tag{4.25a}$$

$$p_1 = \tilde{d}_1 b_1 b_2 - \tilde{d}_0 b_2$$
 (4.25b)

$$\hat{p}_0 = d_1$$
 (4.25c)

$$\hat{p}_1 = \tilde{d}_1(b_1^2 - b_2) - \tilde{d}_0 b_1$$
 (4.25d)

$$q_0 = \tilde{d}_1 b_2 \tag{4.25e}$$

$$q_1 = d_1 b_2$$
 (4.25f)

$$\hat{\mathbf{q}}_{0} = \tilde{\mathbf{d}}_{1} \mathbf{b}_{1} - \tilde{\mathbf{d}}_{0}$$
(4.25g)

$$\hat{q}_1 = d_1 b_1 - d_0$$
 (4.25h)

So that the solutions are given, upon elimination, by

$$k_{P+1} = (p_0 k_{P+2} + p_1) / (q_0 k_{P+2} + q_1)$$
(4.26)

and

$$k_{P+2} = [-s_1 + \sqrt{s_1^2 - 4s_0 s_2}]/2s_2$$
 (4.27)

where

$$s_0 = p_1 \hat{q}_1 - \hat{p}_1 q_1$$
 (4.28a)

$$s_1 = p_1 \hat{q}_0 - \hat{p}_1 q_0 + p_0 \hat{q}_1 - \hat{p}_0 q_1$$
 (4.28b)

$$s_2 = p_0 \hat{q}_0 - \hat{p}_0 q_c$$
 (4.28c)

Evaluation of Coefficients

There are numerous methods of coefficient evaluation that may be considered consistent with the foregoing formulation. Generally this section will present a few of the possibilities for evaluating the coefficients defined by Equation (4.1). In addition, some discussion will be devoted to characterizing the weight function according to Equations (4.3). While performance of the estimation procedure will undoubtedly depend upon the specific method of coefficient evaluation, no one method can be strongly advocated (i.e. to the total exclusion of other methods) at this time; in addition, final selection of a method may be influenced by other ancillary requirements of the specific Because no single explicit procedure is to be application. recommended here the discussion stresses concepts rather than detailed mathematical formulae.

Time domain noise filtering methods usually determine (adaptively) the coefficients of a finite impulse response (FIR) linear filter whose power spectral response is $\hat{H}(\theta)$.

By simply cascading two of these filters a new filter is created whose power spectral response is $\hat{H}(\theta)$ $\hat{H}(\theta)$. The coefficients in Equation (4.1) may then be computed in the usual manner (lag products of the windowed data) from the output of the cascaded filter structure. This scheme, depicted in Figure 2, assumes the relationship expressed by Equation (3.20) although this relationship may generally be avoided by replacing one of the filters in the cascade by a filter with $W(\theta)$ as its power spectral response. For each data window, a "snapshot" of the impulse response of the FIR filter could be used to estimate the parameters of $W(\theta)$. Since the response of the FIR filter may differ slightly from the response of the weight function a somewhat more consistent procedure would use the weight function parameters to implement an infinite impulse response (IIR) filter as the second filter in the cascade.

Frequency domain noise filtering methods generally provide greater flexibility in response function selection than is available with time domain methods. These methods involve an explicit transformation to the frequency domain, often by using the discrete Fourier transform (DFT), and determine the multiplicative response function, $\hat{H}(\theta)$, in sampled form using a formula such as Equation (...19) or (2.20). The sampled form of $\hat{H}(\theta)$ may be used to estimate the parameters of $W(\theta)$. If the noise filtered signal is not required, frequency samples of the weight function may be used multiplicatively before evaluating the coefficients;





alternatively, one may avoid re-evaluating the weight function and simply apply $\hat{H}(\theta)$ twice. This latter alternative is depicted in Figure 3.

A mixed time-frequency domain method is employed to obtain some of the results presented in Chapter V. In this method a Hamming window is applied to the observed data which is then zero-extended before computing the DFT. Α sampled noise spectrum estimate is used together with these transform values to compute a noise filter spectral response, $\hat{H}(\theta)$, according to Equations (2.19) or (2.20).⁴ This frequency sampled noise filter response is applied multiplicatively to the transform values and an inverse DFT of these modified transform values (with their original phase values) is computed. A random phase characteristic is computed and introduced to the frequency sampled noise filter spectral response which is inverse transformed to obtain an impulse response characteristic. Standard (autocorrelation method) LP analysis is applied to this impulse response characteristic to determine the parameters of the weight function. These parameters are used to implement a

⁴It is generally found to be useful to modify the frequency response characteristic slightly by smoothing the response obtained from (2.19) or (2.20) across frequency. The smoother should eliminate features narrower than those expected in the final signal spectrum while retaining broader features; a recursive median filter with a total length of about 2.5% of the single-sided bandwidth is a current favorite of this author. End conditions (near the DC and Nyquist frequencies) can be properly handled using the known periodic nature of the frequency response.



· · · · ·



.

(lattice structure) filter; beginning in the all-zero state the noise filtered (inverse transformed) data values are passed through this filter which is then permitted to "ring" awhile.⁵ Lag products computed from this output then provide the required coefficient estimates; the overall procedure is depicted in Figure 4.

Finally, it is worth mentioning that each of these methods has recommended computing the final coefficient estimates as lagged products. The reason for this is that various quantization effects may occur up to the point of obtaining the modified data samples; however, if full precision is maintained in the final lag product computations, the resulting coefficient estimates will define a positive definite symmetric Töeplitz quadratic form in all but a very few highly exceptional cases (such as all modified data samples being identically zero).

Iterative Techniques

Equations (3.19) may be solved when the weight function has an AR(M) form by using the algebraic procedures described in the first section of this chapter; this method is appropriate if $M \leq 4$. Unfortunately, it is expected that accurate estimation of speech spectra will require weight functions with greater variation than is possible with an

⁵That is to say that a zero input is applied to the filter after all the noise filtered data values have been applied as input.





AR(4) form. The procedures of the first section might be extended by applying numerical polynomial root solving procedures when M becomes large but at present such an approach appears somewhat cumbersome.⁶ In this section alternate numerical formulations are discussed that do not make specific (parametric) assumptions as to the form of the weight function; these techniques are iterative and based upon the notion of a contraction mapping. A good general reference for this section is Collatz [61].

Most (single-step) iterative procedures can be expressed in the form 7

 $\bar{\mathbf{v}}^{(n+1)} = \bar{\varphi}(\bar{\mathbf{v}}^{(n)})$

(4.29)

⁷Parenthesized superscripts shall denote instances of the parameter vector while subscripts shall denote components of the parameter vector.

⁶For the reader wishing to pursue this approach it is worth noting that one stumbling block is that the previous uniqueness theorem has not eliminated the possibility of an unstable (or imaginary) solution to Equations (4.5), (4.6), and (4.8) for which <u>some</u> (but not all) of the reflection coefficients are real and in the interval (-1, 1). If one could devise a method which guarantees that only the solution sought has real parameters isolated in (-1, 1), or some other known interval, the development of a numerical algorithm would be greatly facilitated. The reader is referred to [60, p. 99ff] or any similar discussion of numerical methods for determining real roots of polynomials.

where $\bar{v}^{(n)}$ is the nth iterate of the parameter vector \bar{v} . The solution sought is a fixed point of the map $\bar{\varphi}$. If $\bar{\varphi}$ satisfies a Lipschitz condition⁸

$$\|\bar{\varphi}(\bar{\mathbf{v}}^{(1)}) - \bar{\varphi}(\bar{\mathbf{v}}^{(2)})\| \leq \mathscr{L} \|\bar{\mathbf{v}}^{(1)} - \bar{\mathbf{v}}^{(2)}\|$$
(4.30)

for some $0 < \mathscr{L} < 1$ then $\overline{\varphi}$ is said to be a contraction map. Contraction maps are often used to prove existence theorems because the sequence of iterates generated by (4.29) is Cauchy.

The problem of designing an iterative procedure for solving a system of equations can be viewed as the problem of finding a contraction map whose fixed points coincide with the solutions sought. One usually begins with a map having the appropriate fixed points and then tries to show it satisfies a Lipschitz condition; often one employs the mean value theorem which states that if φ_n is a continuously differentiable function of the parameter vector $\bar{\mathbf{v}}$ then⁹

⁸The map $\overline{\varphi}$ is assumed to have its domain in a Banach space with norm $\|\cdot\|$ and its range contained by the domain.

⁹Two notational conventions are introduced here. First $\varphi_{n/l}$ denotes $\partial \varphi_n / \partial v_l$ and second the Einstein summation convention (with respect to repeated subscripts) is employed. The summation range is 0,1,...,P so that the Einstein convention implies summation with respect to the subscript l (only) over this range on the right hand side of (4.31). These conventions are used in this section only.

$$\varphi_{n}(\bar{\mathbf{v}}^{(1)}) - \varphi_{n}(\bar{\mathbf{v}}^{(2)}) = \varphi_{n/\ell}(\gamma \bar{\mathbf{v}}^{(1)} + [1-\gamma] \bar{\mathbf{v}}^{(2)}) \{\mathbf{v}_{\ell}^{(1)} - \mathbf{v}_{\ell}^{(2)}\}$$
(4.31)

for some $0 \le \gamma \le 1$. If one can determine a constant $\mathscr{L} \le 1$ majorizing the norm of the matrix with components $\varphi_{n/\ell}$ then $\bar{\varphi}$ has been demonstrated to satisfy a Lipschitz condition.

Using Equations (3.22), (3.30), (3.31), (3.32) and (4.1) the system of Equations (3.19) may be expressed as

$$\hat{V}_n = 0$$
; $n = 0, 1, \dots, P$ (4.32)

where

$$\hat{\mathbf{V}}_{\mathbf{n}} = \boldsymbol{\rho}_{\mathbf{n}} - \mathbf{L}_{\mathbf{n}\mathbf{m}} \mathbf{v}_{\mathbf{m}}$$
(4.33)

Defining

$$L_{nm\ell} = \int_{-\pi}^{\pi} W(\theta) [g(\theta)]^3 \cos(n\theta) \cos(m\theta) \cos(\ell\theta) d\theta/2\pi \quad (4.34)$$

and

$$\boldsymbol{\delta}_{nm} = \begin{cases} 0 & n \neq m \\ & & (4.35) \\ 1 & n = m \end{cases}$$

the following relations may be easily verified

$$L_{nm/\ell} = -2 L_{nm\ell} \qquad (4.36)$$

$$\hat{V}_{n/\ell} = -L_{nm/\ell} v_m - L_{nm} v_{m/\ell}$$

$$= 2 L_{nm\ell} v_m - L_{nm} \delta_{m\ell}$$

$$= 2 L_{n\ell} - L_{n\ell} = L_{n\ell} \qquad (4.37)$$

Consider the map $\bar{\varphi}$ with components¹⁰

$$\varphi_{n} = v_{n} - \lambda [L_{o}^{-1}]_{nm} \hat{V}_{m} \qquad (4.38)$$

where λ is a nonzero scalar constant. Use of this map for an iterative procedure is essentially a modified Newton method. First observe that $\bar{\varphi}$ has a fixed point if and only if the second term on the right hand side of (4.38) vanishes. This term vanishes if and only if Equations (4.32) are satisfied since, as shown in Chapter III, L (and so also L^{-1} and L_0^{-1}) is positive definite.

 $^{^{10}{\}rm If}$ L denotes the matrix with entries L_{nm} and L^{-1} the inverse of this matrix then $^{\rm L^{-1}}_{\rm (O)^{0}}$ shall denote L^{-1} evaluated at the initial iterate $\bar{v}^{\rm (O)^{0}}$ and $\left[L_{0}^{-1}\right]_{nm}$ its entries.

Next, using (4.37), consider

$$\rho_{n/\ell} = v_{n/\ell} - \lambda [L_0^{-1}]_{nm} \tilde{V}_{m/\ell}$$
$$= \delta_{n\ell} - \lambda [L_0^{-1}]_{nm} L_{m\ell} \qquad (4.39)$$

which, if evaluated at $\bar{v} = \bar{v}^{(0)}$, is

$$\varphi_{n/\ell}^{(0)} = (1-\lambda) \, \delta_{n\ell} \tag{4.40}$$

Clearly, (4.40) is majorized by $\mathscr{L} = |1-\lambda|$ so that λ should be selected in the range $0 < \lambda < 2$ if the Lipschitz condition is to be satisfied. More generally, since the last term in (4.39) is positive definite, λ should be selected in the range $0 < \lambda < 2/\lambda_{max}$ where

$$\lambda_{\max} \geq \sup_{\|q\|=1}^{\sup} q_n [L_0^{-1}]_{nm} L_{m\ell} q_{\ell}$$

$$(4.41)$$

bounds the matrix norm. With this selection

$$\inf_{\substack{|q||=1}} q_n \varphi_n/\ell q_\ell = 1 - \lambda \sup_{\substack{|q||=1}} q_n [L_0^{-1}]_{nm} L_{m\ell} q_\ell$$

$$\geq 1 - \lambda \lambda_{\max} > -1$$
 (4.42)

and the matrix norm of φ_n/ℓ is bounded by one.

Apparently the choice $\lambda = 1/\lambda_{max}$ would lead to the most rapid convergence while smaller values would lead to slower

convergence and guarantee that $\varphi_{n/\ell}$ is positive definite. Unfortunately, the right hand side of inequality (4.41) is a function of the parameter vector \bar{v} and cannot be bounded by a constant, λ_{max} , for all \bar{v} in \mathscr{R}_p ; consequently the Lipschitz condition cannot be satisfied everywhere in \mathscr{R}_p .

If a solution, $g_*(\theta)$, exists in \mathscr{R}_p it is possible to find a constant G_{max} sufficiently large such that

$$g_{*}(\theta) \leq G_{\max} \tag{4.43}$$

for all $\theta \in [-\pi, \pi)$. For such a constant the solution will be contained in that portion of \mathscr{R}_p for which

$$g(\theta) \leq G_{\max} \tag{4.44}$$

for all $\theta \in [-\pi, \pi)$. Then from

 $\sup_{\|q\|=1} q_n [L_0^{-1}]_{nm} L_{m\ell} q_{\ell}$

 $\leq \sup_{\|q\|=1}^{\sup} q_n L_{nm} q_m / \mathfrak{S}$ $\leq W_{max} G_{max}^2 / \mathfrak{S}$ (4.45)

where

 $W(\theta) \leq W_{max}$ (4.46)

for all
$$\theta \in [-\pi, \pi)$$
 and

$$\mathbf{\mathfrak{S}} = \inf_{\|\mathbf{q}\|=1}^{\inf} q_n \, \mathcal{L}_{nm}^{(0)} \, q_m > 0 \tag{4.47}$$

it is clear that any choice

$$\lambda_{\max} \geq W_{\max} \ G_{\max}^2/6$$
 (4.48)

will suffice to satisfy the Lipschitz condition for that portion of $\mathcal{R}_{\rm P}$.

To recapitulate, the map $\bar{\varphi}$, defined by (4.38), has fixed points coinciding with the solutions to (4.32). Moreover, if there exists a solution in \mathscr{R}_p and the domain of $\bar{\varphi}$ is suitably restricted to a subset of \mathscr{R}_p containing this solution then there exists $\lambda > 0$ sufficiently small such that $\bar{\varphi}$ satisfies a Lipschitz condition on this subset and (4.39) is positive definite. This implies that application of the map $\bar{\varphi}$ to any element of the subset will generate a new parameter vector closer (in norm) to the solution. Hopefully, repeated application of $\bar{\varphi}$ will generate a sequence of parameter vectors approaching the solution; this will be the case if each new parameter vector is also in the restricted domain of $\bar{\varphi}$.

Providing a guarantee that each new parameter vector will be within the restricted domain of $\bar{\varphi}$ is not a simple task. Without such a guarantee it is possible to devise a computational test to check for this condition; then, if the

test is violated, some method must be devised to restart the iterations. In practice the situation is not expected to be quite so pathological; if λ is selected to be conservatively small (smaller if the solution is expected to be a sharply peaked spectrum) and a reasonably good initial estimate is provided, one does not expect to encounter convergence difficulties. This more optimistic approach shall be taken in the following.

To implement the iterative procedure assume $W(\theta)$ is available in sampled form. The components of the nth iterate parameter vector may be used to evaluate

$$g_{n}(\theta) = 1/\{ \sum_{\ell=0}^{P} v_{\ell}^{(n)} \cos(\ell \theta) \}$$
 (4.49)

in sampled form. If the sample mesh is equally spaced at

$$\theta_k = \pi k/N$$
; $k = -N, \dots, 0, 1, \dots, N-1$ (4.50)

then the components $\widehat{V}_{\underline{m}}^{\,(\,n\,)}$ may be computed from

$$\hat{\mathbf{V}}_{\mathbf{m}}^{(n)} = \rho_{\mathbf{m}} - \sum_{\mathbf{k}=-\mathbf{N}}^{\mathbf{N}-1} \mathbf{W}(\boldsymbol{\theta}_{\mathbf{k}}) g_{\mathbf{n}}(\boldsymbol{\theta}_{\mathbf{k}}) \cos(\mathbf{m}\boldsymbol{\theta}_{\mathbf{k}})/2\mathbf{N}$$
(4.51)

and the components of the next iterate are provided by

$$v_{\ell}^{(n+1)} = v_{\ell}^{(n)} - \lambda [L_0^{-1}]_{\ell m} \hat{v}_{m}^{(n)}$$
(4.52)

							_
				END			
				FILWED 10140			



MICROCOPY RESOLUTION TEST CHART NATIONAL BUREAU OF STANDARDS-1963-A A crude test that the n^{th} iterate is in \mathscr{R}_p is provided in the course of these computations by verifying that the denominator of (4.49) is positive on the sample mesh.

The procedure can be initialized by the solution to the Yule-Walker equations where the elements of the coefficient matrix are given by $\rho_{\rm m}$. Equations (3.23) and (3.30) may then be used to evaluate $v_{\ell}^{(O)}$ while the elements $[{\rm L_O^{-1}}]_{\rm nm}$ may be obtained by inverting the real symmetric matrix with entries

$$[L_0]_{nm} = \sum_{k=-N}^{N-1} W(\theta_k) [g_0(\theta_k)]^2 \cos(n\theta_k) \cos(m\theta_k)/2N \quad (4.53)$$

The coefficients ρ_{m} may be evaluated from

$$\rho_{\rm m} = \sum_{\rm k=-N}^{\rm N-1} W(\theta_{\rm k}) \hat{H}(\theta_{\rm k}) \ f(\theta_{\rm k}) \ \cos({\rm m}\theta_{\rm k})/2N \qquad (4.54)$$

Alternatively, the computational methods described in the previous section may be employed to evaluate the ρ_m as lagged products of modified data values.

A simple test for iteration completion is to simply check that

$$\psi_n = \sum_{m=0}^{P} [\hat{v}_m^{(n)}]^2$$
 (4.55)

is less than some small preselected value. Finally, to obtain filter coefficients as are required by many

applications, it is perhaps simplest to first compute correlation values from

$$\mathbf{r}_{\mathbf{m}} = \sum_{\mathbf{k}=-N}^{N-1} g_{\mathbf{n}}(\boldsymbol{\theta}_{\mathbf{k}}) \cos(\mathbf{m}\boldsymbol{\theta}_{\mathbf{k}}) \qquad (4.56)$$

and then solve the Yule-Walker equations.

If at some step prior to iteration completion an iterate falls outside \mathscr{R}_{p} , one may attempt to reinitialize the procedure using one of the last few iterates inside \mathscr{R}_{p} .

Formulae for Vector Quantization

In this section formulae relevant to the problem of minimizing $I_W(\hat{H}f,g)$ over a specified finite collection of AR(P) model spectra are developed. Consider first that according to Equation (3.24) this problem is equivalent to minimizing $I_W(g_1,g)$ where $g_1(\theta)$ is an AR(P) model spectrum satisfying Equation (3.19). Next, observe that minimizing $I_W(g_1,g)$ is equivalent to minimizing

$$J_{W}(g_{1},g) = \int_{-\pi}^{\pi} [W(\theta) g_{1}(\theta)/g(\theta) + W(\theta) \ln g(\theta)] d\theta/2\pi \quad (4.57)$$

Since $g(\theta)$ is an AR(P) model given by Equation (3.22) the first term in Equation (4.57) may be rewritten as

$$\int_{-\pi}^{\pi} W(\theta) g_{1}(\theta)/g(\theta) d\theta/2\pi = \sum_{n=-P}^{P} u_{[n]} \rho_{[n]} \qquad (4.58a)$$

where the fact that $g_1(\theta)$ satisfies Equation (3.19) has been used together with Equation (4.1). Similarly, the second term in Equation (4.57) may be rewritten as

$$\int_{-\pi}^{\pi} W(\theta) \ln g(\theta) d\theta / 2\pi = \ln(\sigma^2) \int_{-\pi}^{\pi} W(\theta) d\theta / 2\pi$$
$$- \int_{-\pi}^{\pi} W(\theta) \ln[A_p(e^{i\theta}) A_p(e^{-i\theta})] d\theta / 2\pi \qquad (4.58b)$$

In general $J_W(g_1,g)$ will be minimized over the finite collection of AR(P) spectra by evaluating this quantity for each model spectrum in the collection. For any given model spectrum the first term may be easily evaluated using (4.58a); the coefficients ρ_n may be determined from the data using one of the methods outlined in the second section of this chapter. The second term presents somewhat greater difficulty; when $W(\theta) = 1$ the last term in (4.58b) may be shown to vanish as a consequence of Jensen's theorem but, in general, this term will not vanish.

When $W(\theta)$ has an AR(M) form an extension of Jensen's theorem, which shall be developed presently, permits the evaluation of this term from a simple formula. In order to establish the general theorem it shall be necessary to first establish the following lemma.

Lemma 4-1. Let

$$A_P(z^{-1}) = \prod_{m=1}^{P} (1 - \eta_m z)$$
 (4.59)

have no roots inside the unit circle, Γ . If τ_k and τ_l are also within the unit circle then

$$T_{\ell k} = \oint_{\Gamma} \{ \ln A_{P}(z^{-1}) \} / \{ (z - \tau_{k})(1 - \tau_{\ell} z) \} dz / 2\pi i$$

= $\{ \ln A_{P}(\tau_{k}^{-1}) \} / (1 - \tau_{k} \tau_{\ell})$ (4.60)

<u>Proof</u>. The method of proof is essentially the same as that used for Jensen's theorem by Hille [62, pp. 256-7]. Assume without loss of generality that a narrow strip from τ_k to $\nu_k = \tau_k / |\tau_k|$ is free of the η_m and consider the integral

$$\mathbf{z}_{\ell k} = \oint_{\mathcal{C}} \{ \ln[(z - \tau_k) / (1 - \tau_\ell z)] \} d[\ln A_p(z^{-1})] / 2\pi i \quad (4.61)$$

around the contour, \mathscr{C} , depicted in Figure 5. The logarithm, determined so that $\ln(-1) = \pi i$, is analytic within \mathscr{C} and $A_p(z^{-1})$ has neither poles nor zeros within \mathscr{C} so $\Xi_{\ell k} = 0$. As the radius of the circular portion of the contour, \mathscr{C} , surrounding the singularity τ_k tends to zero it offers no contribution to this integral. As the distance between the two straight sections of the contour tends to zero they provide the contribution

$$N_{k} = \int_{z=v_{k}}^{z=\tau_{k}} [A_{p}(z^{-1})]^{-1} d[A_{p}(z^{-1})]$$

= $\ln A_{p}(\tau_{k}^{-1}) - \ln A_{p}(v_{k}^{-1})$ (4.62)



Figure 5. The Contour & in the Complex Z-Plane

For the remaining portion of the contour, integration by parts yields

$$(1 - \tau_k \tau_l) T_{lk} = \aleph_k - \Xi_{lk} + \Delta_{\Gamma}/2\pi i \qquad (4.63)$$

where the integrated part is

Ł

$$\Delta_{\Gamma} = \{ \ln A_{P}(z^{-1}) \ln[(z-\tau_{k})/(1-\tau_{\ell} z)] \} | \begin{cases} z = \nu_{k}^{(+)} \\ z = \nu_{k} \end{cases}$$

$$= \ln A_{P}(\nu_{k}^{-1}) \{ 2\pi i + \ln[\nu_{k}-\tau_{k})/(1-\tau_{\ell} \nu_{k})] \}$$

$$-\ln A_{P}(\nu_{k}^{-1}) \{ \ln[(\nu_{k}-\tau_{k})/(1-\tau_{\ell} \nu_{k})] \}$$

$$= 2\pi i \ln A_{P}(\nu_{k}^{-1}) \qquad (4.64)$$

Substitution of (4.62) and (4.64) along with $\Xi_{lk} = 0$ into Equation (4.63) completes the proof.

A simple variable substitution may be used to obtain the related formula

$$T_{k\ell} = \oint_{\Gamma} \{\ln A_{P}(z)\} / \{(z - \tau_{k})(1 - \tau_{\ell} z)\} dz / 2\pi i$$

= $\{\ln A_{P}(\tau_{\ell}^{-1})\} / (1 - \tau_{k} \tau_{\ell})$ (4.65)

which together with (4.60) establishes the

Corollary 4-1.

$$\Gamma_{\ell k} + \Gamma_{k \ell} = \oint_{\Gamma} \{ \ln A_{P}(z) A_{P}(z^{-1}) \} / \{ (z - \tau_{k})(1 - \tau_{\ell} z) \} dz/2\pi i$$
$$= \{ \ln A_{P}(\tau_{k}^{-1}) A_{P}(\tau_{\ell}^{-1}) \} / (1 - \tau_{k} \tau_{\ell})$$
(4.66)

Finally, sufficient background has now been presented to establish the

Theorem 4-1. Let W(θ) have an AR(M) form given by W(θ) = $|\Omega(e^{i\theta})|^2$ (4.67)

where $\Omega(z)$ has the partial fraction expansion

$$\Omega(z) = \sum_{\ell=1}^{M} \omega_{\ell} / (1 - \tau_{\ell} z^{-1}) = \sigma_{W} / B_{M}(z) \qquad (4.68)$$

with $|\tau_{\ell}| < 1$. Then with $g(\theta)$ given by equation (2.3) the second term in (4.57) is

$$\int_{-\pi}^{\pi} W(\theta) \ln g(\theta) d\theta/2\pi = \ln \sigma^2 \int_{-\pi}^{\pi} W(\theta) d\theta/2\pi - T \qquad (4.69)$$

where

$$T = 2 \sum_{k=1}^{M} \omega_{k} \quad \Omega(\tau_{k}^{-1}) \ln A_{p}(\tau_{k}^{-1})$$
 (4.70)

Proof. Using (2.3), (4.67), and (4.68)

$$T = \int_{-\pi}^{\pi} |\Omega(e^{i\theta})|^2 \ln |A_p(e^{i\theta})|^2 d\theta/2\pi$$

$$= \sum_{k,\ell=1}^{M} \omega_k \omega_\ell \oint_{\Gamma} \{\ln A_p(z) |A_p(z^{-1})\}/\{(z-\tau_k)(1-\tau_\ell z)\} dz/2\pi i$$
(4.71)

Together with the above corollary this yields

$$\mathbf{T} = \sum_{k,\ell=1}^{M} \{ \omega_{k} \omega_{\ell} / (1 - \tau_{k} \tau_{\ell}) \} \ln A_{P}(\tau_{k}^{-1}) A_{P}(\tau_{\ell}^{-1})$$
(4.72)

and (upon splitting the logarithm and collecting terms) Equation (4.70).

With $W(\theta) = 1$ this theorem yields

$$\int_{-\pi}^{\pi} \ln g(\theta) \, d\theta/2\pi = \ln \sigma^2 \qquad (4.73)$$

which is a special case of Jensen's theorem [62, Theorem 9.2.5]. The first term in Equation (4.69) is easy to compute while the second term, T, given by Equation (4.70) may offer the reader some difficulty. First observe that (4.70) requires knowledge of the parameters of the partial fraction expansion (4.68). These are fairly easy to determine once

the roots τ_k of $B_M(z)$ are known by recognizing that ω_k equals¹¹

$$(z-\tau_k) z^{-P} \Omega(z) = \sigma_W(z-\tau_k)/\{z^P B_M(z)\}$$
 (4.74)

evaluated at $z = \tau_k$. Hence, the basic difficulty is that of determing the roots, τ_k .

Since extracting the roots of $B_M(z)$ can be a difficult problem for large values of M it is advantageous if $B_M(z)$ is already known as a product of low order factors. To accomplish this, recall that $B_M(z)$ is determined so that $W(\theta)$ approximates $\hat{H}(\theta)$. If $W(\theta)$ is a product of known AR(2) models

$$W(\theta) = W_1(\theta) W_2(\theta) \dots W_{W/2}(\theta)$$

$$(4.75)$$

then $B_M(z)$ is easily known as a product of second order factors. In order to determine $W(\theta)$ in this manner one may first determine $W_1(\theta)$ to approximate $\hat{H}(\theta)$, then $W_2(\theta)$ to approximate $\hat{H}(\theta)/W_1(\theta)$, then $W_3(\theta)$ to approximate $\hat{H}(\theta)/[W_1(\theta) W_2(\theta)]$ and so on. To obtain the best overall approximation it is probably advantageous to develop some simple ad hoc method to force the approximation at each

¹¹This assumes the roots, τ_k , are distinct. The formulae become mildly more complicated when this is not the case.

stage to fit no more than one strong resonance in the function being approximated.

Remarks

This chapter has explored computational procedures related to the weighted information estimation formulation developed in Chapter III; it is worth noting that none of these methods is entirely satisfactory for all applications.

The first section employed an assumed AR(M) form for the weight function which enabled the problem to be cast in the form of a nonlinear system of polynomial equations. Solution of the system was found to be a relatively simple task for small values of M but one that becomes rapidly more complex as M is increased beyond four. As a general approach, the assumption of a parametric form for the weight function has considerable promise for the development of efficient computational methods; the basic difficulty is that of finding a clever parametrization which provides sufficient flexibility in the form of the weight function (for the given application) while leading to a simple and efficient computational algorithm.

The second section discussed the computation of various coefficients that arise within the computational formulae. Choice of a specific procedure will ultimately be influenced by the demands of the specific application; interdependant

factors to be considered include the quantity of data available, rounding/truncation effects, fixed/floating point representation format, algorithm structure, memory requirements, and computational speed. The coefficient evaluation procedures discussed are variants of methods proposed (and sometimes implemented) for real time speech analysis applications.

section discussed single-step iterative The third methods within the general framework provided by the nction of a contraction mapping. Multi-step methods were not discussed; in general, convergence characteristics are more difficult to prove for multi-step methods in spite of the fact that they tend to converge faster in practice.¹² These iterative methods offer significantly more flexibility in the form of the weight function¹³ at the expense of a The notion of a contraction greater computational cost. sometimes employed for nonconstructive existence map, proofs, provides a useful general framework within which a

 $^{^{12}}$ Faster convergence, in terms of a reduced number of iterations, should not be confused with reduced computational cost. Each iteration of a multi-step method generally is more expensive computationally than a comparable single-step method so that a detailed analysis is usually required to compare costs.

¹³That is, compared to the parametric approach to weight function selection discussed in the first section. In this sense one might describe these methods with a seemingly contradictory phrase such as "nonparametric autoregressive estimation".

variety of iterative methods may be discussed; the specific method presented is a modified Newtonian iteration chosen as a tradeoff between simplicity and effectiveness. A possibly more effective iterative procedure would be a steepest descent method; generally such a procedure attempts to minimize a scalar function $U = U(\bar{v})$ by using a map with components

ن کاریند د

$$\varphi_{n} = v_{n} - \lambda \, \partial U / \partial v_{n} \tag{4.76}$$

where the scalar function $\lambda = \lambda(\bar{v})$ is chosen to minimize $U(\bar{v})$ at each iteration.

The fourth section considers the problem of minimizing $I_W(\hat{H}f,g)$ over a given finite collection of AR(P) models. The procedure involves the computation of a cost function for each model in the collection. The cost function involves two terms; the first term is evaluated quite simply (regardless of the form of the weight function) using formula (4.58a) which is identical to one arising in "standard" (unweighted) vector quantization. The second term is usually quite simple in "standard" vector quantization, see Equation (4.73), but becomes far more complex when the weighted information formulation is employed.

An extension of Jensen's theorem provides a formula which may be employed to evaluate this term when $W(\theta)$ has an AR(M) form; however, the reader is admonished to bear in

mind that it is probably far simpler to discretize this integral and evaluate it numerically as a sum of products from

$$\int_{-\pi}^{\pi} W(\theta) \ln g(\theta) d\theta / 2\pi = \sum_{k=-N}^{N-1} W(\theta_k) \mathscr{G}_k \qquad (4.77)$$

where

$$\mathscr{G}_{\mathbf{k}} = [\ln g(\boldsymbol{\theta}_{\mathbf{k}})]/(2\mathbf{N})$$
(4.78)

This has the additional advantage of not imposing an AR(M) form upon the weight function. More generally, $W(\Theta)$ might be expressed as a sum of perhaps only a dozen nonnegative "shape functions" by

$$W(\theta) = \sum_{k} \boldsymbol{\xi}_{k} \boldsymbol{\mathscr{W}}_{k}(\theta)$$
 (4.79)

so that, if the quantities

$$\mathscr{F}_{k} = \int_{-\pi}^{\pi} \mathscr{H}_{k}(\theta) \ln g(\theta) d\theta / 2\pi \qquad (4.80)$$

are precomputed for each AR model in the finite collection, the second term may be easily evaluated from

$$\int_{-\pi}^{\pi} W(\theta) \ln g(\theta) d\theta / 2\pi = \sum_{k} \boldsymbol{\xi}_{k} \boldsymbol{\mathscr{F}}_{k}$$
(4.81)

CHAPTER V

RESULTS

In this chapter the weighted information estimation formulation is demonstrated to provide improved performance relative to the noise filtering formulation. It is worth noting that, although existence has not been proven in previous chapters, several thousand data frames have been analyzed using the weighted information formulation and not one counterexample has been encountered.

Gaussian Signals

In order to study the performance of the weighted information formulation pseudorandom sequences were generated. A zero-mean white Gaussian process was simulated using a congruential multiplicative random number generator; the resulting sequence of independent uniformly distributed samples was transformed to Gaussian form using the Box-Müller transformation followed by Central-Limit averaging.¹ Zero-mean AR(P) Gaussian processes were simulated

¹In theory, the Box-Müller transformation is adequate. However, if the input deviates from a uniform distribution the output will, correspondingly, deviate from a Gaussian distribution; Central-Limit averaging will tend to reduce any such deviations from a Gaussian form.

by applying the simulated white Gaussian process to an allpole (lattice structure) digital filter; the first few thousand output values from the filter were consistently ignored in order to avoid the transient response of the filter.

By adding two independent zero-mean Gaussian AR processes at a specified signal to noise ratio appropriate test data was produced. For many of the examples the "signal" process had an AR(2) spectrum defined by the reflection coefficient values

$$k_1 = -.8$$
 and $k_2 = -.9$ (5.1)

This signal process spectrum, evaluated from these parameter values, is displayed in Figure 6a. While some examples employ a white Gaussian corrupting noise process, others employ an AR(2) process defined by the reflection coefficient values

 $k_1 = +.8$ and $k_2 = -.9$ (5.2)

This "colored" noise process spectrum is displayed in Figure 6b.

As a basis for comparison, the standard autocorrelation analysis method was applied to 100 different 400 sample Hamming windowed frames of data from the uncorrupted signal process. Each resulting estimate is characterized by a pair



Figure 6a. True Spectrum; Test Signal Process



Figure 6b. True Spectrum; Test Noise Process

Figure 6. Test Signal Spectra

of reflection coefficients which define a single dot in Figure 7. For this "scatter plot" (and all subsequent scatter plots) the ordinate and the abcissa correspond to the first and second reflection coefficients, respectively; for convenience, cross-hairs indicate the location of the true parameter values.

Figures 8 and 9 each present various estimates of a single 200 sample Hamming windowed frame of data. In both cases the frame of data consists of the signal and colored noise processes summed at a 10 dB signal to noise ratio. The periodogram estimates in Figures 8a and 9a clearly display the signal resonance (near the fractional frequency value of .8) and the noise resonance (near the fractional frequency value of .2).

Figures 8b and 9b display power spectrum estimates obtained using the noise filtering formulation. The estimate presented in Figure 8b is a result of using the noise filter response displayed in Figure 8c which was determined by using the power subtraction rule;² similarly, Figure 9b results from the use of the noise filter response displayed in Figure 9c which was determined by using the magnitude subtraction rule.

²As indicated in the caption, the noise filter response was smoothed across frequencies before being applied. Although many smoothing algorithms are possible, only a recursive median smoother (with a length about 2.5% of the displayed bandwidth) was ever employed to obtain results presented in this chapter.



و المعالية ا




[

202

•

Figures 8d and 9d display power spectrum estimates obtained using the weighted information formulation. The algebraic solution method, which requires an AR(M) form for the weight function, was used in both cases; coefficient evaluation was performed using the mixed time-frequency domain method presented in Figure 4. The same noise filter response functions were employed and the weight functions, displayed in Figures 8e and 9e, were determined as an AR(2) fit to their respective noise filter response functions.

By comparing Figures 8 and 9 to the true signal spectrum shown in Figure 6a the deficiencies of these typical estimates becomes apparent. In Figure 8b the noise filtering formulation leads to an estimate which is overly flat; the weighted information formulation (Figure 8d) has improved the estimate by raising the peak and lowering the vallevs. In Figure 9b the noise filtering formulation leads to an estimate which is overly sharp; the weighted information formulation (Figure 9d) has improved the estimate by lowering the peak and raising the valleys. Since the weight functions are similar in both figures it is apparent that frequency weighting cannot be simply interpreted as increasing or decreasing the sharpness of a spectral estimate; rather, the weight function reduces distortions in the estimate by requiring a more accurate fit to the data in those spectral regions where the weight function is large.

Figures 10 and 11 present the result of analyzing 100 different 400 sample Hamming windowed frames of data using



Noise (10 $d\overline{B}$)

various different methods. Figure 10 presents the results obtained using the noise filtering formulation; the smoothed noise filter response was determined using different rules ranging (roughly) from the least severe rule in Figure 10a to the most severe in Figure 10e. The results presented in Figure 11 represent an analysis of the same 100 data frames and the same noise filter response functions but the analysis uses the weighted information formulation with an AR(2) weight function fit to the noise filter response function.

It is clear that in each case (a through e) the estimation error is reduced by the weighted information formu-The best results in both figures are obtained by lation. the most severe rules. Figure 10, while exhibiting less variance, shows an increased deviation (bias) of the main cluster from the true values for the more severe rules; apparently, variance error of the noise filtering formulation may be reduced at the expense of increased bias error by using the more severe rules. Comparing, for example, Figures 10e and 11e it is apparent that the weighted information formulation achieves still greater variance reduction while correcting the bias error. Comparison of Figures 11e and 7 indicate that one has little, if any, hope of achieving significantly better performance than that provided by the weighted information formulation in this case.

Figures 12 and 13 show similar results for the same 100 frames of data; the analysis methods used to produce these



٩.

...

figures differ from the method used to produce Figures 10 and 11 only in that no smoothing algorithm was applied to the noise filter response. All the same trends are apparent in figures 12 and 13 as were apparent in Figures 10 and 11; somewhat greater variance is exhibited in these figures indicating that smoothing produces a generally beneficial effect in this case.

Figures 14, 15, 16, and 17 display similar results for the case of white corrupting noise at a 10 dB signal to noise ratio. Again, each plot represents analysis of the same 100 different 400 sample Hamming windowed frames of data. For each method of determining noise filter response, the weighted information formulation leads to less variance and bias error than the comparable (unweighted) noise filtering formulation. As may be expected,³ all these estimators yield poorer performance in this white noise case than in the previous colored noise case.

Figures 18, 19, 20, and 21 again present similar results; while the corrupting noise is still white the signal to noise ratio is now zero dB. One small difference is worth noting: in Figures 10 through 17 the parts b, c, and d employed a soft suppression rule with suppression factors

³The reader will recall that if the signal and noise processes are completely separated in frequency (i.e., do not have overlapping spectra), the Wiener filter can completely eliminate the noise. Since the colored noise case exhibits greater spectral separation from this signal process than the white noise case, an estimate can be expected to provide superior performance.



ļ



Noise (10 dB)

White Noise (10 dB)





Noise (0 dB)

of 4, 6, and 8 respectively; in Figures 18 through 21 the parts b, c, and d again employ a soft suppression rule but with increased suppression factors of 6, 8, and 10 respectively.

Speech and Speech-Like Signals

Many speech waveforms exhibit a nonrandom periodic character; their spectra display a fine harmonic structure (with peaks separated by integral multiples of the pitch frequency) with a roughly AR modulation. The harmonic structure is generally attributed to the periodic glottal pulses while the AR modulation is generally attributed to the response characteristics of the vocal tract.

To simulate such waveforms the all pole filter with frequency response displayed in figure 6a was excited with a periodic stream of impulses (with a period of 79 samples). No figure comparable to Figure 7 is included here since, in the absence of noise, the analysis of 100 different 400 sample Hamming windowed frames of data (with a random distribution of phase displacement) presents no apparent estimation error.⁴ Consequently, while part of the apparent estimation error in the scatter plots of Figures 10 through 21 must be attributed to the random character of the signal

⁴That is, on the scale used for these scatter plots. On a greatly enlarged scale, a small amount of bias and variance error may be observed.

itself, all of the apparent estimation error in the following scatter plots (Figures 24 through 35) may be attributed to the presence of noise.

Figures 22 and 23 each present various estimates of a single 200 sample Hamming windowed frame of data. In both cases the frame of data consists of the aforementioned periodic signal process and a colored Gaussian noise process summed at a 10 dB signal to noise ratio. The periodogram estimates in Figures 22a and 23a clearly display the fine harmonic structure of the signal spectrum near the filter resonance (fractional frequency of .8) while this structure breaks down near the noise resonance (fractional frequency of .2).

Figures 22b and 23b display estimates obtained using the noise filtering formulation; Figures 22c and 23c display the noise filter response characteristics that produced these estimates. Clearly the estimate appearing in Figure 22b is overly flat while the estimate appearing in Figure 23b is overly sharp. Figures 22d and 23d display the obtained estimates using the weighted information formulation; comparison with Figure 6a reveals that both these estimates are improved relative to their counterparts in Figures 22b and 23b. Finally the AR(2) weight functions approximating the noise filter response functions are presented in Figures 22e and 23e.

Figures 24, 25, 26, and 27 display a variety of scatter plots; each scatter plot presents the result of analyzing





CALLER CONTRACTOR CONTRACTOR

P

215







10.75







Figure 32. Noise Filtering: Periodic Signal & White Noise (0 dB)



. . .



•

100 different 400 sample Hamming windowed frames of data; the same 100 data frames were employed for each plot. As mentioned above, because the signal process is periodic and not random in character all of the apparent estimation error can be attributed to the added colored Gaussian noise (SNR = 10 dB).

Figures 24 and 25 employ smoothed noise filter response characteristics while Figures 26 and 27 employ the unsmoothed characteristics.⁵ Figures 24 and 26 display the results obtained with the noise filtering formulation while Figures 25 and 27 display the results obtained with the AR(2) weighted information formulation. Once again, thweighted information formulation leads to less estimation error than the comparable noise filtering formulation; in Figures 25d and 25e the estimation error is so small as to be almost imperceptible on the scale employed for these plots. Smoothing still appears to display a generally beneficial effect.

Figures 28, 29, 30, and 31 present similar results for the case of white Gaussian noise corruption to the periodic signal processes (SNR = 10 dB). As with the Gaussian signal

⁵Some caution is advised regarding the use of smoothers here. The dimensions of the lobes within the fine harmonic structure are controlled by the length and shape of the data window so that a smoother that works well with one frame length may not work well with longer frames or a differently shaped window.

process, all the estimates present degraded performance in this white noise case as compared to the colcred noise case.

•

To complete these simulations, Figures 32, 33, 34, and 35 present analysis results for the case of white Gaussian noise corruption to the periodic signal process at a zero dB signal to noise ratiio. As with the Gaussian signal process, parts b, c, and d of these figures employ soft suppression rules with increased suppression factors of 6, 8, and 10 respectively.

The following summarizes the description of these scatter plots. Figures 10-13 and 24-27 correspond to colored noise corruption at a 10 dB signal to noise ratio; Figures 14-17 and 28-31 correspond to white noise corruption at a 10 dB signal to noise ratio; Figures 18-21 and 32-35 correspond to white noise corruption at a O dB signal to noise ratio. Figures 10-21 correspond to a Gaussian random signal: Figures 24-35 correspond to a periodic (period = 79 samples) signal. Figures 10, 11, 14, 15, 18, 19, 24, 25, 28, 29, 32. and 33 employ a smoothed noise filter response while the remainder employ an unsmoothed response; parts a and e of each of these figures determine the noise filter response using the power and magnitude subtraction rules respectively while parts b, c, and d employ the soft suppression rules. In Figures 10-17 and 24-31 the suppression factors for parts b, c, and d are 4, 6, and 8 respectively; in Figures 18-21 and 32-35 the suppression factors and 6, 8, and 10 respectively. Finally, Figures 10, 12, 14, 16, 18, 20, 24,

26, 28, 30, 32 and 34 display the results of the (unweighted) noise filtering analysis while Figures 11, 13, 15, 17, 19, 21, 25, 27, 29, 31, 33, and 35 display the results of the AR(2) weighted information analysis.

Before concluding this chapter, several examples resulting from the analysis of a real speech segment are provided. Figure 36a shows a periodogram estimate obtained from a Hamming windowed 400 sample segment taken from the vowel portion of the word "wrap";⁶ from the fine harmonic structure it is apparent that the pitch of this segment is about 135 Hz (about 59 samples). Figure 36b shows a tenth order AR estimate of the spectrum obtained as the result of an autocorrelation method analysis of the same Hamming windowed segment; four vocal tract resonances are clearly visible.⁷

Figures 37a and 37b show periodogram and tenth order AR estimates obtained from this same vowel segment after adding white noise at a 10 dB signal to noise ratio. Clearly, the fine harmonic structure of the periodogram estimate has been partially obscured and, while four resonances are still visible, the AR estimate is severely distorted.

⁶The word, spoken in context by an adult male in a quiet environment, was taken from the sentence "Don't gift wrap the tall glass." and was appropriately filtered before sampling at 8 kHz.

⁷Lower and higher order analyses were applied to this segment and it was judged from plots such as these that a tenth order model is appropriate.



 (a) Log Power Spectrum (dB vs Fractional Frequency); Periodogram Estimate of Vowel Spectrum; Noise Free



÷

(b) Log Power Spectrum (dB vs Fractional Frequency); AR(10) Estimate Using Autocorrelation Method; Noise Free

Figure 36. Vowel Spectrum in Quiet Environment



(b) Log Power Spectrum (dB vs Fractional Frequency); AR(10) Estimate Using Autocorrelation Method; SNR = 10 dB

Figure 37. Vowel Spectrum in White Noise

Figures 38 and 39 display the result of applying various other estimators to the same white noise corrupted data frame. Figure 38 shows results obtained with the smoothed power subtraction rule and figure 39 shows results obtained with the smoothed magnitude subtraction rule. Part a of each figure shows the result obtained with the noise filtering formulation; the noise filter response functions are displayed in part b. The weighted information estimates, displayed in part c, were obtained using the modified Newton iteration described in Chapter IV; the weight functions, displayed in part d, were selected as an AR(6) fit to the noise filter response functions displayed in part b.⁸

Comparison of figures 38a and 39a to figure 36b reveals the deficiencies of these noise filtered estimates; in particular, the reader should note the amplitude of the third and fourth (highest frequency) resonance peaks as well as the depth of the valleys near the fractional frequency values of zero and one. These features are partially corrected in figures 38c and 39c by the weighted information formulation; most notable is the correction of the valley depth near the fractional frequency value of zero. Also worth noting is the improved valley depth near the fractional frequency of one in figure 38c and the improved amplitude of the fourth resonance peak in figure 39c.

⁸The weight functions need not be selected to have an AR form; however, with this iterative method, convergence is more difficult to achieve with more complex weight function forms.



CHAPTER VI

CONCLUSION

A new method of spectral estimation has been presented. The method addresses the problem of noise corruption to the time series measurements and assumes knowledge of the noise power spectral density.¹ The method has been demonstrated to yield superior performance, in terms of reduced estimation error, and has been suggested for use in speech analysis applications.

Although the Gaussian assumption is invoked for the theoretical development of the method, examples have been provided that show the method yields superior performance for other signals as well. Similarly, the method is considered to be fairly robust with respect to the other assumptions.² It is worth noting that while the AR signal model has been assumed throughout, this assumption is by no means necessary to the theoretical development so that

¹Actually, only knowledge of the frequency response of a filter designed to eliminate the noise is assumed. Knowledge of the noise power spectral density merely leads to one common method of designing such a filter.

²A possible exception is the assumption of independence between the signal and noise processes for it is this assumption that leads to the model of additive signal and noise power spectral densities.

other (e.g. ARMA, Pisarenko, etc.) models may also be considered. 3

Computational procedures relevant to the problem of AR model estimation (using the weighted information formulation) have been explored. An algebraic method, applicable when the weight function assumes an AR(M) form, has been discussed; when $M \leq 4$, this method will obtain the solution using an algorithm of reasonable complexity for many applications. Iterative techniques have been discussed that obtain the solution while permitting an extremely flexible class of weight functions; the price of this greater flexibility is a considerable increase in complexity as well as the need for much user interaction. Several methods of coefficient evaluation were presented; one was implemented and used to obtain the simulation results.

The problem of AR model detection (vector quantization) requires the evaluation of two integrals for each model in the finite collection. Evaluation of the first integral is accomplished by Equation (4.58a); this equation requires the same number of additions, multiplications, and (read-only) storage locations as is required by the usual (unweighted)

⁹The new formulation would still require minimization of $I_W(\hat{H}f,g)$ and the analogy leading to Equation (3.20) still applies. The only difference is in the selection of a parametric signal model and the system of equations that follows. Uniqueness questions would need to be addressed separately but one may hope to find that similar convexity arguments would apply. Of course, the computational procedures discussed earlier may no longer be appropriate.

methods of vector quantization. The second integral is evaluated as a constant (independent of the data but depending upon the model) by the usual (unweighted) methods of vector quantization; Equation (4.81) is advocated for evaluation of the second integral with the weighted information formulation. With about a dozen terms, as suggested for speech analysis applications, evaluation of the second integral using Equation (4.81) is about equivalent in complexity to evaluation of the first integral.

Suggestions for Future Research

There are numerous ways to extend and refine the ideas and methods presented here. The following suggestions, offered in no particular order, are thought to be worthwhile.

- Extension to other spectral models. As mentioned earlier, the AR model form is not necessary; moreover, for some applications it may not even be appropriate.
- Assuming an AR model, determine the conditions for (and a proof of) existence. Empirical evidence for existence is strong; it is thought that the conditions are quite mild from a practical viewpoint (e.g. that the weight function is bounded). While the question of existence is mostly of theoretical interest by itself; the methods used to prove existence (and the precise

conditions for existence) should have practical value. For example, a proof based upon a contraction map is likely to yield a highly effective iterative solution procedure as well.

- Further investigation of methods of coefficient evaluation. These should be studied in close relation to the specific application in order to select a design offering a reasonable tradeoff between computational effort and performance.
- Investigation of numerical methods for solution of the ideal formulation. It is thought that the ideal formulation should yield still better performance, particularly at very low signal to noise ratios; it is expected that these methods will be very computationally expensive.
- Development of related formulations assuming a correlated noise model. The cross-spectrum (between the signal and noise processes) may be known, say, as a function of the unknown signal model spectrum and the known noise spectrum in some applications; this may occur, for example, if additive independent signal and noise processes were passed through a known nonlinear system prior to observation.

- Further investigation of computational methods appropriate for the AR weight function model; investigation of computational methods appropriate for other parametric weight function models. While the uniqueness result guarantees that only one product model, $C_{P+M}(z)$, satisfying equations (4.13) and (4.15) has all its "additional" reflection coefficients $\{k_{P+1}, k_{P+2}, k_{P+2}\}$..., k_{P+M} inside the interval (-1,1) it is not known if the other product models satisfying these equations have all their "additional" reflection coefficients outside this interval (of course, they must have at least some of their "additional" reflection coefficients outside this interval); if this were true, the development of an efficient algorithm for higher order AR weight function models would be greatly facilitated. In general, it is considered that parametric weight function models provide the greatest hope for procedures yielding a flexible choice of weight function together with an efficient solution algorithm.
- Investigation of the appropriate selection of "snape functions" in connection with use of the weighted information formulation for vector quantization, see equations (4.79), (4.80), and (4.81). For speech analysis applications, it is expected that each shape

function as the power spectral response function of a bandpass filter with response characteristics similar to those filters found in "channel vocoder" systems.

- Performance evaluation in specific (speech analysis and other) applications using (global) measures appropriate to the particular application. In a voice communications system an appropriate measure may be the result of some formal subjective listening test. In a recognition system the recognition error rate may be an appropriate measure. Systems that predict stock market activity might measure overall investment performance.
- Extension of the formulation to problems of multidimensional spectral estimation.
- Use of the basic concepts/ideas of the weighted information formulation to develop a procedure treating the issues of limited data and noise corruption simultaneously, perhaps in combination with notions of Kalman filtering and the Burg algorithm.

CITED REFERENCES

- Robinson, E.A., "A Historical Perspective of Spectrum Estimation," IEEE Proc., Special Issue on Spectral Estimation, 70, 9, September, 1982.
- Schuster, A., "On the Investigation of Hidden Periodicities with Application to a Supposed 26-Day Period of Meterological Phenomena," Terr. Magnet., 3, 1898.
- 3. Yule, G.U., "On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers," Phil. Trans. Roy. Soc. London, Series A, 226, 1927.
- Walker, G., "On Periodicity in Series of Related Terms," Proc. Roy. Soc. London, Series A, 131, 1931.
- 5. Blackman, R.B. and Tukey, J.W., The Measurement of <u>Power Spectra From the Point of View of Communica-</u> <u>tions Engineering</u>, reprint of the two 1958 arti-<u>cles (B.S.T.J., 37)</u>, New York: Dover, 1959.
- 6. Fourier, Jean Baptiste Joseph, The Analytical Theory of Heat, reprint of the 1878 translation from his celebrated treatise (Theorie Analytique de la Chaleur. Paris: Didot, 1822.) elaborating his controversial 1807 manuscript, New York: Dover, 1955.
- 7. Pisarenko, V.F., "The Retrieval of Harmonics from a Covariance Function," Geophys. J. Roy. Astron. Soc., 33, 1973.
- 8. Jaynes, E.T., "Information Theory and Statistical Mechanics - Part I", Phys. Rev., 106, 1957.
- 9. _____, "Information Theory and Statistical Mechanics - Part II", Phys. Rev., 108, 1957.
- 10. _____, "On the Rationale of Maximum-Entropy Methods," IEEE Proc., Special Issue on Spectral Estimation, 70, 9, September, 1982.

11. Burg, J.P., "Maximum Entropy Spectral Analysis," reprint from the Proc. 37th meeting of the Soc. of Exploration Geophys. (1967), in <u>Modern Spectrum</u> <u>Analysis</u>, Childers, D.G., Ed., New York: IEEE Press, 1978.

- 12. Steinnardt, A.O., "An Optimization Theoretic Framework for Spectral Estimation," Ph.D. Thesis, U. of Colorado, 1983.
- 13. Shore, J.E., and Johnson, R.W., "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy," IEEE Trans. on IT, 26, 1, January, 1980.
- 14. Kullback, S., Information Theory and Statistics, New York: Wiley, 1959.
- 15. Shore, J.E., and Johnson, R.W., "Properties of Cross-Entropy Minimization," IEEE Trans. on IT, 27, 4, July, 1982.
- 16. Jeffreys, H., "An Invariant Form for the Prior Probability in Estimation Problems," Proc. Roy. Scc., A186, 1946.
- 17. Shore, J.E., "Minimum Cross-Entropy Spectral Analysis," IEEE Trans. on ASSP, 29, 2, April, 1981.
- 18. Akaike, H., "A New Look at the Statistical Model Identification," reprint from IEEE Trans. on AC, 19, 1974, in <u>Modern Spectrum Analysis</u>, Childers, D.G., Ed., New York: IEEE Press, 1978.
- Parzen, E., "Some Recent Advances in Time Series Modeling," reprint from IEEE Trans. on AC, 19, 1974 in Modern Spectrum Analysis, Childers, D.G., Ed., New York: IEEE Press, 1978.
- 20. _____, "Time Series Model Identification by Estimating Information," Texas A&M Research Foundation, Tech. Rept. N-35, November, 1982.
- 21. Markel, J.D., and Gray, A.H., Jr., Linear Prediction of Speech, New York: Springer-Verlag, 1976.
- 22. Makhoul, J., "Stable and Efficient Lattice Methods for Linear Prediction," reprinted from IEEE Trans. on ASSP, 25, 1977 in <u>Modern Spectrum Analysis</u>, Childers, D.G., Ed., New York: IEEE Press, 1978.
- 23. Burg, J.P., "A New Analysis Technique for Time Series Data," reprinted from a 1968 presentation at the NATO Advanced Study Institute on Signal Processing

in Modern Spectrum Analysis, Childers, D.G., Ed., New York: IEEE Press, 1978.

24. Itakura, F. and Saito, S., "Analysis Synthesis Telephony Based Upon the Maximum Likelihood Method," Reports of the Sixth Int. Cong. on Acoust., Y. Kohasi, Ed., Tokyo, 1968.

- 25. Robinson, E.A., and Treitel, S., Geophysical Signal Analysis, Englewood Cliffs: Prentice-Hall, 1980.
- 26. Kay, S.M., "Recursive Maximum Likelihood Estimation of Autoregressive Processes," IEEE Trans. on ASSP, 31, 1, February, 1983.
- 27. Pinsker, M.S., Information and Information Stability of Random Variables and Processes, translated from the original published in Moscow (1960), San Francisco: Holden Day, 1964.
- 28. Parzen, E., "Time Series Analysis for Models of Signals Plus White Noise," in <u>Spectral Analysis of Time</u> <u>Series</u>, Harris, B., Ed., New York: Wiley, 1966.
- 29. Pagano, M., "Estimation of Models of Autoregressive Signal Plus White Noise," Ann. of Stat., 2, 1, 1974.
- 30. Wiener, N., Extrapolation, Interpolation, and Smoothing of Stationary Time Series, New York: Wiley, 1949.
- 31. Kalman, R.E., "A New Approach to Linear Filtering and Prediction Problems," Trans. of ASME, J. of Basic Eng., March, 1960.
- 32. Lim, J.S., and Oppenheim, A.V., "Enhancement and Bandwidth Compression of Noisy Speech," reprinted from Proc. IEEE, 67, 12, 1979 in Speech Enhancement, Lim, J.S., Ed., Englewood Cliffs: Prentice-Hall, 1983.
- 33. Preuss, R.D., "A Frequency Domain Noise Cancelling Preprocessor for Narrowband Speech Communication Systems," Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, April, 1979.

1

34. Berouti, M., Schwartz, R., and Makhoul, J., "Enhancement of Speech Corrupted by Acoustic Noise," reprinted from Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, April, 1979 in Speech Enhancement, Lim, J.S., Ed., Englewood Cliffs: Prentice-Hall, 1983.

- 35. Boli, S.F., "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," reprinted from IEEE Trans. on ASSP, 27, 2, 1979 in Speech Enhancement, Lim, J.S., Ed., Englewood Cliffs: Prentice-Hall, 1983.
- 36. McAulay, R.J. and Malpass, M.L., "Speech Enhancement Using a Soft-Decision Noise Suppression Filter," reprinted from IEEE Trans. on ASSP, 28, 2, 1980 in Speech Enhancement, Lim, J.S., Ed., Englewood Cliffs: Prentice-Hall, 1983.
- 37. Widrow, B., et al., "Adaptive Noise Cancelling: Principles and Applications," reprinted from Proc. IEEE, 63, 12, 1975 in Speech Enhancement, Lim, J.S., Ed., Englewood Cliffs: Prentice-Hall, 1983.
- 38. Sambur, M.R., "Adaptive Noise Cancelling for Speech Signals," IEEE Trans. on ASSP, 26, 5, 1978.
- 39. Boll, S.F., and Pulsipher, D.C., "Suppression of Acoustic Noise in Speech Using Two Microphone Adaptive Noise Cancellation," reprinted from IEEE Trans. on ASSP, 28, 6, 1980 in Speech Enhancement, Lim, J.S., Ed., Englewood Cliffs: Prentice-Hall, 1983.
- 40. Rife, D.C. and Boorstyn, R.R., "Multiple tone parameter estimation from discrete time observations," Bell Syst. Tech. J., 55, 9, 1976.
- 41. Tufts, D.W. and Kumaresan, R., "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood," Proc. IEEE, Special Issue on Spectral Estimation, 70, 9, September, 1982.
- 42. Cadzow, J.A., Baseghi, B., and Hsu, T., "Singular-value decomposition approach to time series modelling," IEE Proc., Part F., Special issue on Spectral Analysis, 130, 3, 1983.
- 43. Cramer, H., <u>Mathematical Methods of Statistics</u>, Princeton: Princeton University Press, 1946.
- 44. Kaveh, M., and Bruzzone, S.P., "Statistical efficiency of correlation-based methods for ARMA spectral estimation," IEE Proc., Part F., Special issue on Spectral Analysis, 130, 3, 1983.
- 45. Quirk, M. and Liu, B., "Improving Resolution For Autoregressive Spectral Estimation By Decimation," IEEE Trans. on ASSP, 31, 3, June, 1983.

- 46. Nehorai, A. and Morf, M., "Enhancement of Jinusoids in Colored Noise and Whitening Performance of Exact Least Squares Predictors," IEEE Trans. on ASSP, 30, 3, June, 1982.
- 47. Wong, D.Y., Juang, B.H., and Gray, A.H., "An 800 bit/s Vector Quantization LPC Vocoder," IEEE Trans. on ASSP, 30, 5, 1982.
- 48. Abut, H., Gray, R.M., and Rebolledo, G., "Vector Quantization of Speech and Speech-Like Waveforms," IEEE Trans. on ASSP, 30, 3, 1982.
- 49. Gray, R.M., Gray, A.H., Rebolledo, G., and Shore, J.E., "Rate-distortion speech coding with a minimum discrimination information distortion measure," IEEE Trans. on IT, 27, 6, 1981.
- 50. Buzo, A., Gray, A.H., Gray, R.M., and Markel, J.D., "Speech Coding Based Upon Vector Quantization," IEEE Trans. on ASSP, 28, 5, 1980.
- 51. Linde, Y., Buzo, A., and Gray, R.M., "An Algorithm for Vector Quantizer Design," IEEE Trans. on COM, 28, 1, 1980.
- 52. Buzo de la Pena, L.A., "Optimal Vector Quantization for Linear Predictive Coded Speech," Ph.D. thesis, Stanford University, 1978.
- 53. Suresh Babu, B.N., Volz, B.E., Washburn, S.J., and Preuss, R.D., "Software for Vector Quantization," MTR-8414, Bedford, Mass.: The MITRE Corp., 1981.
- 54. Clapp, R.A.J., "Acoustic Segments in Natural Speech: Analysis and Statistics," MTR-8405, Bedford, Mass." The MITRE Corp., 1981.
- 55. Chu, P.L., "Frequency Weighted Linear Predictive Coding of Speech," Ph.D. Thesis, University of Calif., Berkeley, July, 1981.
- 56. _____, and Messerschmitt, D.G., "A Frequency Weighted Itakura-Saito Spectral Distance Measure," IEEE Trans. On ASSP, 30, 4, 1982.
- 57. Levinson, N., "The Wiener RMS Error Criteria in Filter Design and Prediction," J. Math. Phys., 25, pp. 261-78, 1947.
- 58. Durbin, J., "The Fitting of Time-Series Models," Rev. Inst. Int. Statist., 28, 3, pp. 233-43, 1960.

-

۲.

59. Burnside, W.S. and Panton, A.W., <u>The Theory of</u> <u>Equations</u>, 7th Ed., London: L. LEMBIS, Green, & M., 1912. Also available as a reprint, New York: Dover, 1960

- 60. Uspensky, J.V., <u>Theory of Equations</u>, New York: McGraw-Hill, 1948.
- 61. Collatz, L., <u>Functional Analysis and Numerical</u> <u>Mathematics</u>, translated from the Jerman by Oser, F., <u>New York:</u> Academic Press, 1905.
- 62. Hille, E., Analytic Function Theory, Vol. 1, Boston: Ginn and Company, 1959.

PART C

SUMMARY OF PUBLICATION ACTIVITIES

7

.

4

.

د و المراجع مع مع مراجع المراجع المراجع

PUBLICATION ACTIVITIES

In this part, a list of papers that are published and/or to be published is given along with a brief summary for each paper.

List of papers and their Summaries

 L. D. Hoy, D. L. Soldan, and R. Yarlagadda, "An Adaptive Approach to Narrowband Linear Predictive Coding of Speech", proc. of the Fifteenth Annual Asilomar Conference on Circuits, Systems and Computers, Pacific Grove CA, pp. 331-334, 1981.

Summary

Linear predictive coding is efficient an narrowband coding technique for speech signals but degrades significantly in the presence of noise. This paper examines a prefilter consisting of an adaptive digital predictor with pitch period delay. Preliminary results indicating the performance of two adaptive algorithms are presented. It is shown that the ADP can improve speech signal quality, as measured by signalto-noise ratios, when the speech is corrupted by The performance sensitivity to pitch wideband noise. period errors is also examined.

 D. L. Soldan and L. D. Hoy, "Pitch Extraction with an Adaptive Filter", MIDCON/82, Professional Program Session Record 3/2, Dallas, Texas, 1982.

Summary

This paper presents a technique of estimating the fundamental frequency or "pitch" of a voiced speech signal that is based on a tapped delay line adaptive digital filter (TDLADF). This method allows better resolution of the pitch frequency than traditional techniques such as autocorrelation and harmonic analysis. It also appears to have better noise tolerance than these techniques. Advances in VLSI design should allow real-time processing using the TDLADF in the future.

3. L. Hoy, B. Burns, D. Soldan and R. Yarlagadda, "Noise Suppression Methods for Speech Applications", Proc. of the 1985 Int. Conf. on Acoustics, Speech and Signal Processing, pp. 1133-1136, Boston, Mass., 1983.

Summary

This paper presents a discussion and evaluation of several filtering techniques for suppressing narrowband background noise in speech signals. The methods discussed are a modified spectral subtraction

technique, an inverse transform filter, an adaptive notch placement technique, an adaptive predictor, and a modification of the adaptive predictor. Performance of the filter methods are compared using a spectral error measurement and an area ratio parameter error measurement. Although the modified adaptive predictor provided the best improvement in spectral error, results indicate the modified spectral subtraction method to be the most suitable for use with linear predictive coding systems.

4. R. D. Preuss and R. Yarlagadda, "Autoregressive Spectral Estimation in Noise in the Context of Speech Analysis", to be presented at the Second ASSP Workshop on Spectral Estimation, Tampa, Florida, 1983.

Summary

An improved method of spectral estimation is described. The method treats the problem of estimating autoregressive (AR) process parameters from sequential discrete time observations corrupted by additive independent noise with known power spectral density. The method has a theoretical foundation relating it to principles of information theory as well as the linear predictive (LP) procedures popularly employed for speech analysis. Simulation results are used to support the theoretical development and demonstrate the advantages of the method as compared to currently popular methods of estimating speech spectra from noise corrupted observations.

5. C. S. Sims and R. Yarlagadda, "Design of Fast Recursive Estimators", to be published as a full paper in the <u>IEEE</u> <u>Trans. on Acoustics, Speech and Signal Processing</u>.

Ì

2

Summary

A recursive linear estimator is proposed for rapid estimation of a signal in noise. Efficient methods are developed for optimization of the filter coefficients. Optimal selection of data to be processed is shown to be related to a classic integer programming problem.

6. R. Yarlagadda and C. S. Sims, "A Note on the General Discrete-Time Linear Estimation", submitted for publication.

Summary

This paper presents a simple and an efficient algorithm for the solution of a generalized least squares prediction problem. The derivations are presented in terms of matrix point of view.

WARANANANANANANANANA

MISSION

of

Rome Air Development Center

RADC plans and executes research, development, test and selected acquisition programs in support of Command, Control Communications and Intelligence $(C^{3}I)$ activities. Technical and engineering support within areas of technical competence is provided to ESD Program Offices (POS) and other ESD elements. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, solid state sciences, electromagnetics and electronic reliability, maintainability and compatibility.

₢₳₣₡₲₣₡₲₽₡₲₣₡₲₣₡₲₣₡

Printed by United States Air Force Hanscom AFB, Mass. 01731



E N

11-85

DTIC