

Biomedical Signal Processing

Volume I Time and Frequency Domains Analysis

Author

Arnon Cohen, Ph.D.

Associate Professor

Departments of Biomedical Engineering and
Electrical and Computer Engineering

Head

Center for Biomedical Engineering
Ben Gurion University
Beer Sheva, Israel



CRC Press, Inc.
Boca Raton, Florida

Library of Congress Cataloging in Publication Data

Cohen, Arnon, 1938-

Biomedical signal processing.

Bibliography: p.

Includes index.

Contents: v. 1. Time and frequency domains analysis
— v. 2. Compression and automatic recognition.

1. Signal processing. 2. Biomedical engineering.
I. Title.

R857.S47C64 1986 610'.28 85-9626

ISBN 0-8493-5933-3 (v. 1)

ISBN 0-8493-5934-1 (v. 2)

This book represents information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Every reasonable effort has been made to give reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

All rights reserved. This book, or any parts thereof, may not be reproduced in any form without written consent from the publisher.

Direct all inquiries to CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, Florida, 33431.

© 1986 by CRC Press, Inc.

Second Printing, 1987

Third Printing, 1988

International Standard Book Number 0-8493-5933-3 (v. 1)

International Standard Book Number 0-8493-5934-1 (v. 2)

Library of Congress Card Number 85-9626

Printed in the United States

Library of Congress Cataloging in Publication Data

Cohen, Arnon, 1938-

Biomedical signal processing.

Bibliography: p.

Includes index.

Contents: v. 1. Time and frequency domains analysis
— v. 2. Compression and automatic recognition.

1. Signal processing. 2. Biomedical engineering.
I. Title.

R857.S47C64 1986 610'.28 85-9626

ISBN 0-8493-5933-3 (v. 1)

ISBN 0-8493-5934-1 (v. 2)

This book represents information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Every reasonable effort has been made to give reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

All rights reserved. This book, or any parts thereof, may not be reproduced in any form without written consent from the publisher.

Direct all inquiries to CRC Press, Inc., 2000 Corporate Blvd., N.W., Boca Raton, Florida, 33431.

© 1986 by CRC Press, Inc.

Second Printing, 1987

Third Printing, 1988

International Standard Book Number 0-8493-5933-3 (v. 1)

International Standard Book Number 0-8493-5934-1 (v. 2)

Library of Congress Card Number 85-9626

Printed in the United States

PREFACE

Biomedical signal processing is of prime importance not only to the physiological researcher but also to the clinician, the engineer, and the computer scientists who are required to interpret the signal and to design systems and algorithms for its manipulations.

The biomedical signal is, first of all, a signal. As such, its processing and analysis are covered by the numerous books and journals on general signal processing. Biomedical signals, however, possess many special properties and unique problems that render the need for special treatment.

Most of the material dealing with biomedical signal processing methods has been widely scattered in various scientific, technological, and physiological journals and in conference proceedings. Consequently, it is a rather difficult and time-consuming task, particularly to a newcomer to this field, to extract the subject matter from the scattered information.

This book was not meant to be another text or reference on general signal processing. It is intended to provide material of interest to engineers and scientists who wish to apply modern signal processing techniques to the analysis of biomedical signals. It is assumed the reader is familiar with the fundamentals of signals and systems analysis as well as the fundamentals of biological systems. Two chapters on basic digital and random signal processing have been included. These serve only as a summary of the material required as background for other material covered in the book.

The presentation of the material in the book follows the flow of events of the general signal processing system. After the signal has been acquired, some manipulations are applied in order to enhance the relevant information present in the signal. Simple, optimal, and adaptive filtering are examples of such manipulations. The detection of wavelets is of importance in biomedical signals; they can be detected from the enhanced signal by several methods. The signal very often contains redundancies. When effective storing, transmission, or automatic classification are required, these redundancies have to be extracted. The signal is then subjected to data reduction algorithms that allow the effective representation in terms of features. Methods for data reduction and features extraction are discussed. Finally, the topic of automatic classification is dealt with, in both the decision theoretic and the syntactic approaches.

The emphasis in this book has been placed on modern processing methods, some of which have been only slightly applied to biomedical data. The material is organized such that a method is presented and discussed, and examples of its application to biomedical signals are given. Rapid developments in digital hardware and in signal processing algorithms open new possibilities for the applications of sophisticated signal processing methods to biomedicine. Solutions that were cost prohibitive beforehand or impractical because of the lack of appropriate algorithm become available. In such a dynamic environment, the biomedical signal processing practitioner requires a book such as this one.

The author wishes to acknowledge the help received from many students and colleagues during the preparation of this book.

Arnon Cohen

TABLE OF CONTENTS

Volume I

Chapter 1

Introduction

I.	General Measurement and Diagnostic System.....	1
II.	Classification of Signals	3
III.	Fundamentals of Signal Processing	4
IV.	Biomedical Signal Acquisition and Processing.....	5
V.	The Book	6
	References.....	7

Chapter 2

The Origin of the Bioelectric Signal

I.	Introduction.....	9
II.	The Nerve Cell	9
	A. Introduction	9
	B. The Excitable Membrane	10
	C. Action Potential Initiation and Propagation.....	11
	D. The Synapse.....	11
III.	The Muscle	12
	A. Muscle Structure	12
	B. Muscle Contraction.....	12
IV.	Volume Conductors.....	12
	References.....	13

Chapter 3

Random Processes

I.	Introduction.....	15
II.	Elements of Probability Theory	15
	A. Introduction	15
	B. Joint Probabilities	16
	C. Statistically Independent Events	17
	D. Random Variables.....	17
	E. Probability Distribution Functions	18
	F. Probability Density Functions	19
III.	Random Signals Characterization	21
	A. Random Processes.....	21
	B. Statistical Averages (Expectations)	22
IV.	Correlation Analysis	23
	A. The Correlation Coefficient	23
	B. The Correlation Function	25
	C. Ergodicity	26
V.	The Gaussian Process.....	26
	A. The Central Limit Theorem	26
	B. Multivariate Gaussian Process	27
	References.....	28

Chapter 4

Digital Signal Processing

I.	Introduction	29
II.	Sampling.....	29
A.	Introduction	29
B.	Uniform Sampling.....	30
C.	Nonuniform Sampling	31
1.	Zero, First, and Second Order Adaptive Sampling.....	32
2.	Nonuniform Sampling with Run Length Encoding.....	34
III.	Quantization	36
A.	Introduction	36
B.	Zero Memory Quantization.....	36
C.	Analysis of Quantization Noise.....	39
D.	Rough Quantization	40
IV.	Discrete Methods	42
A.	The Z Transform	42
B.	Difference Equations	43
	References.....	44

Chapter 5

Finite Time Averaging

I.	Introduction	45
II.	Finite Time Estimation of the Mean Value	45
A.	The Continuous Case	45
1.	Short Observation Time.....	47
2.	Long Observation Time	48
B.	The Discrete Case	51
III.	Estimation of the Variance and Correlation	53
A.	Variance Estimation — The Continuous Case	53
B.	Variance Estimation — The Discrete Case	54
C.	Correlation Estimation.....	56
IV.	Synchronous Averaging (CAT-Computed Averaged Transients).....	56
A.	Introduction	56
B.	Statistically Independent Responses	58
C.	Totally Dependent Responses	59
D.	The General Case	60
E.	Records Alignment, Estimation of Latencies	61
	References.....	64

Chapter 6

Frequency Domain Analysis

I.	Introduction	65
A.	Frequency Domain Representation	65
B.	Some Properties of the Fourier Transform.....	65
1.	The Convolution Theorem	66
2.	Parseval's Theorem	66
3.	Fourier Transform of Periodic Signals	67
C.	Discrete and Fast Fourier Transforms (DFT, FFT)	68
II.	Spectral Analysis	71
A.	The Power Spectral Density Function	71
B.	Cross-Spectral Density and Coherence Functions.....	72

III.	Linear Filtering	73
A.	Introduction	73
B.	Digital Filters.....	74
C.	The Wiener Filter	74
IV.	Cepstral Analysis and Homomorphic Filtering	76
A.	Introduction	76
B.	The Cepstra	76
C.	Homomorphic Filtering.....	77
	References.....	80

Chapter 7

Time Series Analysis-Linear Prediction

I.	Introduction	81
II.	Autoregressive (AR) Models.....	85
A.	Introduction	85
B.	Estimation of AR Parameters — Least Squares Method.....	85
III.	Moving Average (MA) Models	89
A.	Autocorrelation Function of MA Process	89
B.	Iterative Estimate of the MA Parameters	89
IV.	Mixed Autoregressive Moving Average (ARMA) Models	90
A.	Introduction	90
B.	Parameter Estimation of ARMA Models — Direct Method	90
C.	Parameter Estimation of ARMA Models — Maximum Likelihood Method	93
V.	Process Order Estimation	95
A.	Introduction	95
B.	Residuals Flatness	95
C.	Final Prediction Error (FPE).....	96
D.	Akaike Information Theoretic Criterion (AIC).....	97
E.	Ill Conditioning of Correlation Matrix	98
VI.	Lattice Representation	98
VII.	Nonstationary Processes	99
A.	Trend Nonstationarity — ARIMA.....	99
B.	Seasonal Processes	101
VIII.	Adaptive Segmentation	101
A.	Introduction	101
B.	The Autocorrelation Measure (ACM) Method	102
C.	Spectral Error Measure (SEM) Method.....	103
D.	Other Segmentation Methods.....	105
	References.....	106

Chapter 8

Spectral Estimation

I.	Introduction	109
II.	Methods Based on the Fourier Transform.....	110
A.	Introduction	110
B.	The Blackman-Tukey Method	111
C.	The Periodogram	112
1.	Introduction	112
2.	The Expected Value of the Periodogram	114
3.	Variance of the Periodogram.....	116

4.	Weighted Overlapped Segment Averaging (WOSA)	117
5.	Smoothing the Periodogram.....	119
III.	Maximum Entropy Method (MEM) and the AR Method.....	122
IV.	The Moving Average (MA) Method	125
V.	Autoregressive Moving Average (ARMA) Methods	126
	A. The General Case	126
	B. Pisarenko's Harmonic Decomposition (PHD).....	127
	C. Prony's Method	130
VI.	Maximum Likelihood Method (MLM) — Capon's Spectral Estimation.....	133
VII.	Discussion and Comparison of Several Methods.....	134
	References.....	137
 Chapter 9		
Adaptive Filtering		
I.	Introduction	141
II.	General Structure of Adaptive Filters	142
	A. Introduction	142
	B. Adaptive System Parameter Identification	142
	C. Adaptive Signal Estimation.....	142
	D. Adaptive Signal Correction.....	143
III.	Least Mean Squares (LMS) Adaptive Filter.....	143
	A. Introduction	143
	B. Adaptive Linear Combiner	144
	C. The LMS Adaptive Algorithm.....	145
	D. The LMS Adaptive Filter.....	147
IV.	Adaptive Noise Cancelling.....	147
	A. Introduction	147
	B. Noise Canceller with Reference Input	148
	C. Noise Canceller without Reference Input.....	153
	D. Adaptive Line Enhancer (ALE)	154
V.	Improved Adaptive Filtering	154
	A. Multichannel Adaptive Signal Enhancement.....	154
	B. Time-Sequenced Adaptive Filtering	156
	References.....	158
	Index	161

TABLE OF CONTENTS

Volume II

 Chapter 1		
Wavelet Detection		
I.	Introduction.....	1
II.	Detection by Structural Features	2
	A. Simple Structural Algorithms.....	2
	B. Contour Limiting.....	5
III.	Matched Filtering	6
IV.	Adaptive Wavelet Detection	9
	A. Introduction	9
	B. Template Adaptation	10

C.	Tracking a Slowly Changing Wavelet	12
D.	Correction of Initial Template	12
V.	Detection of Overlapping Wavelets	14
A.	Statement of the Problem	14
B.	Initial Detection and Composite Hypothesis Formulation	15
C.	Error Criterion and Minimization	16
	References	17

Chapter 2

Point Processes

I.	Introduction	19
II.	Statistical Preliminaries	20
III.	Spectral Analysis	24
A.	Introduction	24
B.	Interevent Intervals Spectral Analysis	24
C.	Counts Spectral Analysis	25
IV.	Some Commonly Used Models	26
A.	Introduction	26
B.	Renewal Processes	26
1.	Serial Correlogram	27
2.	Flatness of Spectrum	27
3.	A Nonparametric Trend Test	28
C.	Poisson Processes	28
D.	Other Distributions	31
1.	The Weibull Distribution	31
2.	The Erlang (Gamma) Distribution	32
3.	Exponential Autoregressive Moving Average (EARMA)	32
4.	Semi-Markov Processes	32
V.	Multivariate Point Processes	33
A.	Introduction	33
B.	Characterization of Multivariate Point Processes	33
C.	Marked Processes	35
	References	35

Chapter 3

Signal Classification and Recognition

I.	Introduction	37
II.	Statistical Signal Classification	39
A.	Introduction	39
B.	Bayes Decision Theory and Classification	39
C.	k-Nearest Neighbor (k-NN) Classification	50
III.	Linear Discriminant Functions	53
A.	Introduction	53
B.	Generalized Linear Discriminant Functions	55
C.	Minimum Squared Error Method	56
D.	Minimum Distance Classifiers	58
E.	Entropy Criteria Methods	60
1.	Introduction	60
2.	Minimization of Entropy	60
3.	Maximization of Entropy	62
IV.	Fisher's Linear Discriminant	63

V.	Karhunen-Loeve Expansions (KLE).....	66
A.	Introduction	66
B.	Karhunen-Loeve Transformation (KLT) — Principal Components Analysis (PCA)	67
C.	Singular Value Decomposition (SVD).....	69
VI.	Direct Feature Selection and Ordering	75
A.	Introduction	75
B.	The Divergence	76
C.	Dynamical Programming Methods	77
VII.	Time Warping	79
	References.....	84

Chapter 4

Syntactic Methods

I.	Introduction	87
II.	Basic Definitions of Formal Languages	89
III.	Syntactic Recognizers.....	92
	A. Introduction	92
	B. Finite State Automata	92
	C. Context-Free Push-Down Automata (PDA).....	95
	D. Simple Syntax-Directed Translation	100
	E. Parsing	100
IV.	Stochastic Languages and Syntax Analysis	101
	A. Introduction	101
	B. Stochastic Recognizers	102
V.	Grammatical Inference	104
VI.	Examples	104
	A. Syntactic Analysis of Carotid Blood Pressure	104
	B. Syntactic Analysis of ECG	106
	C. Syntactic Analysis of EEG	110
	References.....	111

Appendix A

Characteristics of Some Dynamic Biomedical Signals

I.	Introduction	113
II.	Bioelectric Signals	113
	A. Action Potential	113
	B. Electroneurogram (ENG)	113
	C. Electroretinogram (ERG)	113
	D. Electro-Oculogram (EOG).....	114
	E. Electroencephalogram (EEG).....	114
	F. Evoked Potentials (EP)	117
	G. Electromyography (EMG)	119
	H. Electrocardiography (ECG, EKG).....	121
	1. The Signal	121
	2. High-Frequency Electrocardiography	124
	3. Fetal Electrocardiography (FECG)	124
	4. His Bundle Electrocardiography (HBE).....	124
	5. Vector Electrocardiography (VCG)	124
I.	Electrogastrography (EGG).....	124
J.	Galvanic Skin Reflex (GSR), Electrodermal Response (EDR).....	125

III.	Impedance	125
A.	Bioimpedance	125
B.	Impedance Plethysmography	126
C.	Rheoencephalography (REG).....	126
D.	Impedance Pneumography.....	126
E.	Impedance Oculography (ZOG)	126
F.	Electroglottography.....	126
IV.	Acoustical Signals	126
A.	Phonocardiography	126
1.	The First Heart Sound	126
2.	The Second Heart Sound	127
3.	The Third Heart Sound	127
4.	The Fourth Heart Sound	127
5.	Abnormalities of the Heart Sound	127
B.	Auscultation	127
C.	Voice.....	128
D.	Korotkoff Sounds	129
V.	Mechanical Signals	130
A.	Pressure Signals	130
B.	Apexcardiography (ACG)	130
C.	Pneumotachography	130
D.	Dye and Thermal Dilution.....	130
E.	Fetal Movements	131
VI.	Biomagnetic Signals	131
A.	Magnetoencephalography (MEG).....	131
B.	Magnetocardiography (MCG)	131
C.	Magnetopneumography (MPG).....	131
VII.	Biochemical Signals	131
VIII.	Two-Dimensional Signals	132
	References.....	134

Appendix B

Data Lag Windows

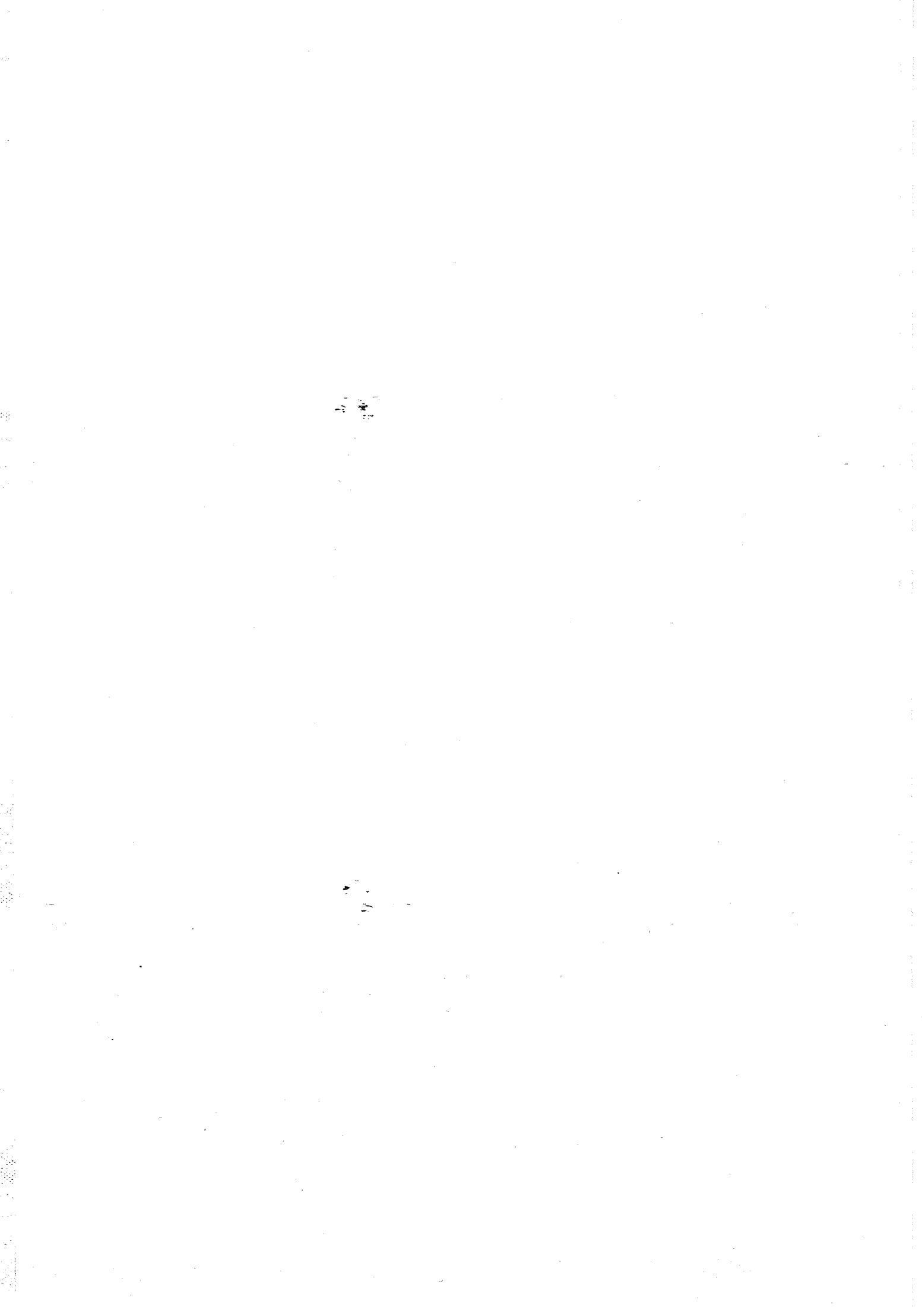
I.	Introduction	139
II.	Some Classical Windows	139
A.	Introduction	139
B.	Rectangular (Dirichlet) Window.....	140
C.	Triangle (Bartlett) Window	140
D.	Cosine _a Windows	141
E.	Hamming Window	143
F.	Dolph-Chebyshev Window	145
	References.....	151

Appendix C

Computer Programs

I.	Introduction	153
II.	Main Programs	154
•	NUSAMP (Nonuniform Sampling)	154
•	SEGMNT (Adaptive Segmentation)	158
•	PERSPT (Periodogram Power Spectral Density Estimation)	162
•	WOSA (WOSA Power Spectral Density Estimation)	162

●	MEMSPT (Maximum Entropy [MEM] Power Spectral Density Estimation)	165
●	NOICAN (Adaptive Noise Cancelling)	167
●	CONLIM (Wavelet Detection by the Contour Limiting Method)	169
●	COMPRS (Reduction of Signal Dimensionality by Three Methods: Karhunen-Loeve [KL], Entropy [ENT], and Fisher Discriminant [FI])	171
III.	Subroutines	174
●	LMS (Adaptive Linear Combiner, Widow's Algorithm).....	174
●	N ^A C ^R (Normalized Autocorrelation Sequence).....	175
●	DLPC (LPC, PARCOR, and Prediction Error of AR Model of Order P)	176
●	DLPC 20 (LPC, PARCOR, and Prediction Error of All AR Models of Order 2 to 20).....	177
●	FTO1A (Fast Fourier Transform [FFT]).....	178
●	XTERM (Maximum and Minimum Values of a Vector)	180
●	ADD (Addition and Subtraction of Matrices).....	180
●	MUL (Matrix Multiplication).....	181
●	MEAN (Mean of a Set of Vectors).....	181
●	COVA (Covariance Matrix of a Cluster of Fectors)	182
●	INVER (Inversion of a Real Symetric Matrix)	183
●	SYMINV (Inversion of a Real Symetric Matrix, [Original Matrix Destroyed]).....	183
●	RFILE (Read Data Vector From Unformatted File)	185
●	WFILE (Write Data Vector on Unformatted File)	186
●	RFILEM (Read Data Matrix From Unformatted File)	187
●	WFILEM (Write Data Matrix on Unformatted File).....	188
	Index	189



Chapter 1

INTRODUCTION

I. GENERAL MEASUREMENT AND DIAGNOSTIC SYSTEM

This book is concerned with the analysis and processing of biomedical signals. It is pertinent to discuss first what, in general, is a signal, what is a biomedical signal, and why process it. A discussion on these general topics is presented in this chapter.

A signal is a means to convey information. It is sometimes generated directly by the original information source. We may then want to learn about the structure or functioning of the source from the extracted information (the signal). The signal available may not yield directly the required information. We then apply some operations on the signal in order to enhance the information needed. This may be the case, for example, when the visual processing mechanism of the brain is of interest. We may present the eye with a flash and monitor the activity of the brain by means of electrodes located on the scalp. We shall find that the required information that is related to the visual activity of the brain is "buried" in the signal which is mainly due to other activities of the brain. Special processing procedures must be applied to the signal, so as to enhance the relevant information. We may want to transmit the signal from point of acquisition to a remote location for monitoring or processing. This may be the case, for example, in the intensive care unit when information on patients is required at the central monitoring station, or when the information concerning a patient at home is required in the hospital or physicians office. In these cases, the processing of the signal is required in order to match it with the requirement of the transmission channel. In other cases, the information is to be stored for later use. Effective storing is needed such that it will require minimum amount of storing space (computer memory, magnetic tapes) and can be later reconstructed at will.

The general measurement and diagnostic system is schematically shown in Figure 1. Usually this system consists of a transducer that is coupled to the information source and extracts the required information. The original information may be in a form that is not suitable for processing, storing, or transmitting (pressure, temperature). The transducer converts the information into (most often) an electrical signal. With current technology, this type of signal is most convenient for the above tasks. The processing of the signal is often required for diagnosis as well. In this case, the processing has to classify the signal into one of many given classes which may be the normal and various abnormal classes. After classification, corrective measures that change the source may be taken.

Several books are available on the topic of biomedical instrumentation and measurements.¹⁻⁵ They deal with the various transducers and the hardware associated with the acquisition and basic preprocessing. In this book, the various topics regarding processing are discussed. The basic preprocessing is sometimes called "signal conditioning" while later processing is sometimes known as "signal manipulation and evaluation". Since no distinct definitions exist we prefer the general term "processing".

The topics discussed in this book are depicted in broken lines in Figure 1 and in more detail in Figure 2.

The first step in the processing is usually that of segmentation. The signal may drastically change its properties during time. We then observe and process the signal only in a finite time window. The length of the time window depends on the signal source and goal of processing. We may use a single "window" with predetermined length as we do, for example, when electrocardiographic monitoring is performed or we may require some scheme for automatically dividing the signal into varying length segments as is often done in electroencephalography.

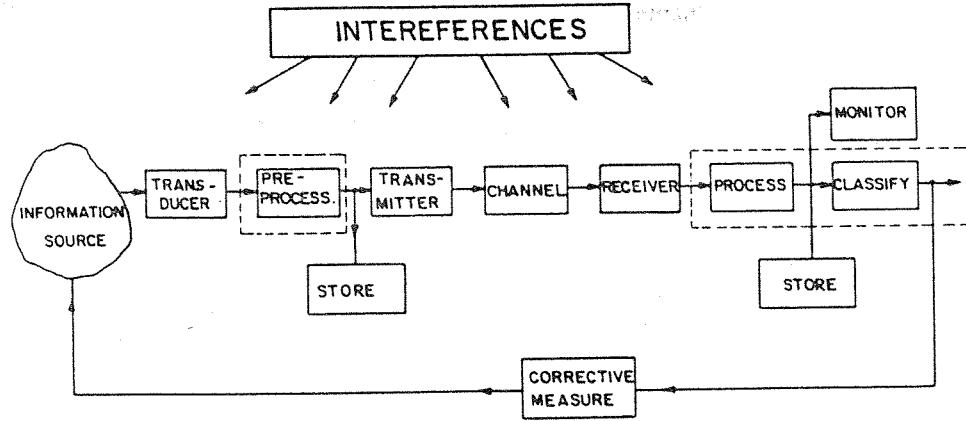


FIGURE 1. General measurement and diagnostic system.

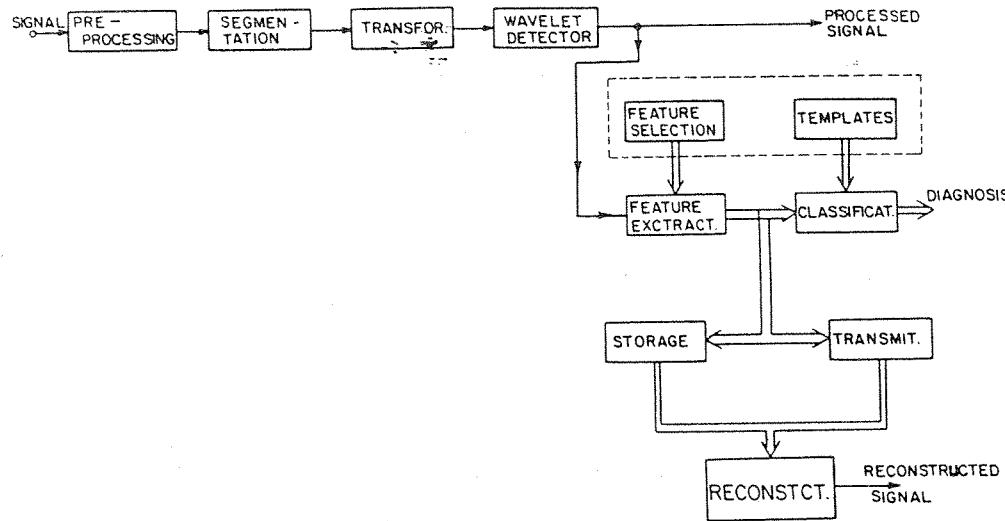


FIGURE 2. Signal processing.

A variety of methods are available for the enhancement of the relevant information in the signal. The signal is either corrupted with additive and multiplicative noise or the information required constitutes only a part of the signal such that irrelevant portions are considered noise. We then apply noise-attenuating-and-cancelling techniques or signal enhancement methods in order to increase the signal-to-noise ratio. To do this, some *a priori* knowledge on the signal and the noise is required. The more *a priori* knowledge that is available, the better the processing. Enhancement methods that are optimal in some sense are discussed as well as adaptive methods that automatically adjust themselves to varying conditions.

Very often the relevant information in the signal possesses some waveshape which is only generally known. A good example is the electrocardiogram where the general shape of the PQRST complex is known. It is required to extract the exact shape of the wavelet present in the signal.

Not all the information conveyed by the signal is of interest. The signal may contain redundancies. When effective storing and transmission are required, or when the signal is to be automatically classified, these redundancies have to be eliminated. The signal can be represented by a set of features that contain the required information. These features are

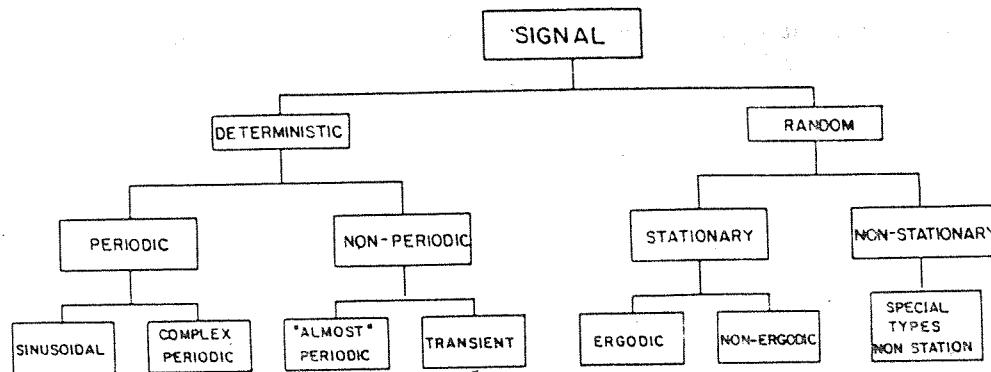


FIGURE 3. Classification of signals.

then used for storage, transmission, and classification. Reconstruction of the signal from its features is often needed. The types of features used and their number dictate, on one hand, the data reduction rate for efficient storing and transmitting and, on the other hand, the error of reconstruction.

The various functions depicted in the blocks of Figure 2 are discussed in the forthcoming chapters. The processing system of Figure 2 is a general system for signal processing independent of application. Geophysical signals,⁶ for example, as well as biomedical signals face the same general steps of processing. In this book, however, the general topics are discussed with emphasis on biomedical signal processing. Additional topics have been introduced, which are specifically oriented to biomedical signals.

II. CLASSIFICATION OF SIGNALS

Signals, extracted from biological and physical systems, may possess various properties and characteristics. It is important to identify the general characteristics of the signal, so that the appropriate processing tools be applied. The various types of signals are introduced in this section.

Signals are classified into two main groups: deterministic and random signals (see Figure 3). Deterministic signals are those that can be described by explicit mathematical relationships. Random signals cannot be exactly expressed. It can be described only in terms of probabilities and statistical averages. A philosophical question may be raised as to whether a random or deterministic signal exists. In reality, we are not able to find a signal that can be accurately predicted by means of an exact mathematical formulation. Even a sine wave from a signal generator is not deterministic in that sense, since no one can tell when a power failure will cause the sine wave to completely disappear or some generator malfunction will cause its shape to change. On the other hand, it can be argued that random signals, in reality, do not exist. Any signal is the result of some physical or chemical phenomena and is governed by some laws. If these laws were completely known to us, we could have exactly expressed the signal and predict its value. Since our interest here is to apply processing methods for the analysis of the signals, we do not have to enter into these philosophical arguments. It is the goals and constraints of the problem at hand that will dictate our decision to consider a given signal as random or deterministic.

When analyzing the electrocardiographic (ECG) signal, for example, we may be interested in the general characteristics of the QRS complex and thus consider the signal deterministic, or we may be interested in the changes of the R-R interval, thereby considering it a random signal.

Deterministic signals are divided into two subgroups: periodic and nonperiodic signals.

Periodic signals are signals for which $x(t) = x(t + T)$, where T is the period. Periodic signals are convenient since one period is sufficient for complete description. In the frequency domain, the description is given by means of the Fourier series, where only the fundamental frequency and its harmonics take part. Nonperiodic signals consist of two classes. "Almost" periodic signals are those that are not periodic in the mathematical sense but have discrete description in the frequency domain. This frequency description differs from the periodic one in that the various frequencies participating are not harmonics of some fundamental frequency. A combination of several unrelated periodic signals creates an "almost" periodic signal.

A transient signal is a deterministic signal not having the properties discussed previously.

Random signals are much more difficult to deal with. A random signal is a sample function of a random process. One sample function of a random process differs from another in their time description. They possess, however, the same statistical properties. The complete (infinite) set of sample functions produced by the random process is called the *ensemble*. The description of the random signal is given by the joint probability density function.

A stationary process is a process the statistical properties of which are not a function of time. For such a process we can calculate, for example, the expectation by averaging the values, $x(t)$, over all the ensemble at any time, t . An important class of random signals is the class of *ergodic signals*. For these signals, statistical averaging over the ensemble equals time averaging over the time axis of any one sample function.

We shall see that stationarity and ergodicity are properties which allow the use of practical processing methods. A process which is nonstationary (and thus nonergodic) is very difficult to process. Very often we are forced to assume the process is ergodic even though it is known *a priori* that the assumption is false.

When processing the electroencephalographic (EEG) signal, for example, we do not have at our disposal the complete ensemble. We have only one sample function. We are thus forced to assume ergodicity and estimate the required statistical properties from time (rather than ensemble) averages. Since the tools for processing nonstationary signals are not very effective, we are often dividing a nonstationary signal into segments, each assumed to be stationary. The length of the segments depends on the properties of the nonstationarities. In speech signals, segments are chosen with durations of about 10 msec while in EEG analysis, segments may be of the order of a few seconds.

Another basis for signal classification, which has great significance from processing point of view, is that of continuous vs. discrete signals. Continuous time signals are signals which, in general, are defined at any point in time. The tools that are applied to the processing of these signals are the Fourier and Laplace transforms and other "analog" methods. In terms of hardware, these signals are treated by analog systems (filters, amplifiers, computers). Discrete signals are signals that are defined only at given points in time. Usually these signals are also "sampled" in amplitude. We usually think of discrete signals as the result of continuous signals that had been time sampled and amplitude quantized, though there may be signals which are discrete by nature. These signals are processed by means of discrete signal processing methods such as the Z transform and the Discrete Fourier Transform (DFT). In terms of hardware, these signals are treated by means of digital systems including digital computers. The advances in digital technology in recent years have created a situation in which most of the signal processing activity is in discrete signals.⁷⁻¹¹

III. FUNDAMENTALS OF SIGNAL PROCESSING

A very basic tool for signal processing is that of filtering. Its use is best applied in the frequency rather than in the time domain.

The Fourier transform¹² is an operator that transfers a signal, $x(t)$, in the time domain into the frequency domain. In the frequency domain, the signal is represented in terms of

its amplitude and phase as a function of frequency. A more practical transformation is into the complex frequency, S, plane. The Laplace transform is a transformation that can transfer a signal $x(t)$ into the complex frequency plane. The spectral properties of the signal can be shaped by means of filters which are designed to attenuate or completely cut off portions of the signals' frequencies.

The same tools exist for discrete signals in which the complex Z domain is defined. Discrete filters can be designed to shape the spectrum of the discrete signal as required.

The fundamentals of continuous and discrete signal processing as well as filter design theory are not discussed in this book. It is assumed that the reader is at least familiar with the basics of signal filtering.

The frequency filtering is a powerful tool for random as well as deterministic signals. When processing random signals, we apply the Fourier transform to the autocorrelation function rather than the sample function itself. We then deal with the power spectral density function. The same filter design techniques are applied here.

IV. BIOMEDICAL SIGNAL ACQUISITION AND PROCESSING

When performing a measurement from a living biological substance, several unique problems arise that deserve special discussion, in addition to the one presented in Section I of this chapter.

The living biological system is a very complex system governed by biochemical, physical, and chemical laws not well understood as of yet. In particular, many aspects of the complex hierarchical control, the genetic control, the neural information transfer and processing, and other systems are still under extensive investigation. Very often we use *a priori* information concerning the system, generating the signal of interest, in order to help us in the analysis and processing procedures. When the underlying mechanism is not well understood, the processing may become less effective.

The complexity of the biological system often introduces difficulties in the measurement and processing procedures. Unlike physical systems, the biological system (most often) cannot be uncoupled in such a way that subsystems can be monitored and investigated individually. Because of the complex hierarchical control linkages among subsystems and due to many feedback paths not always well understood, the biological system under investigation must remain in its natural environment during observation. The signals produced by the system are thus influenced directly by the activity of the surrounding systems. The signal is also inherently contaminated by noise produced by the neighboring systems.

When monitoring, for example, the activity of the visual processing mechanism of the brain by means of the visual evoked potential, we cannot isolate the visual system and perform the clinical test under controlled conditions. The eye, its position, and its sensitivity are controlled by voluntary and nonvoluntary actions of the brain. The visual processing of the brain itself is dependent on many other surrounding activities, most of them beyond our control. The result is that very often we do not know exactly under what conditions the signal was taken and how to interpret the analysis of the signal — whether to attribute it to some significant phenomenon in the underlying mechanism (abnormality in system function) or to some change in measurement conditions.

The large variations that exist in biomedical signals force us to rely heavily on statistical methods. These variations exist in signals acquired from the same individual, and of course among populations. Thus, the accuracies and confidence limits that come out of our biomedical signal processing are usually not very high, at least in terms used in other engineering disciplines.

Biomedical signals are usually extracted from living organisms and, in most applications, from human beings. The measurement system must be designed so as not to damage the

system and not to cause pain whenever possible. *Noninvasive* techniques are always preferable. This means that very often we cannot get the information required directly and we have to infer it from signals that are noninvasively available. Fetal heart monitoring may serve as an example. Rather than apply electrodes directly to the fetal's skin, a procedure requiring invasive methods, we place the electrodes on the mother's abdomen. The signal thus acquired is heavily contaminated with the mother's strong ECG and other muscle activities. The inferring of the fetal ECG from the signal requires a lot of processing efforts.

In many important biomedical applications, the signals of interest are two dimensional.¹³ Conventional X-ray analysis, computer tomography (CT), nuclear magnetic resonance (NMR) imaging, and ultrasonic imaging are examples. In principle, the processing techniques of one- and two-dimensional signals are similar. In practice, however, there are major differences. Sophisticated algorithms for two dimensional signal processing have been developed. These are not discussed here.

Biomedical signals are mechanical, chemical, or electromagnetic in nature. With current technology, almost all transducers used provide electrical output so that the signal to be processed is presented as an electrical signal. One very important group of biomedical signals are signals which are electromagnetical in nature — the bioelectric signals. Chapter 2 is dedicated to the origin and characteristics of these signals.

V. THE BOOK

The material covered in this book follows approximately the main blocks depicted in Figure 2. As mentioned previously, Chapter 2 discusses the origin and characteristics of the bioelectric signal. It is not intended to replace or compete with the excellent books available on the topic. It was felt, however, that because of the importance of these signals, one chapter should be devoted and addressed especially to the engineer or computer scientist who has not been extensively exposed to the biological basis of the signals. Chapters 3 and 4 discuss basic principles of random and digital signal processing. The experienced engineer, and especially the one whose field is measurements and signal processing, will be familiar with the material presented in these chapters. Others, whose expertise lies elsewhere, will find Chapters 3 and 4 a basic reference for two very important topics. The problem of finite time observation records presents itself in almost every signal processing problem. Either the time available for measurement is finite or the signal is nonstationary and short segments of it are used in order to apply stationary processing methods. In both cases, we are faced with the problem of having a finite time observation record (a windowed record) from which we have to estimate the characteristics of the process. Chapter 5 presents and discusses the problems of finite time estimations in continuous and digital signals.

Frequency domain analysis techniques are presented in Chapter 6. This chapter also includes a discussion on cepstral methods and homomorphic filtering which are probably less familiar even to the experienced engineer.

Time series analysis has been originated in the statistical literature and has become an important analysis approach in modern signal processing. The basic idea here is to consider the signal an output of a linear system, whose (inaccessible) input is white in spectrum. Rather than deal with the signal itself, we want to deal with the system that generates it. It turns out that the representation of the signal by means of the coefficients of the system's differential (or difference) equation, or the system's poles and zeroes is a very efficient representation. The features achieved by time series analysis provide effective data reduction for signal storage and transmission as well as signal recognition. Chapter 7 presents the theory and practical algorithms for the implementation of time series analysis.

A related topic, with great importance in random signal processing, is that of spectral estimation. It is presented in Chapter 8, with algorithms that are designed for various types of signals.

One of the problems associated with biomedical signal processing is the complexity of the signal and the lack of *a priori* information that can be used to simplify the processing methods. Adaptive filtering methods have thus found wide applications in biomedical signal processing. Here, one requires a minimum *a priori* information on the signal and noise involved. The adaptive filter "learns" from the running data the required information and automatically adjusts its parameters in some optimal fashion. Chapter 9 deals with adaptive filtering. The discussion is primarily based on Widrow's work, who has also suggested many biomedical applications.

In many biomedical applications, the signal contains a wavelet which is of interest. This may be the case in electrocardiographic analysis where the QRS complex is sought, or in some cases of EEG analysis where K-complexes or spindles are of interest. Methods for wavelet detection are discussed in Chapter 1, Volume II.

Chapter 2 (Volume II) deals with point processes analysis. Point processes are processes which are described by the time of appearance of events. The parameter of interest is appearance time and not the shape of the event. The major application of this theory is in the analysis of neurosignals. Other applications such as R-R interval analysis in ECG or pitch analysis in speech are also presented.

Chapters 3 and 4 (Volume II) deal with the important problem of automatic signal recognition and classification. The more known approach, the *theoretic decision* approach, is presented in Chapter 3 (Volume II) while the more novel approach, that of *syntactic analysis*, is presented in Chapter 4 (Volume II).

A discussion on some of the many biomedical signals appears in the appendices. The aim of the appendices is not to give an exhaustive survey of all signals used in biomedicine, but rather to list a few representative ones with their main characteristics.

Throughout the book, examples and references to various biomedical applications have been provided. The emphasis, however, is on the presentation of various methods (some of them novel which have yet to be massively applied to biomedicine). The framework of the presentation is that depicted in Figure 2 for a general signal processing system.

REFERENCES

1. DeMarre, D. A. and Michales, D., *Bioelectronic Measurements*, Prentice-Hall, Englewood Cliffs, N.J., 1983.
2. Geddes, L. A. and Baker, L. E., *Principles of Applied Biomedical Instrumentation*, John Wiley & Sons, New York, 1968.
3. Strong, P., *Biophysical Measurements*, Tektronix, Beaverton, Or., 1970.
4. Webster, J. G., Ed., *Medical Instrumentation Application and Design*, Houghton Mifflin, Boston, 1978.
5. Cromwell, L., Weibell, F. J., Pfeiffer, E. A., and Usselmann, L. B., *Biomedical Instrumentation and Measurements*, Prentice-Hall, Englewood Cliffs, N.J., 1973.
6. Robinson, E. A. and Treitel, S., *Geophysical Signal Analysis*, Prentice-Hall, Englewood Cliffs, N.J., 1980.
7. Beauchamp, K. G. and Yuen, C. K., *Digital Methods for Signal Analysis*, George Allen and Unwin, Ltd., London, 1979.
8. Gold, B. and Rader, C. M., *Digital Processing of Signals*, McGraw-Hill, New York, 1969.
9. Oppenheim, A. V., Ed., *Application of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1978.
10. Chen, C. T., *One-Dimensional Digital Signal Processing*, Marcel Dekker, New York, 1979.
11. Tretter, S. A., *Introduction to Discrete-Time Signal Processing*, John Wiley & Sons, New York, 1976.
12. Bracewell, R. N., *The Fourier Transform and its Applications*, McGraw-Hill, New York, 1978.
13. Reichenberger, H. and Pfeiler, M., Objectives and approaches in biomedical signal processing, in *Signal Processing II: Theories and Applications*, Shussler, H. W., Ed., Elsevier/North Holland, Amsterdam, 1983.

White A

Chapter 2

THE ORIGIN OF THE BIOELECTRIC SIGNAL

I. INTRODUCTION

The most important information processing mechanism in the living biological system is the neural network.¹ The biological system has several means for information transfer. Probably the most important is the neural information transfer. Neurophysiology, the study of neural function, has been the key field for understanding internal communication and control in the biological system. Basic and applied neurophysiological research heavily relies on our ability to measure chemical and electrochemical activities taking place in the single cell, or collectively by groups of cells.

Many functions of neural and muscular cells are chemical in nature. These functions, however, produce changes in the electric field which can be monitored by electrodes. The so-called *bioelectric* potentials help the neurophysiologist study cell function. Direct measurements of the chemical phenomena, e.g., ion concentration changes,² can be performed by means of special transducers (ion selective electrodes, for example). However, these measurements are more difficult to perform.

6285
The source of the bioelectric signal is the single neural or muscular cell. These, however, do not function alone but in large groups. The accumulated effects of all active cells in the vicinity produce an electric field which propagates in the volume conductor³ consisting of the various tissues of the body. The activity of a muscle, or some neural network, can thus indirectly be measured by means of electrodes placed, say, on the skin. The acquisition of this type of information is easy; electrodes can be conveniently placed on the skin. The information, however, is difficult to analyze. It is the result of all neural and muscular activity in unknown locations transmitted through an inhomogeneous medium. In spite of these difficulties, electrical signals, monitored on the skin surface, are of enormous clinical and physiological importance. Electroencephalographic (EEG), electrocardiographic (ECG), electromyographic (EMG), and other such signals are routinely used for the diagnosis of neural and muscular systems in the clinic. The interpretation of the information is based mainly on the large statistical experience collected throughout the years.

This chapter explains the basic bioelectric phenomena on the cell level both in neural cells and in muscle cells. A brief discussion on the volume conductor problem is then presented to provide a link to the gross surface electric signals.

II. THE NERVE CELL

A. Introduction

The basic processing unit in the neurophysiological system is the nerve cell — the *neuron*. Its task is information processing, transfer, and acquisition. Neurons which are used for information transfer are usually long, and serve to transmit information to and from the central processing body. Special nerve cells have been evolved that serve as sensors.

A variety of mechanisms and sensor shapes exist to transduce many kinds of stimuli (pressure, light, temperature, etc.) into electrical and chemical signals. The central nervous system deals with the task of information processing and control. Though there are many types of neurons, the basic structure of these cells can be generally discussed. Figure 1 schematically depicts the structure of a nerve cell.

The important parts of the neuron are the cell body (*soma*), the *dendrites*, and the *axon*. The cell body consists of the intracellular fluid with the various bodies required for the

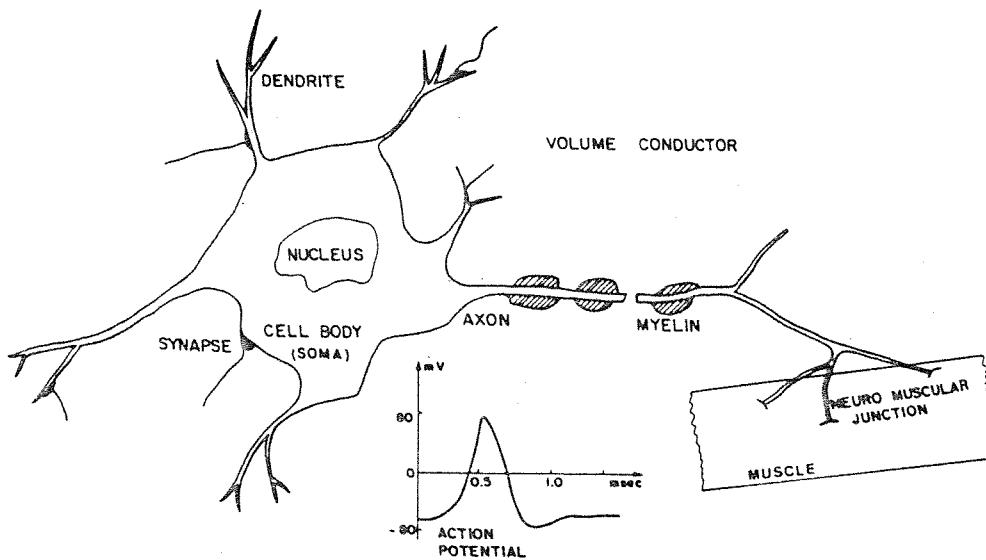


FIGURE 1. The nerve cell.

functioning of a cell. There is a great variance in the size of nerve cells. The diameter can be as small as a few microns or as large as a few tens of microns. It is surrounded by an excitable membrane, the thickness of which is in the range of 50 to 150 Å. The cell membrane is extended in various places to generate root-like structures called *dendrites*. These extensions are used for interconnections with other nerve cells.

The *axon* serves as the output of the nerve unit. It is an extension of the cell with a length that can be about 50 µm (in the cerebral cortex) or up to several meters (in peripheral nerves of large mammals). The diameter of the axon can range from less than 0.5 µm to about 1 mm (in the squid giant nerve fibers). Some axons are covered with interrupted *myelin sheath* which increases the velocity of information transfer.

Information into the neuron from other neurons is introduced through a junction called *synapse*. Synapses are located on the dendrites or on the soma. The synapses can cause an increase or decrease of the voltage across the membrane. The cell function is based on the integrative (in time and space) effects of these potential changes.

The tips of the axon serve as inputs to other neurons through synapses, or to activate muscles through special synapses — the neuromuscular junctions. Peripheral nerves are bounded together into a *nerve trunk*. The electrical activities of the single nerve cell will be discussed in the following sections. The signals that are picked up from the nerve trunk are the result of the electric field generated by the various nerves in the trunk.

B. The Excitable Membrane

The cell membrane can be considered as a dividing medium between the extracellular and intracellular fluids. These two fluids have different ionic concentrations. The membrane has different permeability to the various ions in the solutions. As a result of ions transfer, by means of diffusion and other mechanisms, a voltage is generated across the membrane.

If we consider the effects of only the three main ions, potassium $[K^+]$, sodium $[Na^+]$, and chloride $[Cl^-]$, we get the membrane potential, E , from the *Nernst Equation*:

$$E = \frac{RT}{F} \ln \left\{ \frac{P_K[K^+]_0 + P_{Na}[Na^+]_0 + P_{Cl}[Cl^-]_0}{P_K[K^+]_i + P_{Na}[Na^+]_i + P_{Cl}[Cl^-]_0} \right\} \quad (2.1)$$

where R , T , and F are the universal gas constant, the absolute temperature, and the Faraday

constant, respectively. P_x is the permeability of the resting membrane to the ion X, and $[X]_o$ and $[X]_i$ are the concentrations of the ion X in the extracellular and intracellular fluids.

The calculated resting cross-membrane potential is approximately 80 mV (the inside of the cell being negative with respect to its outside). This value agrees well with neurophysiological measurements.

Some membranes have *excitability* characteristics. When the membrane is excited by means of electrical, mechanical, or chemical stimulus, the permeabilities of the membrane to ionic transfer undergo some changes. These changes cause the resting potential of the membrane to increase, become positive for a short period of time, and later, when the membrane repolarizes, to return to its normal resting potential. The time course of the potential change, the *action potential*, is depicted in Figure 1.

Nerve and muscle cells have excitable membranes. The shape and time durations of the action potentials differ in the various cells. Muscle action potentials are usually much larger in duration.

The excitation of the membrane is caused only if the stimulus exceeds a threshold level (about 20 mV). Once the threshold has been crossed and action potential elicited, the threshold changes. Following the initiation of the action potential is a certain period (of the order of 1 to 2 μ sec) when the threshold level becomes infinite. This period is called the *total refractory period* in which no new action potential can be initiated. The threshold then returns to its resting value according to some decaying function. The period in which the threshold decays to its resting level is called *relative refractory period*. In this period, a new action potential can be elicited provided the stimulus is strong enough to cross the relatively high threshold.

C. Action Potential Initiation and Propagation

The decision as to whether a neuron elicits an action potential is a function of "outside" inputs it receives. In sensors, these may be the amount of photons absorbed or the value of the pressure applied. In regular nerve cells, the "inputs" come from nerve endings of other neurons. Each synapse can cause a slight increase in the resting membrane potential of the neuron (excitatory synapse) or can lower the resting potential (inhibitory synapse). The membrane potential is determined by the integration over all the synaptic effects, both in space and in time. When this cumulative effect causes the threshold to be crossed, an action potential is elicited.

The special properties of the membrane ensure the propagation of the action potential from the cell body along the axon toward its endings. When an action potential has been elicited, internal electric fields are generated sufficiently strong to excite neighboring portions of the membrane. The action potential thus propagates in one direction. It cannot propagate back, since the threshold of that portion of the membrane that has recently been activated is still very high.

The action potential, as propagates along the axon, is locally regenerated. Therefore, it propagates without attenuation. The information carried by the neuron is not in the shape of the action potential but in the interspike intervals. The neuron can be considered a stimulus-to-frequency converter. In most signal processing applications, we are not monitoring single action potentials (spikes) but rather the field generated by a trunk of fibers. We find then that both amplitude and frequency contents of the signal relate to the neural activity.

D. The Synapse

The axon of a neuron terminates with junctions to other neurons or to muscles. One axon can be connected by means of such junctions to many neurons or muscle fibers.

The synapse in the junction between one of the axon endings of one neuron and the dendrite or soma of another. The *presynaptic* region is the axon's ending. It does not actually touch the dendrite (or soma). A spacing of about 200 Å, known as the *synaptic cleft*, exists.

The region in the dendrite (or soma) on the other side of the cleft is called *postsynaptic region*.

When an action potential arrives at the presynaptic region, it causes the membrane characteristics to change. This change increases the ability of certain chemical substances (transmitters) to diffuse from the presynaptic region into the cleft. The transmitters that cross the cleft are captured by receptors in the postsynaptic region and cause membrane potential change. The change may be excitatory (*excitatory postsynaptic potential*, EPSP) or inhibitory (*inhibitory postsynaptic potential*, IPSP) depending on the type of transmitter released.

The complete process of transmitter release, cleft crossing, and postsynaptic receiving is relatively slow and is of the order of 0.5 msec. The transmission of information through the nervous system, though fast when compared with other biological mechanisms (hormones), may be considered slow when compared with electronic or optical systems.

III. THE MUSCLE

A. Muscle Structure

The skeletal muscle consists of cells with excitable membrane. The membrane is similar in principle to the neuron's membrane. Its function, though, is not to transfer or process information but to generate tension. The muscle is constructed from many separate fibers. The fibers contain two kinds of protein filaments, *actin* and *myosin*. These are arranged in parallel interlacing layers which can slide one into the other causing shortening of the muscle length. The sliding of the fibers is caused by chemical reactions that are not yet fully understood.

The generation of motion or force by the muscle is activated when the fiber membrane is excited. An action potential then propagates along the surface membrane of the fiber, triggering chemical reactions that, in turn, cause fiber contraction.

When a muscle contracts, the action potentials generate an electric field that can be monitored by means of surface (skin) electrodes. This field is a result of the contribution of many fibers at different times and with different rates. The signal (EMG) monitored this way will thus be a random signal with statistical properties that depend on the muscle function.

B. Muscle Contraction

The neuron that activates the muscle is called *motor nerve*. The axon endings of the motor nerve are similar to synapses but rather than activate another neuron, they are connected to muscle fibers. The motor neuron-muscle connection is called *neuromuscular junction* or *end plate*.

The chemical substance that serves as a transmitter in the end plate is acetylcholine (ACh).⁵ It is released from the axon endings when an action potential has arrived, diffuses toward the muscle membrane and is absorbed there at the receptors sites, causing muscle membrane potential change. When the change is sufficiently high and threshold level is crossed, an action potential is generated and propagates along the muscle membrane.

The process of transmitter release, diffusion, and reception at the muscle lasts about 0.5 to 1.0 msec. Additional delay in contraction is due to the dynamic properties of the muscle itself.

IV. VOLUME CONDUCTORS

The source of the bioelectric signals are the action potentials generated by single neurons and muscle fibers. The current densities generated by the membrane activity cause current changes in the surrounding medium. The surrounding tissues, in which induced current changes occur, are called the *volume conductor*.

In most clinical applications and in many neurophysiological applications, we monitor the fields of the volume conductor and not the bioelectric source itself. This is certainly the case when skin surface electrodes are used to monitor heart or brain activities. Even in neurophysiological studies, where electrodes are inserted into the tissue, we monitor the volume conductor effects.

It will be of extreme importance to be able to exactly deduct the underlying bioelectric source from gross measurements of the volume conductor. This is, however, a complex task, especially when the characteristics of the complex biological medium are considered.

Mathematical models of the current flow field in volume conductors have been developed with various degrees of success. The reader is referred to Plonsey's work³ for detailed presentation.

REFERENCES

1. Katz, B., *Nerve, Muscle and Synapse*, McGraw-Hill, New York, 1966.
2. Geddes, L. A., *Electrodes and the Measurement of Bioelectric Events*, Wiley-Interscience, New York, 1972.
3. Plonsey, R., *Bioelectric Phenomena*, McGraw-Hill, New York, 1969.
4. Hodgkin, A. L. and Huxley, A. F., A quantitative description of membrane current and its application to conduction and excitation in nerve, *J. Physiol.*, 117, 50, 1952.
5. Mountcastle, V. B., *Medical Physiology*, 13th ed., C.V. Mosby, St. Louis, 1974.

Charlie K

Chapter 3

RANDOM PROCESSES

I. INTRODUCTION

Randomness appears in biomedical signals in two major ways: the source itself maybe stochastic (as are indeed all information conveying signals) or the measurement system introduces external, additive or multiplicative, noise to the signal. Whether a signal is considered stochastic or deterministic is a matter of definition. An ECG signal can be considered deterministic, and even "almost" periodic, when some characteristic of the QRS are of interest, or it can be considered stochastic, where R-R interval variations are of interest.

Probability theory plays an important underlying role in the analysis of random signals. Therefore, we provide a brief review of probability theory in the opening of this chapter. The concepts of probability theory are then extended to the characterization and analysis of random signals. The emphasis in this chapter is on definitions and basic presentation of material directly required for the understanding of the topics discussed in later chapters. For a more detailed and rigorous presentation of the material, the reader is referred to the many textbooks available.¹⁻³

Special attention is given, in this chapter, to the topic of correlation analysis, since it has importance as a detection method often used in biomedical signal processing. The multi-dimensional gaussian process is introduced at the end of the chapter. In several analysis methods discussed in the course of this book, the assumption is made that the signal is gaussian, and reference is made to its distribution and other characteristics.

II. ELEMENTS OF PROBABILITY THEORY

A. Introduction

Consider an experiment, the outcome of which can be one of several events. The outcome of the experiment depends upon the combination of many factors which are unpredictable. The events are called discrete random events. We can not predict the exact result of such an experiment; we can, however, comment about the average outcome of a large number of experiments. A throw of a die serves as a popular example for such "experiments" where the events are the numbers on the face of the thrown die.

Assume we have performed the experiment N times. Out of the N resulted events, the event A_i has occurred n_i times. We define the relative frequency, f_i , as:

$$f_i = \frac{n_i}{N} \quad (3.1)$$

The probability of event A_i , $P(A_i)$, is then given as the limit of the relative frequency:

$$P(A_i) = \lim_{N \rightarrow \infty} (f_i) = \lim_{N \rightarrow \infty} \left(\frac{n_i}{N} \right) \quad (3.2A)$$

with

$$0 \leq P(A_i) \leq 1 \quad (3.2B)$$

Note that we have assumed that the limit in Equation 3.2A does exist.

Two events are called mutually exclusive events if the occurrence of one makes the appearance of the second impossible. If A_i and A_j are mutually exclusive, then the probability that A_i or A_j will occur is $P(A_i \text{ or } A_j)$, with

$$P(A_i \text{ or } A_j) = P(A_i) + P(A_j) \quad (3.3A)$$

and more generally, if the random variables A_i , $i = 1, 2, \dots, M$ are mutually exclusive, then

$$P(A_1 \text{ or } A_2 \text{ or } \dots \text{ or } A_M) = \sum_{i=1}^M P(A_i) \quad (3.3B)$$

For any event A we have an event $B = \text{NOT } A$ which are mutually exclusive, hence:

$$P(A) + P(\text{NOT } A) = P(\text{certain event}) = 1 \quad (3.4)$$

B. Joint Probabilities

When an experiment has many (rather than single) outcomes, we speak about joint probabilities. Consider, for example, the result of a blood test. The test outcome consists of several parameters. We can talk about the probability that the outcome of the blood test will be some given values for all the parameters; the probability of this happening is the joint probability. We denote the joint probability of the random variables A, B, C, \dots, J by $P(ABCD, \dots, J)$ with the meaning: the probability that A and B and C and, ..., and J will occur.

Often the probability of one event is influenced by another event. We may want to consider the probability of one event occurring, given that the other one has already occurred. This is known as conditional probability. The probability of event A occurring, given event B has occurred, is written as:

$$P(A|B) = \text{Probability (A occurs given B has occurred)} \quad (3.5)$$

As an example, consider the following experiment: two cards are successively drawn from a deck (without returning the drawn card to the deck) and the probability of the first being an ace and the second a king is sought. The problem can be posed as follows: what is the probability of drawing a king given an ace was previously drawn?

Consider now the relationship between the joint and conditional probabilities. Assume an experiment the result of which is given by two simultaneous events performed N times. Let n_A denote the number of times event A appeared in the outcome and n_{AB} the number of times the event A and B appeared. The probability of the joint event AB is

$$P(AB) = \lim_{N \rightarrow \infty} \left(\frac{n_{AB}}{N} \right) = \lim_{N \rightarrow \infty} \cdot \left(\frac{n_A}{N} \cdot \frac{n_{AB}}{n_A} \right) \quad (3.6A)$$

Assuming that the number of experiments is sufficiently large such that n_A is also very large, then we can rewrite Equation 3.6A as:

$$P(AB) = P(A) P(B|A) \quad (3.6B)$$

Therefore, we shall get the conditional probability

$$P(B|A) = \frac{P(AB)}{P(A)} \text{ (for } P(A) \neq 0) \quad (3.6C)$$

and

$$P(A|B) = \frac{P(AB)}{P(B)} \text{ (for } P(B) \neq 0\text{)} \quad (3.6D)$$

also:

$$P(A|B) = \frac{P(A) P(B|A)}{P(B)} \quad (3.6E)$$

and

$$P(B|A) = \frac{P(B) P(A|B)}{P(A)} \quad (3.6F)$$

The last equations are known as the Bayes' rule.

C. Statistically Independent Events

Two events are said to be statistically independent if their conditional probability equals the elementary probability of occurrence. Consider the conditional probability $P(B|A)$; if

$$P(B|A) = P(B) \quad (3.7)$$

then the information about the occurrence of event A has added nothing to the knowledge about event B. The two events are said to be statistically independent. Introducing Equation 3.7 into Equation 3.6C, we get:

$$P(AB) = P(A) \cdot P(B) \quad (3.8A)$$

namely, the joint probability of statistically independent events equals the multiplication of the individual probabilities. In general, for the case with n statistically independent events A_i , $i = 1, 2, \dots, n$, we have:

$$P(A_1, A_2, \dots, A_n) = \prod_{i=1}^n P(A_i) \quad (3.8B)$$

In addition, for all combinations $i, i < j < k, \dots, \leq n$

$$\begin{aligned} P(A_i) &= P(A_i) P(A_j) \\ P(A_i A_j A_k) &= P(A_i) P(A_j) P(A_k) \quad (3.8C) \\ \dots &\dots \\ P(A_1, A_2, \dots, A_n) &= P(A_1) P(A_2), \dots, P(A_n) \end{aligned}$$

D. Random Variables

The outcome of an experiment can be a number or some other description of the event. Let us assign a real number, or a set of numbers to each possible outcome of the experiment. We may have some n discrete values for the description of the outcomes, in which case the experiment will be described by a discrete set of numbers (vectors); we denote the set as a discrete random variable. In other cases, continuous values are required to describe the outcomes; this set is termed continuous random variable.

Consider the discrete random variable x . The values x_1, x_2, \dots, x_n are n discrete values constituting the random variable x . The probability of the event to which the number x_i was

assigned is $P_x(x = x_i)$, which denotes the probability that the random variable x will have the value x_i . If, for example, the n outcomes of the experiments are mutually exclusive, then:

$$\sum_{i=1}^n P_x(x = x_i) = 1 \quad (3.9)$$

since Equation 3.9 describes the probability of the certain event. When the outcome of the experiment consists of two events, x and y , we define as before the joint probability $P_{xy}(x = x_i, y = y_j)$, the probability that the random variable x will get the value x_i , and the random variable y get the value y_j . The certain event is the summation over all i and j , hence:

$$\sum_i \sum_j P_{xy}(x = x_i, y = y_j) = 1 \quad (3.10)$$

The conditional probabilities and the Bayes' rule are similar to Equation 3.6:

$$\sum_i P_x(x = x_i | y = y_j) = 1 \quad (3.10A)$$

$$\sum_j P_y(y = y_j | x = x_i) = 1 \quad (3.10B)$$

$$P_x(x = x_i | y = y_j) = \frac{P_{xy}(x = x_i, y = y_j)}{P_y(y = y_j)} \quad (3.10C)$$

$$P_y(y = y_j | x = x_i) = \frac{P_{xy}(x = x_i, y = y_j)}{P_x(x = x_i)} \quad (3.10D)$$

$$P_x(x = x_i | y = y_j) = \frac{P_y(y = y_j | x = x_i) P_x(x = x_i)}{P_y(y = y_j)} \quad (3.10E)$$

E. Probability Distribution Functions

Consider now the continuous case where the real random variable, x , can get any value on the real axis $-\infty \leq x \leq \infty$. We define the probability distribution function (sometimes called the cumulative distribution function) $P(x \leq X)$, namely, the probability that the random variable x will get a value less or equal the constant X . Clearly,

$$P(x \geq \infty) = P(x \leq -\infty) = 0 \quad (3.11)$$

$$P(x \leq \infty) = 1 \quad (3.12)$$

Also, the probability that the random variable x will get a value in the range $X_2 < x \leq X_1$ is nonnegative and is given by:

$$P(x \leq X_1) - P(x \leq X_2) = P(X_2 < x \leq X_1) \quad (3.13)$$

We conclude that for every $X_1 > X_2$ we get $P(x \leq X_1) \geq P(x \leq X_2)$, hence the probability distribution function is a nonnegative, nondecreasing function bounded by zero and 1 (see Figure 1).

Consider the case where two random variables, x and y , are participating in the range $(-\infty \leq x \leq \infty, -\infty \leq y \leq \infty)$. Define the joint probability distribution function, $P(x \leq X, y \leq Y)$. The joint probability distribution function of X and Y is the probability that the random variable x will get a value $x \leq X$ and the variable y will get a value $y \leq Y$. The following relations are obvious:

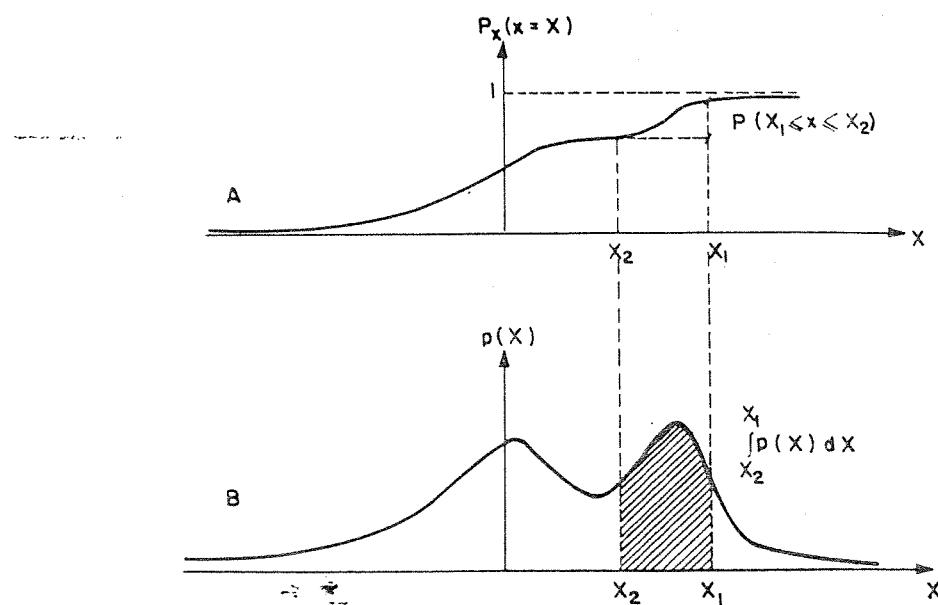


FIGURE 1. Probability functions. (A) Probability distribution function; (B) probability density function (PDF).

$$P(x \leq X, y \leq \infty) = P(x \leq X) \quad (3.14A)$$

$$P(x \leq \infty, y \leq Y) = P(y \leq Y) \quad (3.14B)$$

$$P(x \leq -\infty, y \leq Y) = 0 = P(x \leq X, y \leq -\infty) \quad (3.14C)$$

$$P(x \leq \infty, y \leq \infty) = 1 \quad (3.14D)$$

F. Probability Density Functions

Consider the continuous random variable x and the probability $P(X - \Delta X \leq x \leq X)$, namely, the probability that x will get a value close to X ; denote this probability by $p(X) \cdot \Delta X$. We shall consider an infinity small neighborhood ΔX , and use Equation 3.13 to get:

$$p(X) = \lim_{\Delta X \rightarrow 0} \frac{P(x \leq X) - P(x \leq X - \Delta X)}{\Delta X} \quad (3.15)$$

which, by definition, is the derivative operator, hence:

$$p(X) = \frac{d}{dx} P(x \leq X) \quad (3.16A)$$

$$P(x \leq X) = \int_{-\infty}^X p(x) dx \quad (3.16B)$$

In Equation 3.16A, we have assumed that the derivative exists, or can be expressed in terms of delta functions.² The following relations are obvious:

$$\int_{-\infty}^{\infty} p(x) dx = 1 \quad (3.17A)$$

$$P(X_2 < x \leq X_1) = \int_{X_2}^{X_1} p(x)dx \quad (3.17B)$$

Note also that for continuous random variable, the probability of getting a certain value, say X_0 , is zero:

$$P(x = X_0) = 0 \quad (3.17C)$$

When the experiment consists of more than one random variable, we shall use the joint probability density function defined on the joint distribution function in a similar manner to Equation 3.16:

$$p(X, Y) = \frac{\partial^2}{\partial X \partial Y} P(x \leq X, y \leq Y) \quad (3.18A)$$

$$P(x \leq X, y \leq Y) = \int_{-\infty}^Y \int_{-\infty}^X p(x, y) dx dy \quad (3.18B)$$

In Equation 3.18A, we have assumed that the partial derivatives exist, or can be expressed in terms of delta functions.²

We shall be interested in situations where a random variable is to be investigated given some conditions on the other random variable. Consider the probability of the random variable y being less or equal to some Y , given that x is in the range $X - \Delta X < x \leq X$:

$$P(y \leq Y|x = X) = \frac{\int_{-\infty}^Y p(X, y) dy}{p(X)} \quad (3.19)$$

We define Equation 3.19 as the conditional probability distribution function. If this function has derivatives, we define the *conditional probability density function*, $p(Y|X)$, by:

$$p(Y|X) = \frac{dP(y \leq Y|x = X)}{dY} \quad (3.20A)$$

$$P(y \leq Y|x = X) = \int_{-\infty}^Y p(y|X) dy \quad (3.20B)$$

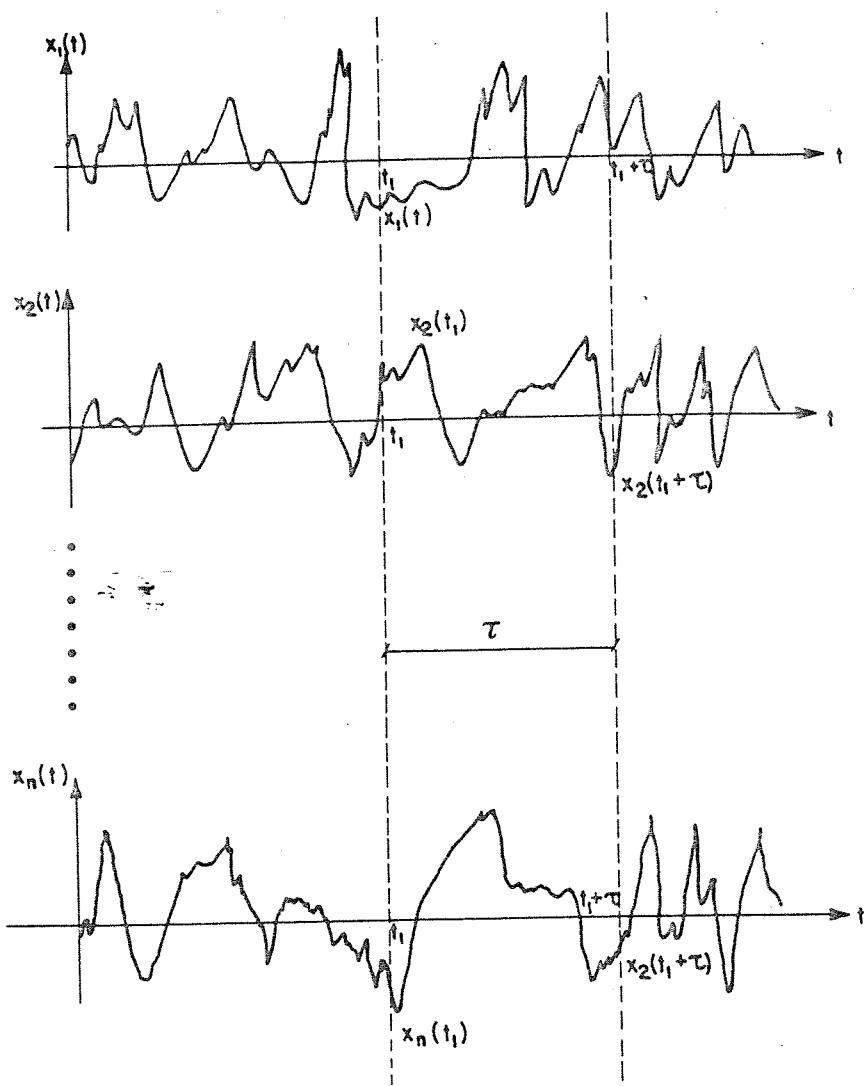
The following relationships are easily shown from the last definitions and previous relations:

$$p(Y|X) \geq 0 \quad (3.21A)$$

$$P(Y_1 < y \leq Y_2|X) = \int_{Y_1}^{Y_2} p(y|X) dy \quad (3.21B)$$

$$\int_{-\infty}^{\infty} p(y|X) dy = 1 \quad (3.21C)$$

$$p(Y|X) = \frac{p(X, Y)}{p(X)} \quad (3.21D)$$

FIGURE 2. The ensemble of the random process $x(t)$.

III. RANDOM SIGNALS CHARACTERIZATION

A. Random Processes

Up until now we have considered random variables that were "the outcome of an experiment". We now consider a time function $x(t)$. The value of the function at any time t_1 , $x(t_1)$, is a random variable. The variable t is chosen since most of the signals that will be considered here are time-dependent signals. In general, of course, x may be a function of distance or any other variable.

We assume that we have a source generating the random function $x(t)$, which is denoted as a sample function. The source generates many sample functions which together are known as the ensemble. At any time, t_1 , we can observe the values of all sample functions, to get many "outcomes of the experiment". Figure 2 depicts n sample functions out of the ensemble of random process x .

For example, consider the EEG signal (see Appendix A) taken by means of surface electrodes located at a certain location on the scalp. We want to investigate the properties of the EEG, recorded at the particular location, of a given population segment, say, healthy

children in a certain age group. We hypothesize that the EEG recorded is a sample function of a common random process. After recording n sample functions, we have for each time t_i , n values of the random variable $x(t_i)$. We can use these values to estimate the probability distribution function of $x(t_i)$.

Assume we have an ensemble of n sample functions of $x(t)$. For large n we can use the ensemble for estimating the probability, $P(x(t) \leq X)$. We can also estimate the joint probability, $P(x(t_1) \leq X_1, x(t_2) \leq X_2, \dots, x(t_N) \leq X_N)$; we shall denote this as $P(x_1, x_2, \dots, x_N)$ and the joint probability density function of the random process $x(t)$ by $p(x_1, x_2, \dots, x_N)$. It is easy to show that the probability density function has the following properties:

$$p(x_1, x_2, \dots, x_N) \geq 0 \quad (3.22A)$$

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} p(x_1, x_2, \dots, x_N) dx_1, \dots, dx_N = 1 \quad (3.22B)$$

$$p(x_{j+1}, x_{j+2}, \dots, x_N | x_1, x_2, \dots, x_j) = \frac{p(x_1, x_2, \dots, x_N)}{p(x_1, x_2, \dots, x_j)} \quad (3.22C)$$

Suppose we have two random processes $x(t)$ and $y(t)$. These are said to be *statistically independent random processes* if:

$$\begin{aligned} p(x_1, x_2, \dots, x_N, y_1, y_2, \dots, y_M) &= \\ &= p(x_1, x_2, \dots, x_N)p(y_1, y_2, \dots, y_M) \end{aligned} \quad (3.23)$$

The joint probability density function of $x(t)$ was estimated at the time $t = t_1$ (refer to Figure 2). Let us also estimate the joint probability density function at another time, $t = t_1 + \tau$.

A process in which the joint probability density functions are identical at all times, and for all N is called a *stationary process*.

B. Statistical Averages (Expectations)

Consider two random variables, x and y , the joint probability function of which is given. Let a new random variable, $z = f(x, y)$, be a single-valued function of these two variables. We define the *expectation*, $E\{z\}$, by:

$$E\{z\} = E\{f(x, y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) p(x, y) dx dy \quad (3.24)$$

The expectation is also known as the statistical average or mean. If x and y are discrete such that x gets the values x_i , $i = 1, 2, \dots, I$ and y gets y_j , $j = 1, 2, \dots, J$, then the expectation is

$$E\{z\} = E\{f(x, y)\} = \sum_{i=1}^I \sum_{j=1}^J f(x_i, y_j) P(x_i, y_j) \quad (3.25)$$

It is easy to show that the expectation is a linear operator.

Consider for example the function $z = x^n$. The expectation of x^n is known as the n th moment of x :

$$E\{x^n\} = \int_{-\infty}^{\infty} x^n p(x) dx \quad (3.26)$$

The first moment of x , $m = E\{x\}$, is called the *mean*. In a stationary random process it is the "dc" component. The n th *central moment*, μ_n , is defined by:

$$\mu_n = E\{(x - m)^n\} = \int_{-\infty}^{\infty} (x - m)^n p(x) dx \quad (3.27)$$

The second central moment has a special importance; it is called the *variance* and is denoted by σ_x^2 :

$$\sigma_x^2 = \mu_2 = E\{(x - m)^2\} = \int_{-\infty}^{\infty} (x - m)^2 p(x) dx \quad (3.28)$$

The square root of the variance is called the *standard deviation*. The $(n + k)$ th order *joint moment*, $E\{x^n y^m\}$, and *joint central moments*, μ_{nm} , are similarly defined:

$$E\{x^n y^m\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^n y^m p(x, y) dx dy \quad (3.29)$$

$$\begin{aligned} \mu_{nm} &\equiv E\{(x - m_x)^n (y - m_y)^m\} = \\ &\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - m_x)^n (y - m_y)^m p(x, y) dx dy \end{aligned} \quad (3.30)$$

Of a special importance is the joint central moment, μ_{11} , called the *covariance*:

$$\mu_{11} = E\{(x - m_x)(y - m_y)\} \quad (3.31)$$

IV. CORRELATION ANALYSIS

A. The Correlation Coefficient

Consider two random variables x and y . Let us assume that the two have some linear dependence. By this we mean that if we were to plot the samples of the random variables on the x - y plane, the scatter diagram of the points thus plotted will be distributed around a straight line:

$$y_p = ax + b \quad (3.32)$$

The random points (x, y) will not fall exactly on the line (points (x, y_p)) but will be scattered in its vicinity. The mean square error, ϵ , between the random points and the line is given by:

$$\epsilon = E\{(y - y_p)^2\} = E\{(y - ax - b)^2\} \quad (3.33)$$

The best line that is drawn through the scattered points in such a way that ϵ is minimized is known as the *regression-line*. The parameters a and b of the regression line are given by minimizing Equation 3.33:

$$\begin{aligned} \frac{\partial \epsilon}{\partial a} &= 0 = -2E\{(y - ax - b)x\} = -2E\{xy\} + 2aE\{x^2\} + 2bE\{x\} \\ \frac{\partial \epsilon}{\partial b} &= 0 = -2E\{(y - ax - b)\} = -2E\{y\} + 2aE\{x\} + 2b \end{aligned} \quad (3.34)$$

By solving, we get:

$$a = \frac{\mu_{11}}{\sigma_x^2} \quad (3.35A)$$

$$b = m_y - \frac{\mu_{11}}{\sigma_x^2} m_x \quad (3.35B)$$

Hence, the regression line is

$$y_p = \frac{\mu_{11}}{\sigma_x^2} (x - m_x) + m_y \quad (3.36)$$

Note that the line passes through the point (m_x, m_y) . It is often useful to normalize the random variable such that we work with variables that have zero mean and unit variance; hence, define:

$$\xi = \frac{x - m_x}{\sigma_x}; \phi = \frac{y - m_y}{\sigma_y}; \phi_p = \frac{y_p - m_y}{\sigma_y} \quad (3.37)$$

Then the regression line (Equation 3.36) can be written in terms of the new variables:

$$\phi_p = \frac{\mu_{11}}{\sigma_x \sigma_y} \xi = \rho \xi \quad (3.38)$$

where the *correlation coefficient*, ρ , is defined:

$$\rho = E\{\xi \phi\} = \frac{\mu_{11}}{\sigma_x \sigma_y} \quad (3.39)$$

The correlation coefficient is also known as the *normalized covariance*, and is the slope of the regression line of the normalized variables. Consider now the mean square difference, $E\{(\xi - \phi)^2\}$. Since it is a quantity squared it must be nonnegative:

$$E\{(\xi - \phi)^2\} = E\{\xi^2\} - 2E\{\xi \phi\} + E\{\phi^2\} = 2(1 - \rho) \geq 0 \quad (3.40A)$$

Similarly,

$$E\{(\xi + \phi)^2\} = E\{\xi^2\} + 2E\{\xi \phi\} + E\{\phi^2\} = 2(1 + \rho) \geq 0 \quad (3.40B)$$

From Equation 3.40 we get the range of the correlation coefficient:

$$-1 \leq \rho \leq 1 \quad (3.41)$$

The variables x and y are said to be *uncorrelated* or *linearly independent* if their correlation coefficient, ρ , is zero. Note that from Equation 3.39 we get for uncorrelated variables:

$$\mu_{11} = E\{\xi \phi\} = 0 = E\{(x - m_x)(y - m_y)\} = E\{xy\} - E\{x\}E\{y\} \quad (3.42A)$$

Hence,

$$E\{xy\} = E\{x\}E\{y\} \quad (3.42B)$$

Recall (Equation 3.8A) that Equation 3.42B is true for statistically independent variables. Statistically independent random variables, therefore, are also uncorrelated.

Conversely

B. The Correlation Function

Consider the two real random processes $x(t)$ and $y(t)$. Let us denote $x(t_1) = x_1$ and $y(t_2) = y_2$. The joint expectation, $E\{x_1 y_2\}$, is, in general, a function of t_1 and t_2 and is thus called the *cross correlation function*. In the case where $y(t) = x(t)$, the joint expectation $E\{x_1 x_2\}$ is called the *auto correlation function*. The correlation function, $r_{xy}(t_1, t_2)$, is given by:

$$r_{xy}(t_1, t_2) = E\{x(t_1), y(t_2)\} \quad (3.43)$$

We define the *normalized correlation function*, $\rho_{xy}(t_1, t_2)$, similar to the way the correlation coefficient was defined (Equation 3.39), hence:

$$\rho_{xy}(t_1, t_2) = \frac{E\{(x_1 - m_{x1})(y_2 - m_{y2})\}}{\sigma(x_1)\sigma(y_2)} = \frac{r_{xy}(t_1, t_2) - m_{x1}m_{y2}}{\sigma(x_1)\sigma(y_2)} \quad (3.44)$$

where $m_{x1} = E\{x(t_1)\}$, $m_{y2} = E\{y(t_2)\}$, and σ is the standard deviation.

The correlations of two random processes are described by the autocorrelations of each one and the cross correlation between them. These can be written in a matrix form, denoted as the *correlation matrix*:

$$R_{xy}(t_1, t_2) = \begin{bmatrix} r_x(t_1, t_2) & r_{xy}(t_1, t_2) \\ r_{yx}(t_1, t_2) & r_y(t_1, t_2) \end{bmatrix} \quad (3.45)$$

Similarly, we can consider the joint covariances of the two processes:

$$\mu_{xy} = E\{(x(t_1) - m_{x1})(y(t_2) - m_{y2})\} \quad (3.46)$$

and the *covariance matrix*, W_{xy} ,

$$W_{xy} = \begin{bmatrix} \mu_x(t_1, t_2) & \mu_{xy}(t_1, t_2) \\ \mu_{yx}(t_1, t_2) & \mu_y(t_1, t_2) \end{bmatrix} \quad (3.47)$$

If the random processes are stationary, the expectations are not a function of the actual times t_1 and t_2 , but only on the time difference $\tau = t_1 - t_2$. Hence, for stationary processes we have:

$$r_{xy}(t_1, t_2) = r_{xy}(t, t - \tau) = r_{xy}(\tau) = E\{x(t)y(t - \tau)\} \quad (3.48A)$$

and the normalized correlation:

$$\rho_{xy}(\tau) = \frac{r_{xy}(\tau) - m_x m_y}{\sigma_x \sigma_y} \quad (3.48B)$$

Also, for stationary processes we have:

$$r_{xy}(t_1, t_2) = r_{xy}(\tau) = r_{yx}(-\tau) \quad (3.49A)$$

$$r_x(\tau) = r_x(-\tau) \quad (3.49B)$$

$$|r_{xy}(\tau)| \leq (r_x(0))^{1/2} (r_y(0))^{1/2} \quad (3.49C)$$

$$|r_x(\tau)| \leq r_x(0) \quad (3.49D)$$

Random processes, which are not stationary in the (strict) sense that their joint probability density functions are time independent, but for which Equation 3.48 holds, are called *stationary in the wide sense*.

The meaning of the autocorrelation function can be demonstrated by Figure 2. We take the value of the random process $x_1(t_1)$ from the first sample function and multiply it by the value of the sample function τ seconds later. The product is calculated for each one of the sample functions. These products are averaged over the complete ensemble to yield the autocorrelation at time t_1 . If the process is stationary (wide sense stationarity is sufficient), then this procedure will yield the same result for every t_1 .

C. Ergodicity

We define the time average as the average over the sample function (rather than over the ensemble). Consider the function $z(t) = f(x(t))$; then the time average of the random sample function $z(t)$ is defined by:

$$\langle z(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\tau}^{\tau} z(t) dt = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\tau}^{\tau} f(x(t)) dt \quad (3.50)$$

Hence, the first time average of $x(t)$ is simple:

$$\langle x(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\tau}^{\tau} x(t) dt \quad (3.51)$$

The time average of the two random processes, $x(t)$ and $y(t)$,

$$\langle r_{xy} \rangle(t, \tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-\tau}^{\tau} x(t + \tau) y(t) dt \quad (3.52)$$

is called the *time cross correlation*. For stationary processes, Equation 3.52 is a function of τ only.

The *ergodicity theorem* states that, under certain conditions, the time averages of a stationary random process equal its statistical expectations, with probability one. The proof and detailed description of the ergodicity conditions are beyond the scope of this book.

The ergodicity theorem is of extreme importance. Statistical analysis requires the estimations of various expectations. For nonergodic processes, this means the need to average over the ensemble. In practical cases, however, it is often difficult to get a sufficient number of sample functions to perform such averaging. For ergodic processes, the expectation can be estimated from the time averages performed over a single sample function. Consider, for example, the statistical analysis of EEG. The expectations of this process can be estimated if we have at hand many sample functions recorded under similar conditions from identical subjects. These are practically impossible to achieve. If, however, the process is ergodic, we can use a single EEG sample function recorded from one subject and estimate the expectations by performing time averages over it.

V. THE GAUSSIAN PROCESS

A. The Central Limit Theorem

In the practical case, the probability density functions of the processes under investigation are not given. These have to be estimated from available sample functions. A very important

distribution is the *gaussian*, or *normal distribution*. A random variable x is said to have a gaussian (normal) distribution if:

$$p(x) = \frac{1}{(2\pi)^{1/2}\sigma_x} \exp\left(-\frac{(x - m_x)^2}{2\sigma_x^2}\right) \quad (3.53)$$

with the corresponding probability distribution function:

$$P(x \leq X) = \frac{1}{(2\pi)^{1/2}\sigma_x} \int_{-\infty}^X \exp\left(-\frac{(x - m_x)^2}{2\sigma_x^2}\right) dx \quad (3.54)$$

Note that the gaussian distribution is completely described by the first and second moments. The normal distribution is often denoted by $N(m_x, \sigma_x)$.

We often normalize the random variable, x , to get a new variable, ξ (Equation 3.37). The probability density function of the new variable will be $N(0, 1)$.

It can be shown that (for statistically independent samples) the probability distribution function of the finite time mean of the sample tends to become gaussian (as the number of samples is increased without limit), regardless of the probability distribution of the random variable. This is known as the *central limit theorem*. This theorem is one of the reasons the gaussian distribution is commonly encountered in real signal processing. Since the sum of independent random variables tends to be gaussian distributed, we would expect, for example, the EEG to be almost gaussian distributed. This is because the EEG can be, in some cases, considered the sum of fields generated by the activity of many neurons operating independently of one another.

B. Multivariate Gaussian Process

Consider the n random variables x_1, x_2, \dots, x_n . Define the n dimensional random vector, \underline{x} such that:

$$\underline{x}^T = [x_1, x_2, \dots, x_n] \quad (3.55)$$

The expectations of the random variables can be expressed in a vector form by:

$$\underline{m}_x^T = E\{\underline{x}^T\} = [E\{x_1\}, E\{x_2\}, \dots, E\{x_n\}] \quad (3.56)$$

The covariance matrix W_x (see Equation 3.47) is given by:

$$W_x = E\{(\underline{x} - \underline{m}_x)(\underline{x} - \underline{m}_x)^T\} \quad (3.57)$$

The random variables are said to be *jointly gaussian distributed* if their joint probability density function is

$$p(x_1, x_2, \dots, x_n) = p(\underline{x}) = \frac{1}{(2\pi)^{n/2}|W|^{1/2}} \exp\left[-\frac{1}{2}(\underline{x} - \underline{m}_x)^T W^{-1} (\underline{x} - \underline{m}_x)\right] \quad (3.58)$$

where $|W|$ is the determinant of the covariance matrix.

If the random variables x_i are uncorrelated, the covariance matrix is diagonal. If all x_i also have zero mean, then the joint probability density function becomes:

$$p(\underline{x}) = \frac{1}{(2\pi)^{n/2}|\Omega|^{1/2}} \exp\left[-\frac{1}{2} \underline{x}^T \Omega^{-1} \underline{x}\right] \quad (3.59)$$

where Ω denotes the diagonal covariance matrix.

A gaussian random process is a random process in which for every time instant, t_i , the random variables $x_i(t_i)$, $i = 1, 2, \dots, n$ are jointly gaussian distributed.

Since the gaussian process is completely described by the first and second moments, it follows that a wide sense stationary gaussian process is also stationary in the strict sense.

It can also be shown that a linear transformation of a random gaussian process yields gaussian random process. This property of the gaussian process is important in signal processing since any linear amplification and filtering will not alter the gaussian nature of the process.

REFERENCES

1. Papoulis, A., Probability, *Random Variables and Stochastic Processes*, McGraw-Hill, New York, 1965.
2. Davenport, W. B. and Root, W. L., *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York, 1958.
3. Bendat, J. S. and Piersol, A. G., *Random Data: Analysis and Measurement Procedures*, Wiley-Interscience, New York, 1971.

Chapter 4

DIGITAL SIGNAL PROCESSING

I. INTRODUCTION

Most of the signals of interest in biomedicine are continuous (analog); namely, they are defined over a continuous range of the variable (usually time). It is important, however, to analyze discrete signals, namely, signals that are defined only at discrete instants. Modern digital technology, both in terms of hardware and software, make discrete time processing advantageous over analog processing. The advantages are such that usually it is worthwhile to convert the analog signal into a discrete one so that discrete processing can be applied. The conversion is done by analog to digital (A/D) conversion systems that sample and quantize the signal at discrete times. Usually the sampling is performed uniformly; however, sometimes nonuniform sampling is used.

Some of the main tasks of digital signal processing are to apply filtering, to estimate various signal parameters, and to perform transformations of the signal (i.e., Fourier transform). When the results of the processing are not required immediately following the signal, *off-line* processing methods are used. When results are required during and immediately after the signal sample has been acquired, *real-time* or *on-line* processing methods are used.

Depending on the application, the processing time and the size of memory required are of importance in digital signal processing. Off-line processing can be performed on general purpose computers. Real-time processing usually requires special dedicated machines or systems such as fast correlators, dedicated Fourier transform machines, array processors, or hardware multiply-accumulate circuits.

This chapter briefly discusses the problem of sampling and quantization. The z transform is introduced and applied to digital signal processing.

The material presented in this chapter is used as a background for later chapters. For a comprehensive discussion of this material and other important topics such as the fast Fourier transform (FFT) and digital filtering, the reader is referred to the literature on digital signal processing.¹⁻⁶

II. SAMPLING

A. Introduction

Consider a band-limited continuous signal, $x(t)$, which is bounded in amplitude by $|x(t)| \leq A$. A band-limited signal has a Fourier transform, $X(w)$, with $X(w) = 0$ for all $|w| \geq w_{\max}$. It is desired to process the signal by digital processing means, in general, by a digital computer. The digital machine requires the conversion of the continuous signal, $x(t)$, into a series of discrete numbers $\{x(k)\}$ (refer to Figure 1). The first stage of the conversion is sampling or time discretization. Assume that we apply uniform sampling with interval T_s , namely, with sampling frequency $f_s = 1/T_s$. We then generate out of the signal a sequence of sampled data $\{x(nT_s)\}$, $n = 0, 1, \dots$. Note that each sample, $x(iT_s)$, is continuous in amplitude in the range $-A \leq x(iT_s) \leq A$. A discretization of amplitude is now required to get the necessary sequence of discrete numbers. A quantizer device is used for this method. The combined process of sampling and quantization is known as *analog-to-digital conversion* (A/D). We consider the quantization as a transformation of the amplitude continuous quantity, $x(iT_s)$, into the discrete number, $x_q(iT_s)$. A detailed discussion on the quantization process is given in the next section.

In the general processing system, the signal is required in an analog continuous form after

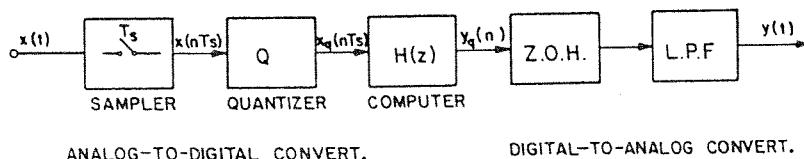


FIGURE 1. Digital processing of analog signals.

processing. Consider (see Figure 1) the processed sequence of numbers, $y_q(n)$. To get an analog signal, we perform the *digital-to-analog conversion* (D/A). The D/A most often consists of *zero order hold* (ZOH) circuit, followed by a *low pass filter* (LPF).

B. Uniform Sampling

The ideal sampler, described in Figure 1, consists of a switch that is closed for an infinitely short duration every T_s seconds. It can be considered as an ideal switch driven by a train of impulse function, $\delta_{T_s}(t)$, where:

$$\delta_{T_s}(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \quad (4.1)$$

and where $\delta(t)$ is the delta (or dirac) function. The sampled signal can be considered as the multiplication of the continuous signal, $x(t)$, with the train, $\delta_{T_s}(t)$. The time discrete sampled signal is written, thus, as a time function, denoted by $x^*(t)$:

$$x^*(t) = x(t)\delta_{T_s}(t) = \sum_{n=-\infty}^{\infty} x(t)\delta(t - nT_s) \quad (4.2)$$

It can be shown that if $x(t)$, the band-limited signal, has a Fourier transform given by $X(w)$, then the Fourier transform of the sampled signal, $X^*(w)$, is given by:

$$X^*(w) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} X(w + nw_s) \quad (4.3)$$

where $w_s = 2\pi f_s = 2\pi/T_s$. Figure 2 shows an example of the transform for two cases. In Figure 2b, the sampling frequency obeys $w_s > 2w_{max}$ where w_{max} is the largest frequency of the signal $x(t)$. We note that the sampled signal in the frequency domain consists of nonoverlapping functions. Consider the effect of a low pass filter that will pass all frequencies in the range $-w_{max} \leq w \leq w_{max}$ undistorted, while zeroing all frequencies outside this range. The Fourier transform of the signal at the output of the filter equals that of $x(t)$. Since the Fourier transform is unique, we can restore the original signal from its samples by such low pass filtering operation, provided the sampling frequency obeys:

$$w_s \geq 2w_{max} \quad (4.4)$$

This is known as the *sampling theorem*. Condition 4.4 is known as the *Nyquist rate*. Figure 2c shows the Fourier transform of the sampled signal when the sampling frequency is less than the Nyquist rate. In this case, the functions in the frequency domain overlap and low-pass filtering cannot restore the signal without distortions. The phenomenon of overlapping is called *aliasing in frequency*. Note that when sampling a continuous signal that is not band limited, aliasing always occurs no matter how large w_s is.

In practical cases, when a signal has a large w_{max} , it is often preprocessed by analog low

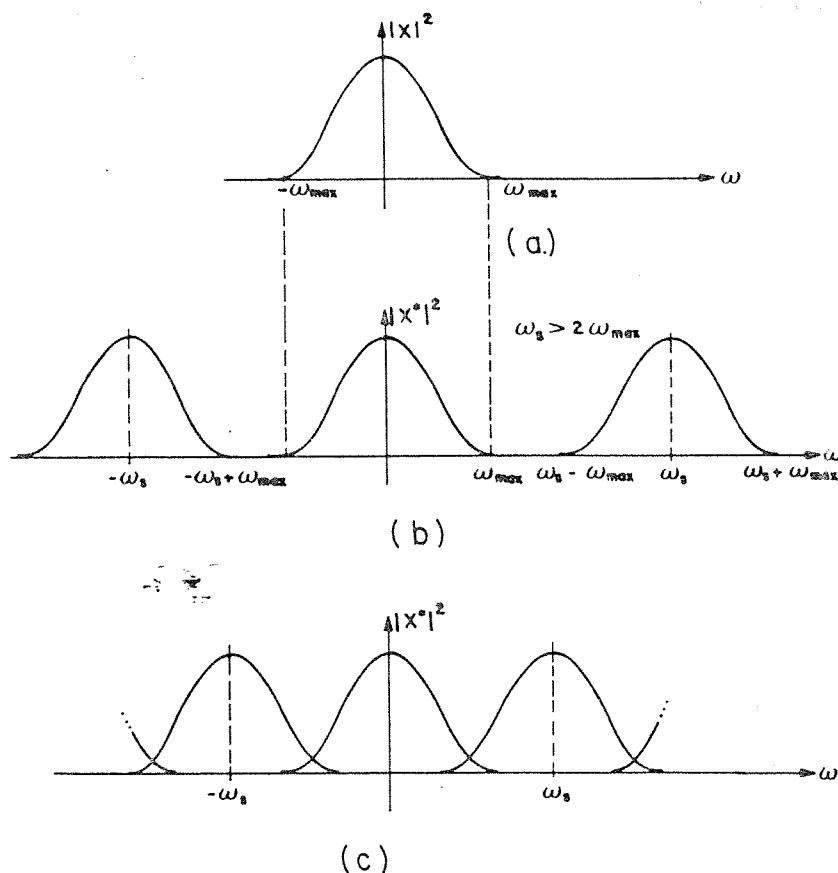


FIGURE 2. Sampled band-limited signal in the frequency domain. (a) Spectrum of the band-limited signal; (b) spectrum of the sampled signal, $w_s > 2w_{\max}$; (c) spectrum of the sampled signal, $w_s < 2w_{\max}$. Note aliasing.

pass filter in such a way that the high frequencies are eliminated so as not to cause aliasing problems. In theory, the signal can be sampled at the lowest Nyquist rate, $w_s = 2w_{\max}$.

The reconstruction of a signal such sampled requires an ideal rectangular low pass filter, which is impossible to implement. The need to use realizable filters for the reconstruction of the signal makes it necessary to sample at frequencies higher than the Nyquist rate. Sampling at frequencies of 2.5 to 10 times w_{\max} are often used.

C. Nonuniform Sampling

Uniform sampling rate is convenient since the information is contained in the value of the sample only. No time information is required since it is known *a priori* that samples are equally spaced by T_s seconds. Sometimes, however, the signal consists of some intermittent occurrences of fast changing and relatively quiescent intervals. One would then tend to sample at a high rate during fast changing periods, while reducing sampling rate during the quiescent intervals. This calls for an adaptive, nonuniform sampling. The ECG signal is an example where such sampling scheme may be effective.

Two main reasons exist for using nonuniform, adaptive sampling. The first is when effective storage is required. The problem is to store the signal using minimum storage size, retaining the ability to reconstruct the signal within a given error. The second is when an effective transmission is required. The problem here is to reduce the transmission rate (bit per second), retaining the ability to reconstruct the signal at the receiver side within a given error.

Several data compression techniques to reduce transmission rate and storage requirements have been developed for communications application. The differential pulse code modulation (DPCM) is one of the most popular schemes. An error signal is generated and nonuniformly quantized. The error is the difference between the original signal and a signal estimated from the output of the quantizer. Thus, only the error is quantized and transmitted reducing the amount of information. The output of the quantizer is uniformly sampled. An improvement to the accuracy of the above scheme is the introduction of an adaptive quantizer that automatically adapts the step size of the quantizer, q , according to the signal. Adaptive delta modulation (ADM) is such a scheme often used in synchronous communication systems.

A significant reduction in data compression can be achieved by using nonuniform sampling.⁷⁻⁸ Consider a scheme in which information is sent only when the source signal crosses a threshold level. This will cause periods in the signal, where fast changes exist, to be sampled at a higher rate than periods with slow variations. Note, however, that the transmission is now asynchronous, since the receiver does not know *a priori* the exact location of the sample on the time axis. In storage application, information must be added to indicate the time of the sample.

1. Zero, First, and Second Order Adaptive Sampling

Three basic schemes for adaptive, nonuniform sampling are discussed in this section. A comparative study of these three methods, with applications to the ECG, has been reported.⁹ The zero order adaptive method is also known as the *voltage triggered method*. Assume that at the time t_i , a sample was sent. The next time a sample will be sent is the time t_{i+1} , when the absolute value of the difference, $(x(t_{i+1}) - x(t_i))$, first exceeds a given threshold, R_o . Hence, the i th sampling interval is given by τ_i , such that:

$$|\Delta x(t_i, \tau_i)| = |x(t_i + \tau_i) - x(t_i)| > R_o \quad (4.5)$$

Since the signal is assumed to be band limited by w_{max} , there is no use to sample it at a rate higher than $w_s = kw_{max}$ (where k is a constant in the (empirical) range $2 \leq k \leq 10$). When Equation 4.5 yields sampling interval $\tau_i < 2\pi/kw_{max}$, we replace it by $\tau_s = 2\pi/kw_{max}$. Thus the maximum instantaneous sampling frequency of the adaptive scheme is bounded by kw_{max} . An example of the voltage-triggered nonuniform sampling of the ECG is given in Figure 3.

The first order method is known also as the *two points projection method*. Here, the first samples are used to estimate the slope of the signal. As long as subsequent samples fall within some specified error of this slope, they are ignored. The first sample that falls outside the error tolerance is stored (or transmitted) and used to estimate the next slope. Denote the derivative of the signal, at time t_i , by $\dot{x}(t_i)$. Assume that at t_i , the sample $x(t_i)$ has been stored. The next sample to be stored is the sample at time $t_i + \tau_i$, $x(t_i + \tau_i)$, for which the absolute value of the slope's difference first crosses the threshold, R_1

$$|\Delta \dot{x}(t_i, \tau_i)| = |\dot{x}(t_i + \tau_i) - \dot{x}(t_i)| > R_1 \quad (4.6)$$

Note that here we compare the slope of points at time $(t_i + \tau)$ with the slope at the last point to be stored. When R_1 is crossed, we store the sample $x(t_i + \tau_i)$ and use the new slope $\dot{x}(t_i + \tau_i)$ as a new reference.

The slope of the signal has to be estimated. Consider the uniform sampling of the signal at a maximum rate of $w_s = kw_{max}$, yielding the samples $\{x(nT_s)\}$, $n = 0, 1, \dots$. The slope can be estimated by:

$$\hat{\dot{x}}(nT_s) = \frac{x(nT_s) - x((n-1)T_s)}{T_s} \quad (4.7A)$$

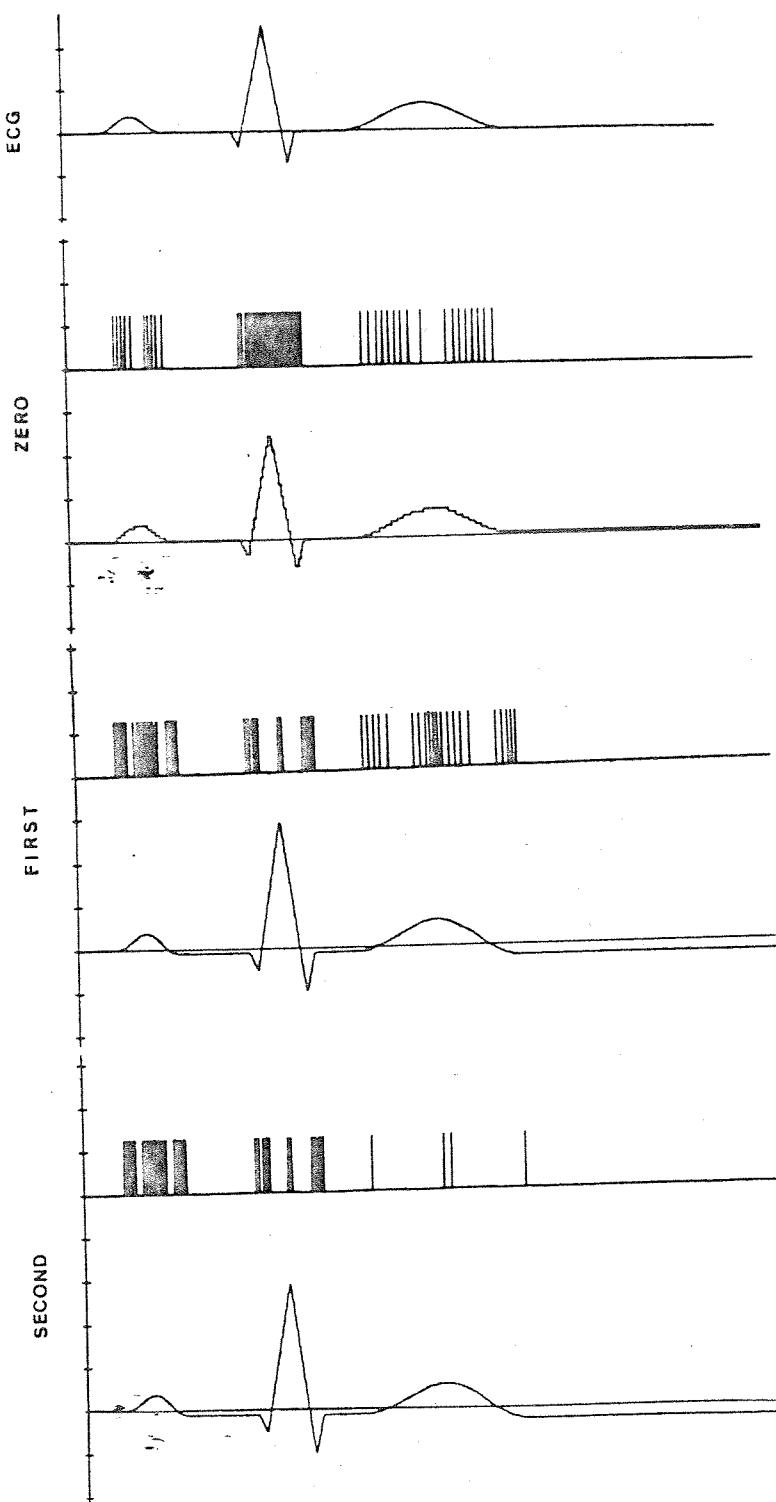


FIGURE 3. Nonuniform sampling of ECG. Synthesized ECG and sampling instances and reconstructed signal for zero, first, and second order adaptive sampling methods.

If the signal contains additive noise, the estimation 4.7A can be modified by:

$$\hat{x}(nT_s) = \frac{\frac{1}{M} \sum_{j=0}^{M-1} x((n+j)T_s) - \frac{1}{M} \sum_{j=0}^{M-1} x((n-j)T_s)}{(M-1)T_s} \quad (4.7B)$$

where $(2M-1)$ is the number of samples used to smooth the data. The slope is then estimated every $(2M-1)T_s$ seconds.

The application of the two points projection method to the ECG is demonstrated in Figure 3.

The second order nonuniform sampling method is known as the *second differences method*.¹⁰ It examines the slope just before the current sample and just after it. If the absolute value of these two adjacent slopes is larger than a given threshold, R_2 , the sample is stored. Hence, here we are considering the local change of slopes. The method is formulated as follows: The sample $x(t_i)$ is stored if:

$$|\dot{x}(t_i^+) - \dot{x}(t_i^-)| > R_2 \quad (4.8)$$

In practice, we have to estimate the time derivatives. This can be done, again, by uniformly sampling the signal at a maximum rate. The examination of the sample $X(nT_s)$ at time $t_n = nT_s$ by Equation 4.8 can be done by using the slope estimation of Equation 4.7B. If we choose again a window of $(2M-1)$ samples for smoothing the data, we get (assuming M is odd):

$$\dot{x}(t_n^+) = \frac{\frac{2}{M+1} \sum_{j=0}^{(M-1)/2} x\left(\left(n + \frac{M-1}{2} + j\right)T_s\right) - \frac{2}{M+1} \sum_{j=0}^{(M-1)/2} x\left(\left(n + \frac{M-1}{2} - j\right)T_s\right)}{\left(\frac{M-1}{2}\right)T_s} \quad (4.9A)$$

$$\dot{x}(t_n^-) = \frac{\frac{2}{M+1} \sum_{j=0}^{(M-1)/2} x\left(\left(n - \frac{M-1}{2} + j\right)T_s\right) - \frac{2}{M+1} \sum_{j=0}^{(M-1)/2} x\left(\left(n - \frac{M-1}{2} - j\right)T_s\right)}{\left(\frac{M-1}{2}\right)T_s} \quad (4.9B)$$

The application of the second difference method to ECG is shown in Figure 3. Blanchard and Bar⁹ showed that the voltage triggered and second differences method had serious defects in sampling their synthetic ECG. They concluded that the two points projection method best suited the signal yielding average compression reduction of about 1:14.

2. Nonuniform Sampling with Run Length Encoding

The asynchronous transmission of the nonuniformly sampled signal does not offer compatibility with digital storage. Information concerning the time interval between samples must be provided. Several methods have been suggested for such schemes; we shall present here a method suggested by Mark and Todd¹¹ in which the time intervals are coded by means of run lengths.

We employ a *nonuniform sampling encoder* (NSE), which samples the signal each time it crosses a threshold level. The NSE divides the dynamic range of the signal $\pm A$ into equally spaced threshold levels and transmits a sample each time one of the levels is crossed. The information encoded for each sample is the time interval from the last level crossing and the direction of crossing.

Let the time of the i th crossing be denoted by t_i , and the time interval between the i th

and $(i - 1)$ crossings, denoted by $\tau_i = t_i - t_{i-1}$. We shall quantize the time axis such that the time quantum be t_q . The quantized i th interval, τ_i^q , is given by:

$$\tau_i^q = \begin{cases} t_q & \text{if } \tau_i \leq t_q \\ m_i t_q & \text{if } t_q < \tau_i \end{cases} \quad (4.10)$$

$m_i = [\tau_i/t_q + 0.5]$, with $[\cdot]$ being the integer smaller than or equal to the argument. We shall encode each sample as follows: Denote each time quantum, t_q , in τ_i^q by a pair of zeroes (00), an upward crossing by (01), and a downward crossing by (10). A digital word is now generated by placing the pairs of zeroes (minus one pair) followed by the crossing code. For example, a downward crossing that occurred $3t_q$ sec after the previous crossing will be denoted by 00 00 10. We now code this word by letting the first digits be the number of pairs of zeroes followed by a single bit — one for downward and zero for upward crossings. For the example above, we shall get the coded number 101. This coding is the run length of the initial word.⁷ The length of the encoded word in the example is $n = 3$ bits. This allows for two digits for the registration of the interval. If no crossing occurs in the period of $n \cdot t_q$, a word of n times "00" is generated (for $n = 3$ this word is coded into 110 denoting 3 groups of "00"). Long quiescent periods with no level crossings will be coded into a succession of the words 110.

Consider the compression ratio provided by the method. Assume that the probability of downward crossing equals the probability of upward crossing. We then get:

$$P(01) = P(10) = \frac{1}{2} (1 - P(00)) \quad (4.11)$$

since $P(01) + P(10) + P(00) = 1$, the certain event.

If we further assume that the NSE output sequence is independently and identically distributed, then the mean run length,⁷ $\bar{\ell}$, is

$$\bar{\ell} = 2 \cdot \frac{1 - p^m(00)}{1 - P(00)} \text{ bits} \quad (4.12)$$

where $m = 2^{n-1} - 1$ is the number of time quanta that can be coded with block length of n bits. The rate of information is given by the required bits per seconds. If the rate of the coded information is denoted by r_c , then:

$$r_c = \frac{\bar{\ell}}{m \cdot t_q} \quad (4.13)$$

Consider now the case where the signal is sampled uniformly with sampling frequency, $f_s = 1/T_s$, and where each sample is described by a word length of N bits. The rate of information, r , is

$$r = \frac{N}{T_s} = Nf_s \quad (4.14)$$

The compression ratio achieved by the nonuniform sampling with run length coding is

$$\frac{r_c}{r} = \frac{\bar{\ell}}{m \cdot N} \frac{T_s}{t_q} = \frac{\bar{\ell}}{m \cdot N} \beta \quad (4.15A)$$

where we have denoted $f_q = 1/t_q = \beta f_s$.

Since the probability, $P(00)$, is much larger than that of $P(01) = p(10)$, it is interesting to check the compression ratio when $P(00)$ approaches one. The limiting value of the compression is

$$\lim_{P(00) \rightarrow 1} \left(\frac{r_c}{r} \right) = \frac{2}{N} \beta \quad (4.15B)$$

Mark and Todd⁷ have shown that for a band-limited gaussian signal with uniform spectrum, under certain assumptions, a limit on β exists:

$$\beta < \frac{\sigma_x}{d} \frac{8\pi}{(3)^{1/2}} \quad (4.16)$$

where σ_x is the standard deviation of the signal and d is the spacing between level thresholds. Average compression ratios of more than ten can be achieved under the above assumptions.

The nonuniform scheme described here is attractive because it provides significant compression with relatively simple implementation. The run-length coding can be easily implemented by means of digital counters. Note also that the method uses synchronous transmission for the coded blocks. The method can be further improved by employing predictive coding and nonuniform threshold levels.

III. QUANTIZATION

A. Introduction

The sampling process provides discretization of the time axis. Each sample, however, is continuous in amplitude in the range $|x(iT_s)| \leq A$, where A is the amplitude bound of the signal. Since digital circuits and computers can handle only finite length words, discretization of amplitudes is also required. Amplitude quantization is the process of replacing the analog values of the signal samples, with approximate values taken from a finite set of allowed values.¹¹ The quantizer is a nonlinear device that transfers a sequence of analog samples into a sequence of binary words compatible with digital systems, for storage transmission or processing. If the binary word is of n bits, the quantized sample can be one of the 2^n members of the quantizer's output set.

The most common quantizer is the *zero memory quantizer* in which each sample is decoded individually. A more complex quantizer is the *block quantizer* which "looks" at a block of input samples and provides a corresponding block of quantized symbols. *Sequential quantizers* store some information about an input sample and quantize only additional information provided by incoming samples. Delta modulation (DM), differential pulse code modulation (DPCM), and other similar schemes belong to this class of quantizers.

The discussion in this section will concentrate on zero memory quantizers. The quantization noise is discussed followed by a presentation of rough quantization with some applications.

B. Zero Memory Quantization

A zero memory quantizer of N points is defined by specifying a set of $N - 1$ levels, x_1, \dots, x_{N-1} , a set of N output points $x_{q1}, x_{q2}, \dots, x_{qN}$ and an operator $Q(x)$ that relates them.

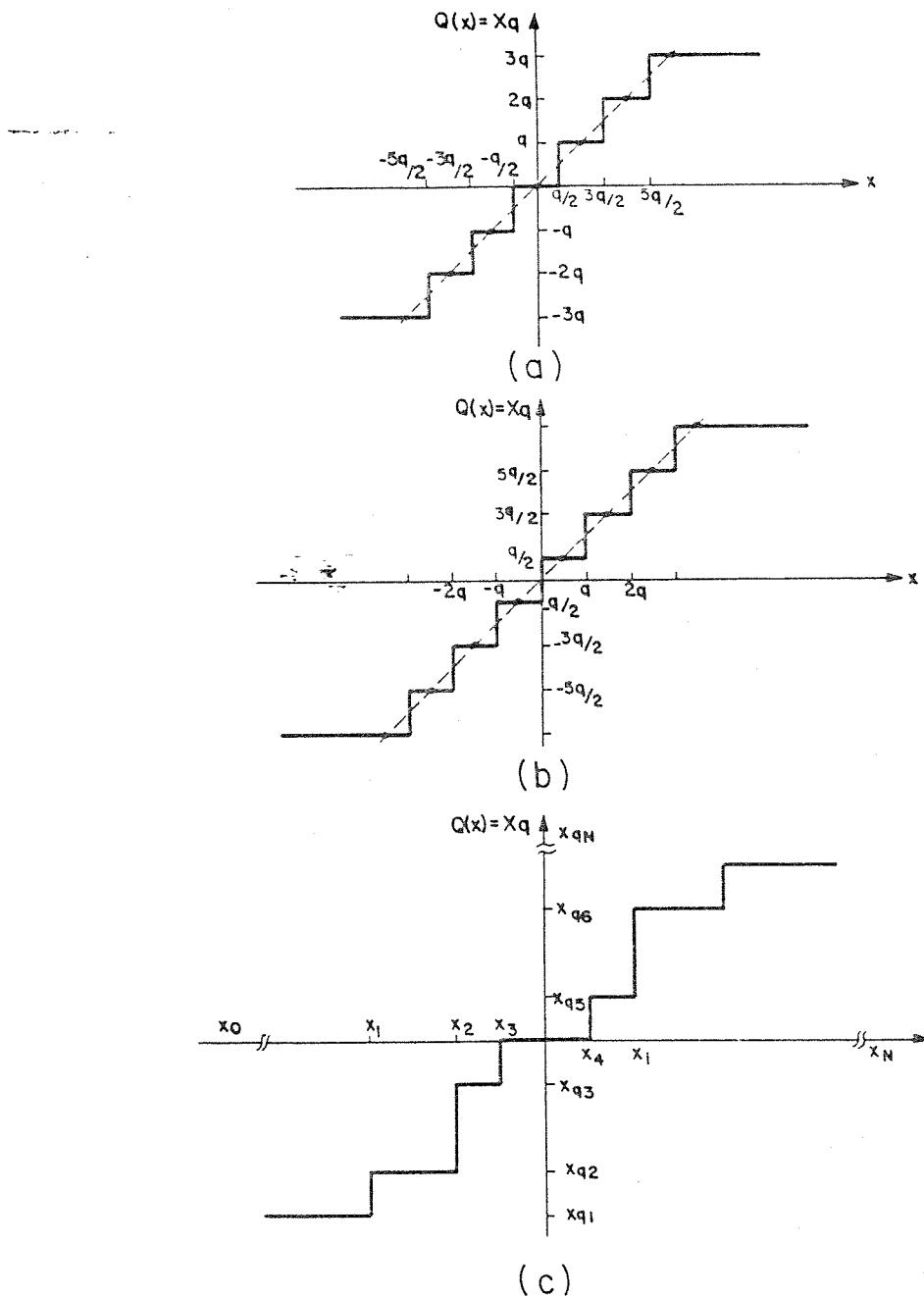


FIGURE 4. Zero memory quantizers. (a) Midtread uniform quantizer; (b) midriser uniform quantizer; (c) nonuniform quantizer.

Several types of zero memory quantizers are depicted in Figure 4. The first two quantizers (Figures 4a and b) are uniform quantizers, in which the quantization steps are of equal size, q . Figure 4a shows a midtread uniform quantizer with $N = 7$ where the quantization operation, $Q(x)$, is given by:

$$Q(x): \begin{cases} x_q = -3q; & x \leq -\frac{5}{2}q \\ x_q = mq; & \frac{2m-1}{2}q < x \leq \frac{2m+1}{2}q \\ x_q = 3q; & \frac{5}{2}q < x \end{cases} \quad (4.17A)$$

where m is an integer, $m = -2, -1, 0, 1, 2$.

Figure 4b shows a midviser uniform quantizer with $N = 6$ and with:

$$Q(x): \begin{cases} x_q = -\frac{7}{2}q; & x \leq -3q \\ x_q = \frac{2m-1}{2}q; & (m-1)q < x \leq mq \\ x_q = \frac{7}{2}q; & 3q < x \end{cases} \quad (4.17B)$$

where m is an integer, $m = -2, -1, 0, 1, 2, 3$.

Note that in two of the examples the average slope of the quantization steps is one. In general, the average slope can be made any positive quantity a , by changing the scale of x or x_q . Note also that the "average" line passes through the origin. In general, bias can be introduced to shift the average line from the origin.

Figure 4c depicts a nonuniform quantizer where the size of quantization steps is a function of x .

The quantizers described here can be considered as a "memoryless" nonlinearity whose characteristics are viewed as "staircase" approximations to the function $Q(x) = x_q \equiv ax$. We can consider the quantizer as the noisy operator:

$$Q(x) = x_q = ax + n_q \quad (4.18)$$

where n_q is the quantization additive noise which is dependent on the amplitude of the signal. The quantization noise resulted from the midtread quantizer of Figure 4a, as a function of x is shown in Figure 5. In the range $|x| < 7/2q$ the quantization noise is bounded by $|n_q| \leq q/2$. Outside this range, the noise increases linearly with x . In the case of a nonuniform quantizer, the noise will be bounded by the nonuniform quantization levels ($q_i/2$) and will behave like a saw waveform with variable amplitude.

Let us assume that the quantizer under test is unbounded, namely, $x_o = -\infty$, $x_N = \infty$ (see Figure 4c). This is a valid assumption since it can be considered true for every case where the amplitude of the signal $x(t)$ is bounded $|x(t)| \leq A$ and A is smaller than the quantizers' saturation levels. If $N = 2^n$, we can associate a unique n bit binary word with each output point. We then call the quantizer an *n-bit quantizer*.

Often the question arises of how to design the distribution of the quantization levels for a given n bit quantizer. Intuitively, it is clear that the uniform quantizer is not always the best and that the distribution of levels should be designed according to the probability distribution of the input signal, $x(t)$. Regions where the amplitude of the signal is highly probable should be quantized with finer levels than regions less probable. A uniform quantizer can still be used if we preprocess the signal by a memoryless nonlinear element (*compressor*). The compressor characteristics are determined in some optimal manner, depending on the input's probability distribution. After uniform quantization, the quantized signal is subjected

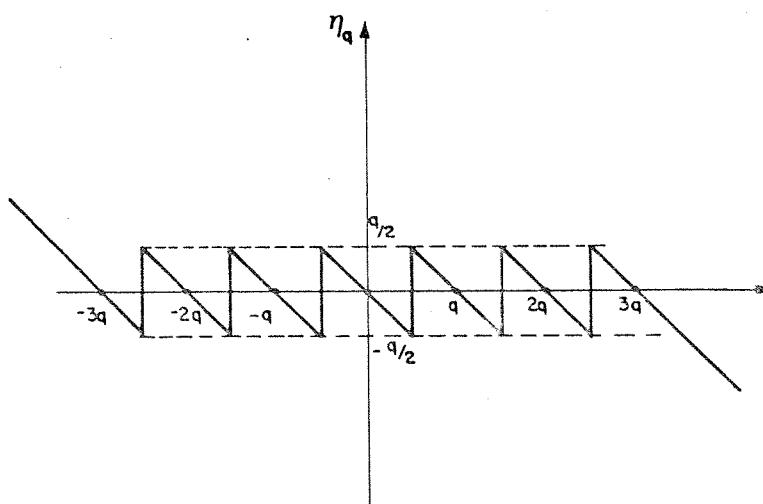
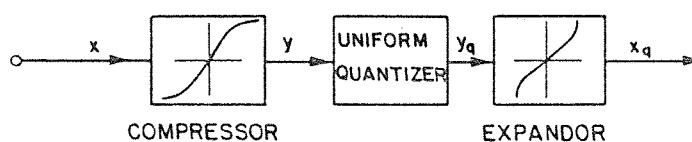
FIGURE 5. -Quantization noise for midread uniform quantizer with $N = 7$.

FIGURE 6. Block diagram of a companding system.

to an inverse memoryless nonlinear transformation (*expandor*) to obtain the quantized samples of the original input. The process is called *companding* and is often used in communication systems. Figure 6 depicts the block diagram of a companding system.

C. Analysis of Quantization Noise

Quantizers which have a large number of small quantization steps produce quantization noise which is approximately white. For the quantization noise to be approximately white, we also require that successive input signal samples will only be moderately correlated.

Consider the mean square error, σ_n^2 , of the quantizer for a given input signal, $x(t)$, with probability density function, $p(x)$:

$$\sigma_n^2 = \int_{-\infty}^{\infty} [Q(x) - x]^2 p(x) dx \quad (4.19A)$$

which can also be written as:

$$\sigma_n^2 = \sum_{i=1}^{N-1} \int_{x_{i-1}}^{x_i} (x_{qi} - x)^2 p(x) dx \quad (4.19B)$$

Note that Equation 4.19A and B describe the variance of the quantization noise.

Assume now that the quantization steps are small such that $p(x) \approx p(x_{qi})$ in the range $x_{i-1} \leq x \leq x_i$. Denote the step $q_i = x_i - x_{i-1}$; then, Equation 4.19 can be estimated by:

$$\sigma_n^2 \approx \frac{1}{12} \sum_{i=2}^{N-1} p(x_{pi}) q_i^3 + \int_{-\infty}^{x_1} (x_{qi} - x)^2 p(x) dx + \int_{x_{N-1}}^{\infty} (x_{qi} - x)^2 p(x) dx \quad (4.20A)$$

The last two terms on the right side of Equation 4.20A are due to the saturation parts of the quantizers. In Equation 4.20A it was assumed that $x_0 = -\infty$ and $x_N = \infty$. In general, we shall adjust the input signal such that $p(x) \equiv 0$ for $x > x_{N-1}$ and $x < x_1$; hence, the last two terms of Equation 4.20A are zero and:

$$\sigma_n^2 = \frac{1}{12} \sum_{i=2}^{N-1} p(x_{q_i}) q_i^3 \quad (4.20B)$$

The most common quantizer is the uniform quantizer for which $q_i = q$, $i = 2, \dots, N-1$. For these quantizers we can further simplify the expression for the noise variance:

$$\sigma_n^2 = \frac{q^2}{12} \sum_{i=2}^{N-1} p(x_{q_i}) q \approx \frac{q^2}{12} \quad (4.21)$$

since:

$$\sum p(x_{q_i}) q \approx \int p(x) dx = 1$$

The approximate value for the quantization noise variance given by Equation 4.21 is the one most often used. Uniform quantizers, with small quantization step q , and with saturation levels that can be ignored, have uniformly distributed noise,¹² in the range $-q/2 \leq n_q \leq q/2$ with variance $\sigma_n^2 = q^2/12$.

To justify the assumption that saturation effects can be neglected, we make sure the extreme quantization levels x_1 and x_{N-1} are some multiple of the input signal variance.

Define the *loading factor*, L_q (for symmetric uniform quantizers):

$$L_q = \frac{x_{N-1}}{\sigma_x} = \frac{-x_1}{\sigma_x} \quad (4.22)$$

where σ_x is the standard deviation (rms) of the input signal. A common choice for the loading factor is $L_q = 4$ (*four sigma loading*). For such a quantizer, we have $N = 2$ levels for $-4\sigma_x \leq x \leq 4\sigma_x$, hence:

$$q = 8\sigma/(N - 2) \quad (4.23)$$

A signal-to-noise ratio, SNR, for the quantizer can be defined by:

$$\text{SNR} = 10 \log (\sigma_x^2 / \sigma_n^2) \quad (4.24)$$

The SNR of the symmetric uniform quantizer is given by introducing Equations 4.21 and 4.23 into Equation 4.24:

$$\text{SNR} = 10 \log((N - 2)^2 \cdot 0.1875) \approx 20 \log N - 7.3 \approx 6n - 7.3 \quad (4.25)$$

In Equation 4.25, we have used the relation $N = 2^n$ and the assumption $N \gg 2$. In practical cases, we may use a 10-bit word, which results in a high SNR of more than 50 db.

D. Rough Quantization

In the previous section, the quantization noise was analyzed under the assumption that the number of quantization levels is high and quantization step is small. In many applications, these assumptions are not valid; we then talk about rough quantization. A more detailed

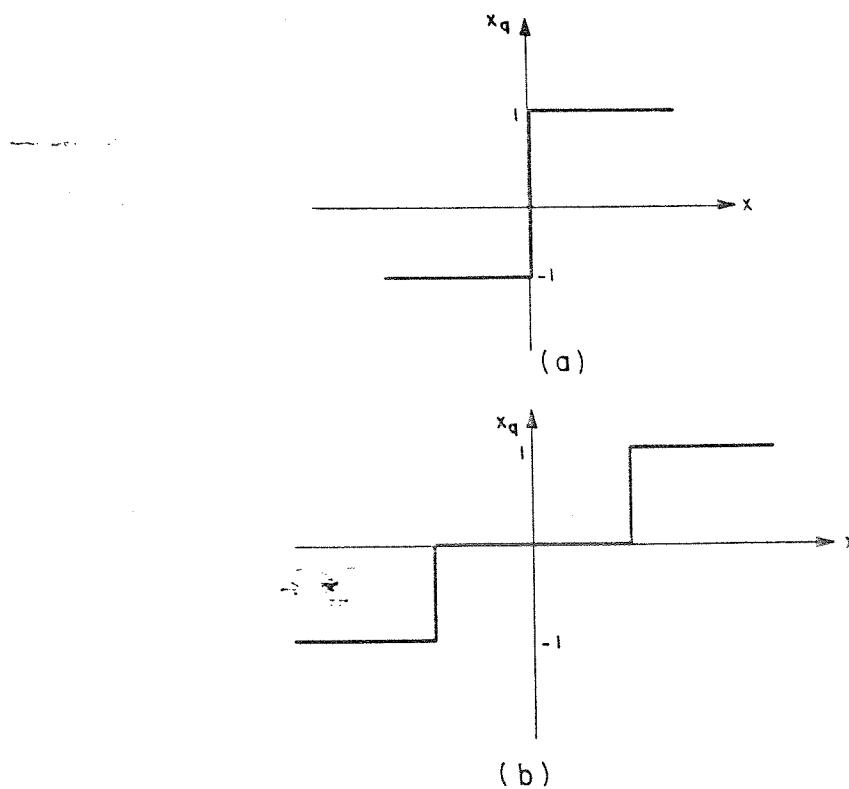


FIGURE 7. Rough quantizers (clippers). (a) One bit quantizer ($N = 2$). (b) Ternary quantizer ($N = 3$).

analysis is required for the noise generated by rough quantization. Figure 7 shows two rough quantizers. The first is a one bit quantizer ($N = 2$) in which the quantized sample is the sign of the input, $x_q = \text{Sgn}(x)$. The second rough quantizer has $N = 3$, where the description of the quantized output requires 2 bits. The importance of these two rough quantizers (also known as clippers) is due to the fact that processing of the quantized data is extremely simple. Digital correlation, for example, of signals quantized by these quantizers requires no multiplication. Such correlators have been suggested¹³ and applied¹⁴ to biomedical signal processing.

The statistical analysis of rough uniform quantization noise has been given by Widrow.¹⁵⁻¹⁶ Widrow has proven the *quantization theorem* which is, in some sense, analog to the Nyquist sampling theorem. The time samples of $x(t)$ have continuous amplitude probability density function $p(x)$. The quantized output, x_q , assumes only discrete amplitudes and thus has discrete probability density function $p_q(x_q)$. This function consists of a series of uniformly distributed impulses, each one centered in a quantization region. Figure 8 depicts the two density functions.

Widrow has considered the output density function as the sampled form of the input density function. If the input density, $p(x)$, is bounded by frequency (namely, its Fourier transform $P(u)$ has the property $P(u) = 0$ for all $|u| \geq u_{\max}$), then there exists a quantization level, q_s , such that quantized signals, with quantization levels $q \leq q_s$, contain all the information on the original distribution, $p(x)$. In other words, we can generate the original probability density function of the input signal from the quantized one, provided the quantizer obey the quantization law, $q \leq q_s$. The quantization law states that in order to have all information on the probability density function, the quantization step must obey:

$$q \leq q_s = \Pi/u_{\max} \quad (4.26)$$

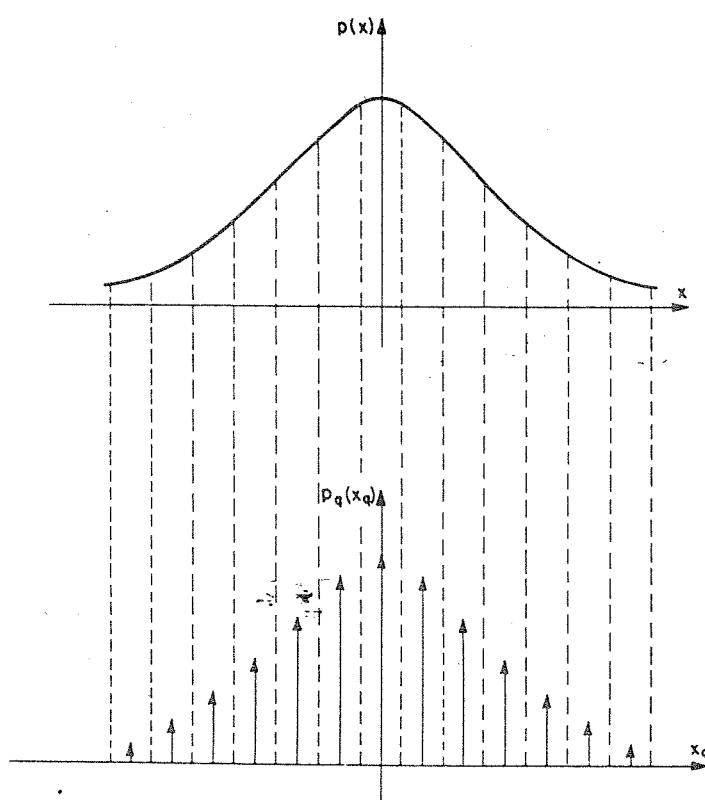


FIGURE 8. Probability density functions (PDF). Upper trace: PDF of the random variable, x ; lower trace: PDF of the quantized random variable, x_q .

Having developed the probability density function of the quantized signal, the noise statistics (such as the variance and the correlation) can be calculated in general. The variance of the quantized noise of Equation 4.21 is a special case of the general result given by Widrow.

IV. DISCRETE METHODS

A. The Z Transform

Consider the sampled signal, $x^*(t)$, given by Equation 4.2. If its (one-sided) Laplace transform is denoted by $X^*(S)$, then:¹⁷

$$X^*(S) = \sum_{n=0}^{\infty} x(nT) \exp(-nTS) \quad (4.27)$$

where T is the sampling interval and S is the complex frequency.

Define a new complex variable, Z :

$$Z = \exp(ST) \quad (4.28)$$

and the Z transform $X(Z)$:

$$Z\{x\} = X(z) = X^*(S)|_{\exp(ST)=z} = \sum_{n=0}^{\infty} x(nT)Z^{-n} \quad (4.29)$$

Equation 4.29 is known as the *one-sided Z transform* in which we assume that the signal $x(t) = 0$ for $t < 0$. It is easily shown that the Z transform is a linear operator. Several important properties of the transform make this operator an important tool for the solution of difference equation and the analysis of sampled data system.

One of the important properties is the shift property. It can easily be shown that:

$$Z\{x(t + mT)\} = Z^m X(Z) - Z^m x(0) - Z^{m-1} x(T) - \dots - Z x(mT - T) \quad (4.30)$$

For example, for $m = -1$, Equation 4.30 yields the Z transform of the sampled signal, $x(nT)$, delayed by one interval, in terms of the Z transform of the original signal:

$$Z\{x(t - T)\} = Z^{-1} X(Z) - Z^{-1} x(0) \quad (4.31)$$

where $x(0)$ is the initial conditions, the value of the signal at $t = 0$.

Note that Equation 4.29 describes the Z transform in terms of a series:

$$X(Z) = x(0) + x(T)Z^{-1} + x(2T)Z^{-2} + \dots + x(nT)Z^{-n} + \dots \quad (4.32)$$

The inverse transform, $x(nT)$, can be determined by inspection. The inverse transform can be determined also analytically by the residue theorem, through an integration in the complex plane.¹⁷

B. Difference Equations

A time invariant linear system, with input $u(t)$ and output $y(t)$ which are defined only at discrete instances $t = kT$, can be described by a difference equation:

$$\begin{aligned} y(kT) + a_1 y(kT - T) + a_2 y(kT - 2T) + \dots + a_p y(kT - pT) = \\ b_0 u(kT) + b_1 u(kT - T) + \dots + b_q u(kT - qT) \end{aligned} \quad (4.33)$$

The difference equation 4.33 can be solved by means of the Z transform in a similar manner in which differential equations are solved via the Laplace transform. Denote $Z\{y(t)\} = Y(Z)$, $Z\{u(t)\} = U(Z)$ and transfer both sides of Equation 4.33 into the Z domain using the shift properties. Assuming all intial conditions to be zero, we get:

$$\begin{aligned} (1 + a_1 Z^{-1} + a_2 Z^{-2} + \dots + a_p Z^{-p})Y(Z) = \\ (b_0 + b_1 Z^{-1} + b_2 Z^{-2} + \dots + b_q Z^{-q})U(Z) \end{aligned} \quad (4.34A)$$

or:

$$Y(Z) = H(Z)U(Z) = \frac{b_0 + b_1 Z^{-1} + \dots + b_q Z^{-q}}{1 + a_1 Z^{-1} + \dots + a_p Z^{-p}} U(Z) \quad (4.34B)$$

The output signal is given in the Z domain by the ratio of the two polynomials, $H(Z)$ (the Z domain transfer function) describing the system, and the input, $U(Z)$. Applying the inverse transform operation on Equation 4.34B will give the required output signal in the time domain.

The transfer function, $H(Z)$, can represent a digital filter operating on the signal, $u(t)$, to improve its quality in some sense. We shall also see in later chapters that we sometimes use $H(Z)$ as a means for effective description of the signal, $y(t)$. In these cases, we assume that $y(t)$ is the output of a linear system driven by $u(t)$, a white noise source. We identify $H(Z)$ and use the parameters a_i and b_i to represent the signal $y(t)$.

REFERENCES

1. Gold, B. and Rader, C. M., *Digital Processing of Signals*, McGraw Hill, New York, 1969.
2. Beauchamp, K. G. and Yuen, C. K., *Digital Methods for Signal Analysis*, George Allen and Unwin, Ltd., London, 1979.
3. Tretter, S. A., *Introduction to Discrete Time Signal Processing*, John Wiley & Sons, New York, 1976.
4. Chen, C.-T., *One Dimensional Digital Signal Processing*, Marcel Dekker, New York, 1979.
5. Oppenheim, A. V., Ed., *Application of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N. J., 1978.
6. Ahmed, N. and Rao, K. R., *Orthogonal Transforms for Digital Signal Processing*, Springer-Verlag, Berlin, 1975.
7. Mark, J. W. and Todd, T. D., A nonuniform sampling approach to data compression, *IEEE Trans. Commun.*, 29, 24, 1981.
8. Plotkin, E., Royzman, L., and Swamy, M. N. S., Nonuniform sampling of band limited modulated signals, *Signal Process.*, 4, 295, 1982.
9. Blanchard, S. M. and Barr, R. C., Zero, first and second order adaptive sampling from ECG's, in *Proc. of the 35th ACEMB*, Philadelphia, 1982, 209.
10. Pahlm, O., Borjesson, P. D., and Werner, O., Compact digital storage of ECG's, *Comput. Programs Biomed.*, 9, 293, 1979.
11. Gersho, A., Principles of quantization, *IEEE Trans. Circuits. Syst.*, 25, 427, 1978.
12. Sripad, A. B. and Synder, D. L., A necessary and sufficient condition for quantization errors to be uniform and white, *IEEE Trans. Acoust. Speech Signal Process.*, 25, 442, 1977.
13. Landsberg, D. and Cohen, A., Fast correlation estimation by a random reference correlator, *IEEE Trans. Instrum. Meas.*, 32, 438, 1983.
14. Cohen, A. and Landsberg, D., Adaptive real-time wavelet detection, *IEEE Trans. Biomed. Eng.*, 30, 332, 1983.
15. Widrow, B., A study of rough amplitude quantization by means of Nyquist sampling theory, *IRE Trans. Circuit Theory*, 3, 266, 1956.
16. Widrow, B., Statistical analysis of amplitude quantized sampled-data systems, *AIEE Trans.*, (Applications and Industry), II, 555, 1961.
17. Derusso, P. M., Roy, R. J., and Close, C. M., *State Variables for Engineering*, John Wiley & Sons, New York, 1965.

Chapter 5

FINITE TIME AVERAGING

I. INTRODUCTION

It is often necessary to estimate statistical properties of a random process given only a finite time sample function. This may be the case, for example, when the membrane potential is measured by means of electrodes and a high input impedance amplifier. The recorded signal can be considered a constant (or "almost" constant) voltage corresponding to the membrane signal, corrupted with additive, zero mean noise. The corrupting signal is the result of the electrode and amplifier noise as well as background noise. To recover the constant membrane voltage out of the noisy signal, its mean must be estimated from a finite time sample.

Other applications call for the estimation of the variance, the auto- or cross-correlation, and other statistical parameters.

The need to estimate statistical parameters, from a finite time sample, arises mainly from one or two of the following constraints:

Availability of long records — Often only short time records are available for processing. This may be due to the fact that the phenomenon monitored existed only for a short time or due to the fact that the acquisition system has allocated only a given time slot to the signal at hand.

Stationarity — Most often the signal to be processed is nonstationary. It is convenient, however, to assume stationarity so that powerful (stationary) signal processing techniques can be employed. The signal, therefore, is divided into segments, such that each can be considered stationary. Rather than estimating the statistics of a nonstationary signal, the problem now is to estimate the statistics of several "stationary" signals represented by finite time segments.

This chapter deals with the problems associated with finite time estimation. The errors involved with these types of estimators are discussed, as well as the improvements in signal-to-noise ratio achieved by the estimation.¹⁻⁴

An important case, when the signal to be processed is a repetitive one, is analyzed. In this case, synchronous averaging⁵⁻⁹ (known also as coherent averaging) techniques are employed in order to estimate the averaged waveshape of the repetitive signal.

EEG evoked potentials (EP) are classical examples of a signal treated by means of synchronous averaging.

Finite time averaging techniques are implemented by software on general purpose computers, on dedicated computers, and on special digital circuits.¹⁰⁻¹² In practice, all signal processing is time bounded, hence the importance of the knowledge of the estimation errors involved.

II. FINITE TIME ESTIMATION OF THE MEAN VALUE⁴

A. The Continuous Case

Consider a single sample record, $x(t)$, from a stationary ergodic random process, $\{x(t)\}$. The record is given for the time $t \in (0, T)$. Without any loss of generality one can assume that the process is given by

$$x(t) = \mu_x + n(t) \quad (5.1)$$

where μ_x is the expectation of $\{x(t)\}$

$$\mu_x = E\{x(t)\} \quad (5.2)$$

and $n(t)$ is a sample record of a stationary zero mean process.

Estimation of μ_x by means of the given finite time record is required. Define the estimate

$$\hat{\mu}_x = \frac{1}{T} \int_0^T x(t) dt \quad (5.3)$$

Clearly,

$$E\{\hat{\mu}_x\} = \mu_x \quad (5.4)$$

hence, the estimator 5.3 is an unbiased estimator.

The variance of the estimator yields the mean squared error, hence

$$\text{Var}[\hat{\mu}_x] = E\{(\hat{\mu}_x)^2\} - \mu_x^2 \quad (5.5)$$

The first term of the right side of Equation 5.5 can be rewritten, using Equation 5.3:

$$E\{\hat{\mu}_x^2\} = \frac{1}{T^2} \int_0^T \int_0^T E\{x(\xi)x(\eta)\} d\xi d\eta \quad (5.6)$$

by definition of the correlation function

$$E\{x(\xi)x(\eta)\} = r_x(\eta - \xi) = r_x(\tau) \quad (5.7)$$

where $\tau \triangleq \eta - \xi$.

Since stationarity was assumed $r_x(\tau)$ is independent of η and ξ , is an even function of τ , and has a maximum at $\tau = 0$.

Equation 5.6 in terms of the new variable τ can be rewritten as:

$$E\{\hat{\mu}_x^2\} = \frac{1}{T^2} \int_0^T \int_{-\xi}^{T-\xi} r_x(\tau) d\tau d\xi \quad (5.8)$$

Note that the integration of Equation 5.8 is carried out on the variables τ and ξ over the region shown in Figure 1.

Changing the orders of integration leads to:

$$E\{\hat{\mu}_x^2\} = \frac{1}{T^2} \left[\int_{-\tau}^0 \int_{-\tau}^T r_x(\tau) d\tau d\xi + \int_0^T (T - \tau) r_x(\tau) d\tau \right] = \\ \frac{1}{T^2} \int_{-\tau}^T (T - |\tau|) r_x(\tau) d\tau \quad (5.9)$$

since

$$\frac{1}{T^2} \int_{-\tau}^T (T - |\tau|) d\tau = 1 \quad (5.10)$$

We get from Equations 5.5, 5.9, and 5.10:

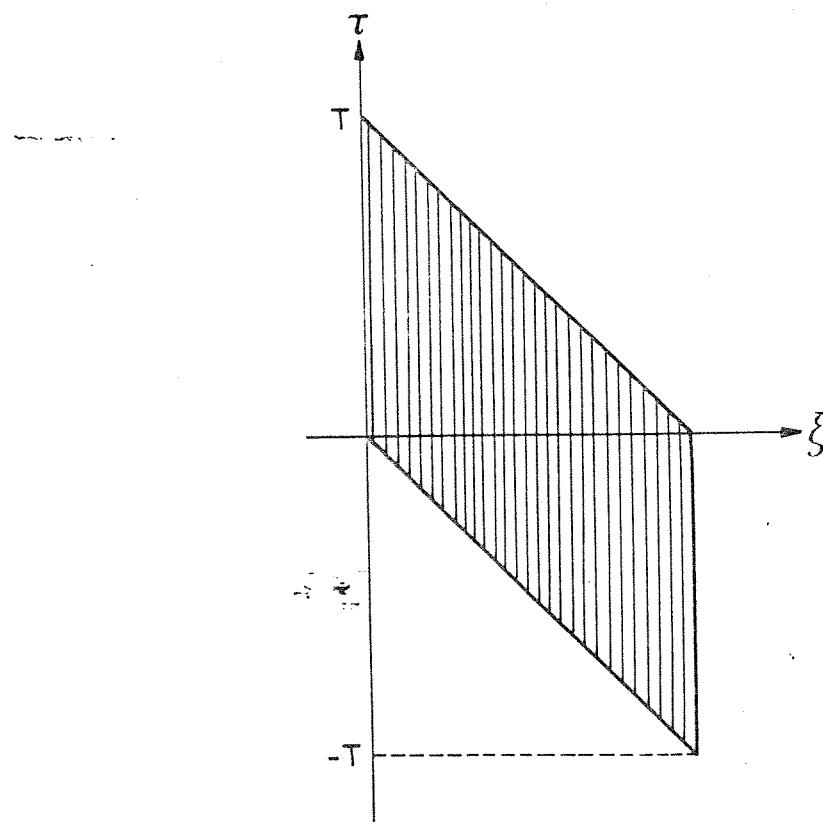


FIGURE 1. Integration boundaries.

$$\text{Var} [\hat{\mu}_x] = \frac{1}{T} \int_{-T}^T \left(1 - \frac{|\tau|}{T}\right) (r_x(\tau) - \mu_x^2) d\tau \quad (5.11)$$

In order to have the mean square error of the estimated quantity, $\hat{\mu}_x$, one has to have the autocorrelation function $r_x(\tau)$. This function is, however, usually unavailable. Note that when Equation 5.1 is used, one has

$$r_n(\tau) = r_x(\tau) - \mu_x^2 \quad (5.12)$$

where $r_n(\tau)$ is the autocorrelation function of the noise process; hence,

$$\text{Var} [\hat{\mu}_x] = \frac{1}{T} \int_{-T}^T \left(1 - \frac{|\tau|}{T}\right) r_n(\tau) d\tau \quad (5.13)$$

For gaussian processes and most physical random processes, Equation 5.3 is a consistent estimator. We shall deal first with two general cases.

1. Short Observation Time

In this case, the given record is very short so that the following approximation holds:

$$r_n(\tau) \approx r_n(0); \text{ for all } \tau \in (0, T) \quad (5.14)$$

Using the last approximation to solve Equation 5.13, we get

$$\lim_{T \rightarrow 0} \text{Var} [\hat{\mu}_x] = r_n(0) = r_x(0) - \mu_x^2 = \text{Var} [n(t)] \quad (5.15)$$

which states the trivial conclusion that for very short observation times, the variance of the estimated quantity equals that of the record itself, hence no improvement in signal-to-noise ratio is achieved.

2. Long Observation Time

Assume the noise autocorrelation function obeys

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^{T} |r_n(\tau)| d\tau = 0 \quad (5.16)$$

Then, for large observation times we get from Equation 5.13

$$\lim_{T \rightarrow \infty} \text{Var} [\hat{\mu}_x] \cong \frac{1}{T} \int_{-\infty}^{\infty} r_n(\tau) d\tau \quad (5.17)$$

The signal-to-noise ratio may be defined as the ratio between the expectation $E\{x(t)\} = \mu_x$ (the signal) and the variance, $\text{Var} [x(t)]$. The signal-to-noise ratio of the signal is thus

$$\text{SNR}_i = \frac{\mu_x}{\text{Var} [x(t)]} = \frac{\mu_x}{\text{Var} [n(t)]} \quad (5.18)$$

and that of the estimator

$$\text{SNR}_0 = \frac{\mu_x}{\text{Var} [\hat{\mu}_x]} \quad (5.19)$$

The ratio between the SNRs can be considered a figure of merit for the estimator.

Thus, for very large observation times, the expected improvement achieved by the estimator 5.3 is

$$\frac{\text{SNR}_0}{\text{SNR}_i} \cong \frac{\text{Var} [n(t)]}{\frac{1}{T} \int_{-T}^{T} \left(1 - \frac{|\tau|}{T}\right) r_n(\tau) d\tau} \quad (5.20)$$

For large observation times, the improvement in SNR approaches infinity.

Example 5.1

The resting potential of a cell membrane is to be estimated by means of the system depicted in Figure 2. The cell membrane potential is measured by the two high impedance glass electrodes and the amplifier. The signal $s(t)$ is the amplified membrane potential and the noise $n(t)$ is the noise (referred to the output) of the electrodes, the amplifier, and the noise picked up by the electrodes.

We assume that the additive noise is zero mean with autocorrelation function

$$r_n(\tau) = \sigma_n^2 \text{Exp} (-\alpha |\tau|) \quad (5.21)$$

Using Equation 5.13, the variance of the $\hat{\mu}_x$ can be calculated:

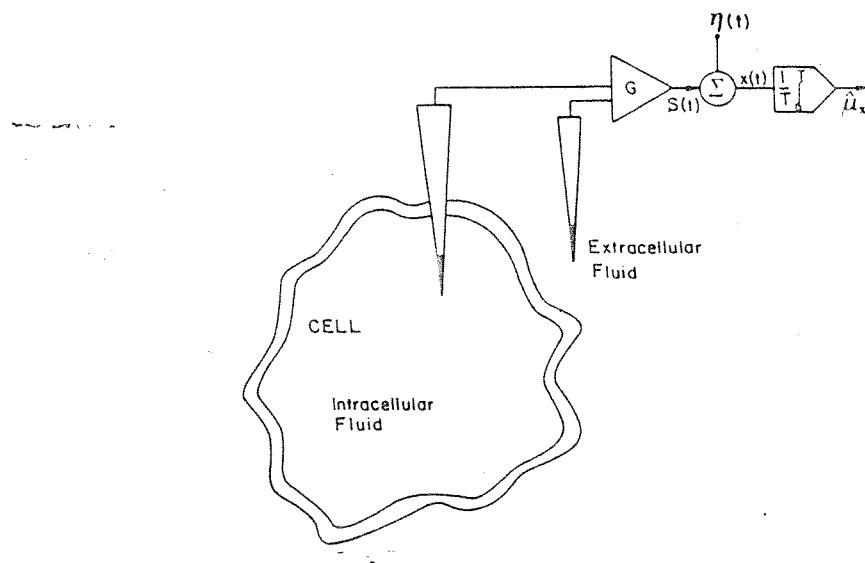


FIGURE 2. Measurement system for membrane potential.

$$\text{Var} [\hat{\mu}_x] = \frac{2\sigma_n^2}{\alpha T} \left[1 - \frac{1}{\alpha T} (1 - \text{Exp}(-\alpha)) \right] \quad (5.22)$$

and by using Equation 5.20, the signal-to-noise ratio improvement is written:

$$\frac{\text{SNR}_0}{\text{SNR}_i} = \frac{\alpha T}{2 \left[1 - \frac{1}{\alpha T} (1 - \text{Exp}(-\alpha T)) \right]} \quad (5.23)$$

For very large observation times, the improvement becomes

$$\lim_{T \rightarrow \infty} \frac{\text{SNR}_0}{\text{SNR}_i} = \frac{\alpha T}{2} \quad (5.24)$$

For large observation times, T , the improvement in SNR becomes linear.

The estimator in Equation 5.3 uses a finite time integrator. For long observation times, Equation 5.3 is difficult to implement. Consider, instead, a low pass filter with an impulse response $h(t)$. This filter replaces the integrator such that the estimation of the quantity μ_x in Equation 5.2 is given by

$$\hat{\mu}_x(T) = \int_0^T h(\tau) x(T - \tau) d\tau \quad (5.25)$$

Define a new filter impulse response, $h(t, T)$, which is equal to the low-pass filter's impulse response throughout the observation interval and is zero elsewhere; hence,

$$h(t, T) = \begin{cases} h(t); & 0 \leq t \leq T \\ 0; & \text{elsewhere} \end{cases} \quad (5.26)$$

Introducing Equation 5.26 into Equation 5.25,

$$\hat{\mu}_x(t) = \int_{-\infty}^{\infty} h(\tau, T) x(T - \tau) d\tau \quad (5.27)$$

We would like to investigate the mean square error of the estimator. The expectation of the filter's output is

$$E\{\hat{\mu}_x\} = \mu_x \int_{-\infty}^{\infty} h(\tau, T) d\tau \quad (5.28)$$

which is biased. The expectation of its square is

$$E\{(\hat{\mu}_x)^2\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(v, T) h(u, T) r_x(u - v) dv du \quad (5.29)$$

hence, the variance

$$\begin{aligned} \text{Var}[\hat{\mu}_x] &= E\{(\hat{\mu}_x)^2\} - [E\{\hat{\mu}_x\}]^2 = \\ &\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(v, T) h(u, T) [r_x(u - v) - \mu_x^2] dv du \end{aligned} \quad (5.30)$$

Substituting $\tau = u - v$ and defining

$$r_h(\tau, T) = \int_{-\infty}^{\infty} h(v, T) h(v + \tau, T) dv \quad (5.31)$$

we get from Equation 5.30

$$\begin{aligned} \text{Var}[\hat{\mu}_x] &= \int_{-\infty}^{\infty} r_h(\tau, T) [r_x(\tau) - \mu_x^2] d\tau = \\ &\int_{-\infty}^{\infty} r_h(\tau, T) \cdot r_n(\tau) d\tau \end{aligned} \quad (5.32)$$

Comparing Equations 5.32 and 5.13 we see that when an ideal integrator is used, the variance was calculated by integrating $r_n(\tau)$ with a weighting function

$$\frac{1}{T} \left(1 - \frac{|\tau|}{T} \right)$$

while here, with a low-pass filter, the weighting function becomes $r_h(\tau, T)$.

Example 5.2

Repeat the problem discussed in Example 5.1 with a simple RC low-pass filter replacing the integrator.

Denoting $(RC)^{-1} = \beta$, the impulse response for the filter is given by

$$h(t) = \beta \exp(-\beta t); \quad t \geq 0 \quad (5.33)$$

The weighting function $r_h(\tau, T)$ for this case is

$$r_h(\tau, T) = \begin{cases} 0; & \tau > T \\ \int_0^{T-\tau} h(v) h(v + \tau) dv; & 0 \leq \tau \leq T \\ \int_{-\tau}^T h(v) h(v + \tau) dv; & -T \leq \tau \leq 0 \\ 0; & \tau < -T \end{cases} \quad (5.34)$$

Introducing Equation 5.33 we get

$$r_h(\tau, T) = \frac{1}{2} \begin{cases} 0; & \tau > T \\ \beta (\exp[-\beta\tau] - \exp[-\beta(2T - \tau)]); & 0 \leq \tau \leq T \\ \beta (\exp[\beta\tau] - \exp[-\beta(2T + \tau)]); & -T \leq \tau \leq 0 \\ 0; & \tau < -T \end{cases} \quad (5.35)$$

Assume, again, that the noise is exponentially correlated (Equation 5.21); then, from Equation 5.32:

$$\text{Var} [\hat{\mu}_x] = \sigma_n^2 \int_{-\infty}^{\infty} r_h(\tau, T) \exp[-\alpha|\tau|] d\tau$$

Defining the dimensionless parameters,

$$\psi = \alpha/\beta \text{ and } \phi = T \cdot \beta \quad (5.36)$$

The improvement in signal-to-noise ratio can be written as

$$\frac{\text{SNR}_o}{\text{SNR}_i} = \begin{cases} \frac{(\psi^2 - 1)}{(\psi - 1)(1 - \exp[-\phi(\psi + 1)] - (\psi + 1)\exp(-2\phi)(1 - \exp(-\phi(\psi - 1)))}} & \psi \neq 1 \\ \frac{1}{0.5 - \exp[-2\phi](0.5 + \phi)}; & \psi = 1 \end{cases} \quad (5.37)$$

For long observation times, we get from Equation 5.37

$$\lim_{\phi \rightarrow \infty} \frac{\text{SNR}_o}{\text{SNR}_i} = \psi + 1 = \frac{\alpha}{\beta} + 1 \quad (5.38)$$

Comparing Equations 5.24 and 5.38, we note that while the signal-to-noise ratio can be infinitely improved by the integrator, it is bounded when using the low-pass filter estimator. Note also that in both cases the improvement is linearly proportional to the noise correlation exponential factor α .

B. The Discrete Case

Consider the case where the signal $x(t)$ of Equation 5.1 is given in the discrete form. We shall denote $x(t = k\Delta t)$ by x_k

$$x_k = \mu_x + n_k; \quad k = 0, 1, 2, \dots \quad (5.39)$$

where $\{x_k\}$ and $\{n_k\}$ are the signal and noise sample sequences.

The estimate of the mean $\hat{\mu}_x$ at time $k\Delta t$ is defined as the average over the current and $M - 1$ previous samples:

$$\hat{\mu}_x(k, M) = \frac{1}{M} \sum_{i=0}^{M-1} x_{k-i}; \quad k \geq M - 1 \quad (5.40)$$

The estimator is clearly unbiased. The variance of the estimator (Equation 5.5) is given by

$$\begin{aligned} \text{Var}[\hat{\mu}_x(k, M)] &= \\ &\frac{1}{M^2} E \left\{ \left(\sum_{i=0}^{M-1} x_{k-i} \right)^2 \right\} - \mu_x^2 = \\ &\frac{1}{M^2} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} E\{x_{k-i} x_{k-j}\} - \mu_x^2 \end{aligned} \quad (5.41)$$

Assuming $\{x_k\}$ is a stationary process, the last equation can be expressed in terms of the correlation function $r_\tau^x = E\{x_k x_{k+\tau}\}$

$$\text{Var}[\hat{\mu}_x(M)] = \frac{1}{M^2} \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} r_{|j-i|}^x - \mu_x^2 \quad (5.42)$$

Consider the symmetric ($M \times M$) correlation matrix

$$R_x = \begin{bmatrix} r_0^x & r_1^x & r_2^x & \dots & r_{M-1}^x \\ r_{-1}^x & r_0^x & r_1^x & \dots & r_{M-2}^x \\ \vdots & & & & \vdots \\ \vdots & & & & \vdots \\ r_{1-M}^x & \dots & \dots & \dots & r_0^x \end{bmatrix} \quad (5.43)$$

the elements of which are the correlation coefficients used in the first term in the right side of Equation 5.42. This term, therefore, is the sum of the matrix elements. Due to the symmetry of the matrix, the variance becomes

$$\text{Var}[\hat{\mu}_x(M)] = \frac{1}{M} \left[r_0^x + \frac{2}{M} \sum_{\tau=1}^{M-1} (M - \tau) r_\tau^x \right] - \mu_x^2 \quad (5.44)$$

and by introducing Equation 5.12 in its discrete form,

$$\text{Var}[\hat{\mu}_x(M)] = \frac{1}{M} \left[r_0^n + \frac{2}{M} \sum_{\tau=1}^{M-1} (M - \tau) r_\tau^n \right] \quad (5.45)$$

The last equation is the equivalent of Equation 5.13 for the discrete case.

The improvement in SNR for the discrete case is thus given by

$$\frac{\text{SNR}_0}{\text{SNR}_i} = \frac{\text{Var}[n_k]}{\frac{1}{M} \left[r_0^n + \frac{2}{M} \sum_{\tau=1}^{M-1} (M-\tau) r_\tau^n \right]} \quad (5.46)$$

Example 5.3

Consider again the problem given in Example 5.1. Assume now that the noisy signal $x(t)$ is given in terms of its finite samples sequence $\{x_i\}$: $t = k - M + 1, k - M + 2, \dots, k - 1, k$.

Assume also the noise process has an exponential correlation function as in Example 5.1 (Equation 5.21). From Equations 5.18, 5.19, and 5.46, we get the improvement in signal-to-noise ratio due to estimator 5.40:

$$\frac{\text{SNR}_0}{\text{SNR}_i} = \frac{M}{1 + 2 \sum_{\tau=1}^{M-1} (1 - \tau/M) \exp(-\alpha\tau)} \quad (5.47)$$

For long observation times, Equation 5.47 becomes

$$\lim_{M \rightarrow \infty} \frac{\text{SNR}_0}{\text{SNR}_i} = M \frac{1 - \exp(-\alpha)}{1 + \exp(-\alpha)} \quad (5.48)$$

The improvement in signal-to-noise ratio is linearly proportional to the observation time M , as is the case in Example 5.1. Figure 3 depicts the effect of estimating the expectation of a random process by means of the estimator 5.40.

III. ESTIMATION OF THE VARIANCE AND CORRELATION¹

A. Variance Estimation — The Continuous Case

Consider the sample record from an ergodic random process given in Equation 5.1. It is required to estimate the variance of $\{x(t)\}$ given finite observation time.

Denote the desired variance by S_x such that

$$S_x = E\{(x(t) - E\{x(t)\})^2\} \quad (5.49)$$

The finite time estimator $\hat{S}_x(T)$ is given by

$$\hat{S}_x(T) = \frac{1}{T} \int_0^T (x(t) - \hat{\mu}_x)^2 dt \quad (5.50)$$

where $\hat{\mu}_x$ is the estimator for the expectation of $\{X(t)\}$ given by Equation 5.3. It can be shown, using Equations 5.8 and 5.12 that the expectation of Estimator 5.50 can be written as

$$E\{\hat{S}_x\} = S_x - \frac{2}{T^2} \int_0^T \xi r_n(\xi) d\xi \quad (5.51)$$

The variance estimator (Equation 5.50) is a biased estimator. In most practical cases, however, the noise process is such that

$$\lim_{T \rightarrow \infty} \frac{1}{T^2} \int_0^T \xi r_n(\xi) d\xi = 0 \quad (5.52)$$

For these cases, the bias term approaches zero for large observation times.

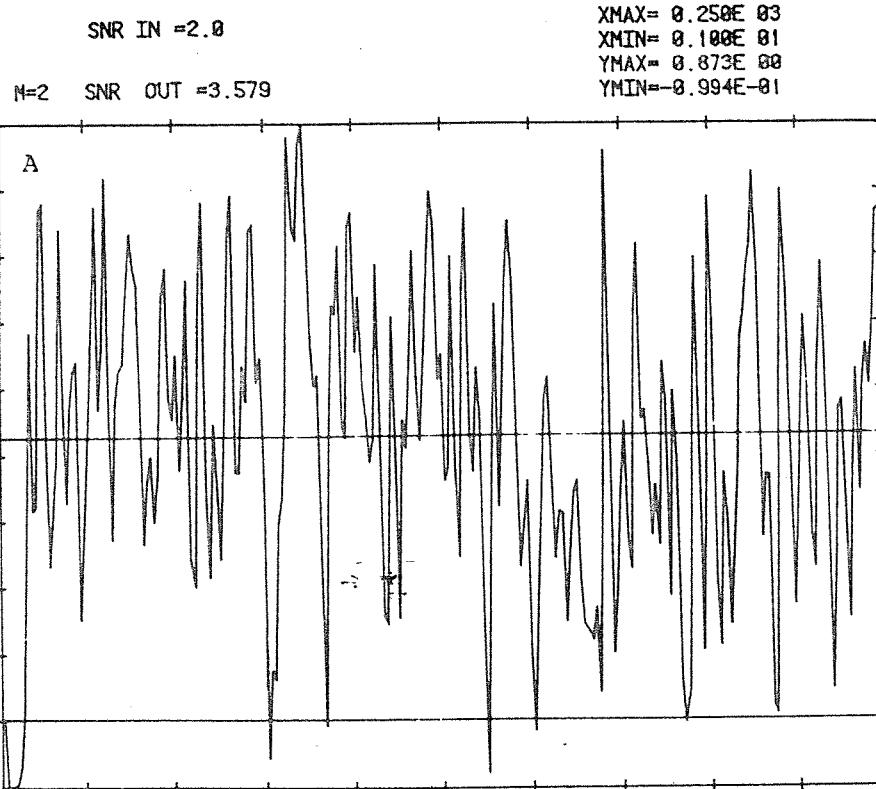


FIGURE 3. Finite time averaging of membrane voltage. (A) $M = 2$, $\text{SNR}_0 = 3.58$; (B) $M = 16$, $\text{SNR}_0 = 28.15$; (C) $M = 64$, $\text{SNR}_0 = 52.72$.

The variance of the estimator (Equation 5.50) is a function of the fourth moment of the process $\{x(t)\}$ and thus its calculation is of a limited practical value.

B. Variance Estimation — The Discrete Case

Consider now the discrete case where the variance of $\{x(t)\}$ is to be estimated from the finite sample sequence

$$\{x_i\}; \quad t = k - M + 1, \quad k - M + 2, \dots, \quad k - 1, \quad k$$

Define the variance estimator as

$$\hat{S}_x(M) = \frac{1}{M - 1} \sum_{i=0}^{M-1} (x_{k-i} - \hat{\mu}_x(k - i))^2 \quad (5.53)$$

where $\hat{\mu}_x$ is the discrete estimation of $E[x(t)]$ given by Equation 5.40.

The expectation of the estimator 5.53 is

$$\begin{aligned} E\{\hat{S}_x(M)\} &= \frac{1}{M - 1} \sum_{i=0}^{M-1} E\{(x_{k-i} - \hat{\mu}_x(k - i))^2\} = \\ S_x &= \frac{2}{M(M - 1)} \sum_{\tau=1}^{M-1} \tau r_n(\tau) \end{aligned} \quad (5.54)$$

Equation 5.54 is the discrete equivalent of Equation 5.51. The estimator is a biased one.

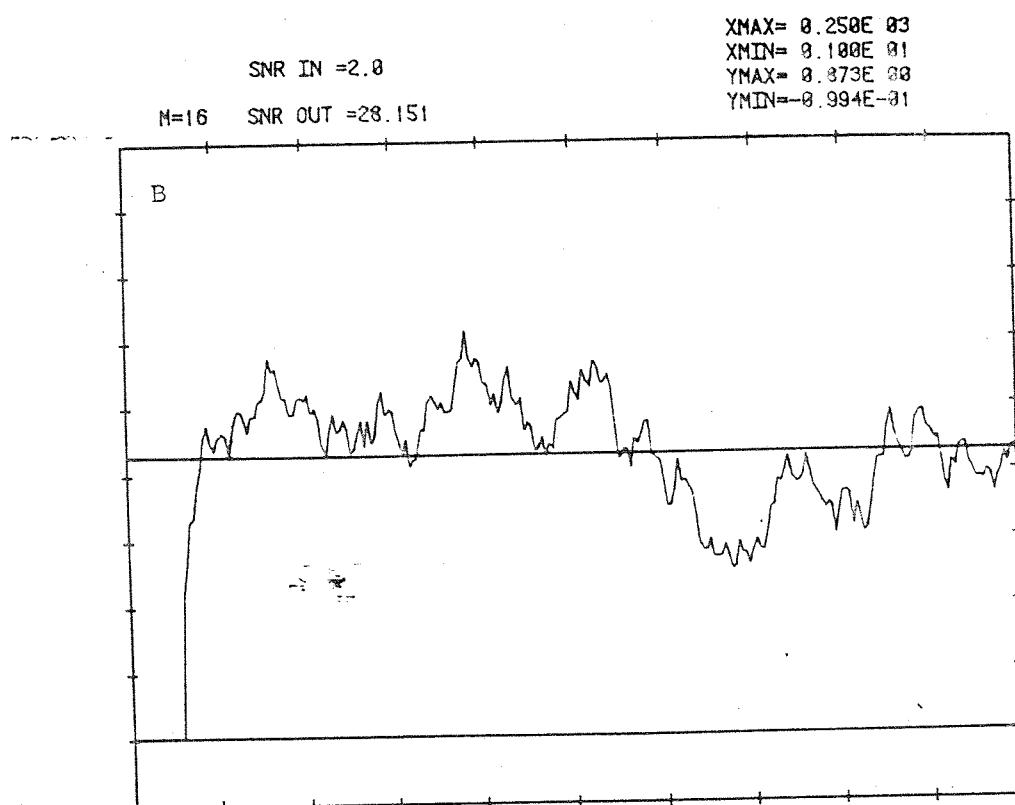


FIGURE 3B.

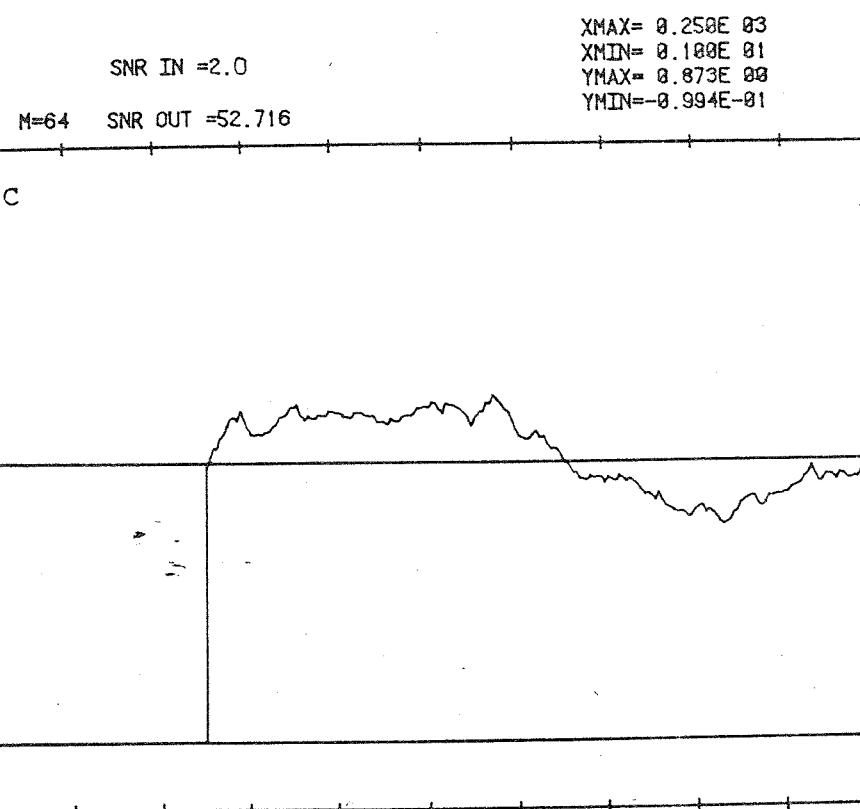


FIGURE 3C.

For most practical noise processes, however, the bias term in Equation 5.54 approaches zero for long observation times. As in the continuous case, the variance of the estimator (Equation 5.53) depends on the fourth moment of the process. Its calculation is thus of very limited practical value.

Note that for uncorrelated noise process, both the continuous estimator (Equation 5.50) and the discrete one (Equation 5.53) are unbiased.

C. Correlation Estimation¹

The autocorrelation function $r_x(\tau)$ of the ergodic process $\{x(t)\}$ and the function $r_{xy}(\tau)$ (which is the cross-correlation function between the two ergodic processes $\{x(t)\}$ and $\{y(t)\}$) are of great importance in signal processing. It is often required to estimate the correlation functions given only finite sample functions of the processes.

When the sample functions $x(t)$, $y(t)$ are given in the finite time $t \in (0, T + \tau)$ the estimator for the cross correlation can be defined by

$$\hat{r}_{xy}(\tau, T) = \frac{1}{T} \int_0^T x(t) y(t + \tau) dt \quad (5.55)$$

The expected value of the estimator is

$$E\{\hat{r}_{xy}(\tau, T)\} = r_{xy}(\tau) \quad (5.56)$$

Hence, the estimator (Equation 5.55) is an unbiased estimator of the correlation function $r_{xy}(\tau)$. The mean square error is given by the variance of the estimator. The variance depends on the fourth moment of the processes.

It can be shown that for jointly gaussian processes with zero means the variance becomes

$$\text{Var}[\hat{r}_{xy}(\tau, T)] = \frac{1}{T} \int_0^T \left[\left(1 - \frac{|\xi|}{T} \right) r_x(\xi) r_y(\xi) + r_{xy}(\xi + \tau) r_{yx}(\xi - \tau) \right] d\xi \quad (5.57)$$

If $r_x(\xi)$, $r_y(\xi)$, and $r_{xy}(\xi)$ are absolutely integrable over $(-\infty, \infty)$, the variance of the estimator approaches zero as the observation time increases. Thus, for these cases, the estimator is a consistent one.

The estimation of the autocorrelation function is performed in a similar manner with results given by Equations 5.56 and 5.57 with $\{y\}$ replaced by $\{x\}$. The correlation function can be efficiently estimated indirectly. The power spectrum density function is first estimated (Chapter 8) and then subjected to an inverse FFT to yield the correlation estimation.

IV. SYNCHRONOUS AVERAGING (CAT-COMPUTED AVERAGED TRANSIENTS)

A. Introduction

In many biomedical applications, the problem of detecting a repetitive random signal heavily corrupted with noise arises. It is often possible to trigger the signal by controlling the cause. This may be the case when analyzing evoked potentials (EP) in the EEG. The stimulus can be repeated at known times, and the response can be analyzed by methods discussed in this chapter.

Let the finite, stochastic response to the i th stimulus be denoted by

$$S_i(t - t_i) = \begin{cases} S_i(\tau); & 0 \leq \tau \leq T_i \\ 0; & \text{otherwise} \end{cases}$$

where t_i is the time of initiation of the i th stimulus and T_i is the length of the i th response.

The observed signal to be analyzed, $z(t)$, can be expressed as a series of the responses $s_i(t)$ corrupted with additive noise

$$z(t) = \sum_{i=1}^{\infty} S_i(t - t_i) + n(t) \quad (5.58)$$

where $n(t)$ is a zero mean process, statistically independent of $S_i(t)$. Let the observation $z_i(t)$ be defined as

$$z_i(t) = s_i(t - t_i) + n_i(t - t_i); \quad t_i \leq t \leq t_i + T \quad (5.59)$$

where $n_i(t) = n(t)$; $t \in [t_i, t_i + T]$ and T is larger or equal the length of the longest response

$$T \geq \text{Max } (T_i) \quad (5.60)$$

Assume also that the autocorrelation function of the noise obeys

$$r_n(T_j - T_i) = 0; \quad \text{for all } i, j \quad (5.61)$$

We are interested in estimating the average response $S(N,t)$ over N stimuli;

$$S(N,t) = \frac{1}{N} \sum_{i=1}^N E\{S_i(t - t_i)\} \quad (5.62)$$

Let the estimate of the required average response be

$$\hat{S}(N,t) = \frac{1}{N} \sum_{i=1}^N z_i(t) = \frac{1}{N} \sum_{i=1}^N (S_i(t - t_i) + n_i(t - t_i)) \quad (5.63)$$

Assuming stationary zero mean noise process, with variance σ_n^2 , the expectation of the estimator is

$$E\{\hat{S}(N,t)\} = S(N,t) \quad (5.64)$$

hence, the estimator is unbiased.

The variance of the estimator, σ_s^2 , is given by

$$\begin{aligned} \sigma_s^2 &= E\{\hat{S}^2(N,t)\} - (E\{S(N,t)\})^2 = \\ &\frac{1}{N^2} \left[E\left\{ \sum_{i=1}^N z_i^2(t - t_i) + \sum_{i=1}^N \sum_{j=1, j \neq i}^N z_i(t - t_i) z_j(t - t_j) \right\} \right] - (E\{S(N,t)\})^2 \quad (5.65) \end{aligned}$$

In general, the responses $S_i(t)$ are dependent on one another. This may be (in the case of EP analysis) due to phenomena like learning or fatigue.

Define the nonstationary spatial cross correlation function $r_s(\tau, t)$ by

$$r_s(\tau, t) = E\{S_i(t - \tau) S_j(t)\}$$

$$\tau = j - i \quad (5.66)$$

Denote the expectation of the response by

$$E\{S_i(t)\} = m(t); \text{ for all } i \quad (5.67)$$

and its variance

$$\sigma_s^2 = E\{S_i^2(t)\} - m^2(t); \text{ for all } i \quad (5.68)$$

Noting that

$$r_s(0, t) = E\{S_i^2(t)\} = \sigma_s^2 + m^2(t) \quad (5.69)$$

and

$$E\{z_i^2(t - \tau)\} = r_s(\tau, t) + \sigma_n^2 \quad (5.70)$$

and

$$\sum_{\substack{i,j=1 \\ \tau \neq 0}}^N r_s(\tau, t) = \sum_{\tau=1}^{N-1} (N - \tau) r_s(\tau, t) \quad (5.71)$$

it can be shown that

$$\sigma_s^2 = \frac{1}{N} (\sigma_s^2 + \sigma_n^2) + \frac{2}{N^2} \sum_{\tau=1}^{N-1} (N - \tau) [r_s(\tau, t) - m^2(t)] \quad (5.72)$$

Equation 5.72 gives the variance of the estimator in the general case, with the assumption of statistical independence between noise and response.

Consider now two important cases.

B. Statistically Independent Responses

For EP analysis, for example, this assumption can be made if the number of stimuli, N , is such that the influence of learning and fatigue can be neglected. Hence, we shall assume the above assumption to be valid for $N \leq N_{max}$.

For this case, we have for $N \leq N_{max}$

$$r_s(k, \tau) = \begin{cases} E\{S_i^2(t - \tau)\}; & k = 0 \\ m^2(t); & \text{otherwise} \end{cases} \quad (5.73)$$

hence, for Equation 5.72

$$\sigma_s^2 = \frac{1}{N} (\sigma_s^2 + \sigma_n^2) = \frac{1}{N} \sigma_s^2 ; \quad N \leq N_{\max} \quad (5.74)$$

which states that the variance of the estimator can be made as small as required by increasing the number of responses participating in the averaging window. In most practical cases, the variance cannot be made zero due to the limit on N .

The signal-to-noise ratio of the observation $z_i(t)$ is defined by the ratio of the mean response (the "signal") to the variance of the observation:

$$\text{SNR}_z = \frac{S(N,t)}{\sigma_z^2} \quad (5.75)$$

Similarly, the signal-to-noise ratio of the estimator is

$$\text{SNR}_s = \frac{S(N,t)}{\sigma_s^2} \quad (5.76)$$

Using Equations 5.74, 5.75, and 5.76, the improvement in signal-to-noise ratios can be written as

$$\frac{\text{SNR}_s}{\text{SNR}_z} = N; \quad N \leq N_{\max} \quad (5.77)$$

Note that the improvement is independent of the signal to noise ratio of the observation $z(t)$.

C. Totally Dependent Responses

In this case we assume that $S_i(t) = S(t)$, namely, all responses are samples of the same random process. Hence, Equation 5.66 becomes

$$r_s(\tau, t) = E \{ S_i(t - t_i) S_j(t - t_j) \} = E \{ S_i^2(t - t_i) \} \quad (5.78)$$

Substituting Equation 5.78 into Equation 5.72 we get

$$\sigma_s^2 = \sigma_s^2 + \frac{1}{N} \sigma_n^2 \quad (5.79)$$

Note that in this case the estimator is not a consistent estimator. Its variance can only approach the variance of the signal; as N approaches infinity, it will not approach zero.

The improvement in signal-to-noise ratios in this case can be written as

$$\frac{\text{SNR}_s}{\text{SNR}_z} = N \frac{\rho + 1}{N\rho + 1} \quad (5.80)$$

where $\rho = \frac{\sigma_s^2}{\sigma_n^2}$ is the ratio of variances of signal and noise in $Z_i(t)$. Note that in this case the improvement does depend on the signal-to-noise ratio of the observation. The improvement is always greater than 1. For very noisy observations (ρ approaches zero), the improvement approaches N .

D. The General Case

In some applications, it may not be possible to assume the basic assumptions made in the last sections.

An evoked response to a photic stimulus may be followed by a long persisting rhythmic wave. Assumption 5.60 may not hold in some cases. For example, when analyzing EPs, the main source of noise is the background EEG. The assumption that the EEG and EP are independent is not always valid. Correlation between the two may arise when the subject under examination is aware of the experiment and is expecting a stimulus.

The synchronous averaging method described in this chapter estimates the average response assuming the responses lie in the T time window. It often happens that the responses are not exactly synchronized with stimuli but have a latency τ such that

$$z_i(t) = S_i(t - \tau_i - t_i) + n_i(t) \quad (5.81)$$

where τ_i is the latency of the i th response. Synchronizing the averaging process in Equation 5.63 to the known stimuli times t_i means that the averaged responses will not be properly aligned with one another. The estimator will thus yield a "smeared" template of the response. To overcome this problem, one must estimate the latency, and synchronize the observations z_i to times $(t_i + \tau_i)$ rather than t_i . One way to overcome this problem is discussed in the following section.

In some applications, the knowledge of the average response in Equation 5.62 is not sufficient. In these cases, one is required to analyze the single EP. Techniques such as sophisticated adaptive filtering (Chapter 9) and waveform detection (Chapter 1, Volume II) have to be applied.

Example 5.4

The estimator 5.63 is a random variable with mean that equals the desired quantity $S(n,t)$. The probability distribution of the estimator is unknown; however, with the use of very well-known bounds, confidence limits can be set for the design of the synchronized averaging.

Consider the Chebychev inequality:

$$\text{Prob}[(m(t) - k\sigma_s) \geq \hat{S}(N,t) \geq (m(t) + k\sigma_s)] \leq \frac{1}{k^2} \quad (5.82)$$

which states that the probability of the estimate to be outside the range of $\pm k\sigma_s$ from the mean $m(t)$ is less than or equal to $1/k^2$. Hence, the probability to have an estimate outside the range $\pm 3\sigma_s$ is less than or equal to 0.11. With a probability (confidence) of 0.889 ($=0.9$ or 90%), the error in the estimate will be in the range of $\approx 3\sigma_s$. The experimental requirement can be phrased as follows: determine the number of trials (N) required such that with certainty of 90%, the error in the estimate will be less than or equal to $\approx \epsilon$. For statistically independent responses, we have (Equation 5.74)

$$3\sigma_s = 3 \frac{\sigma_s}{(N)^{1/2}} \leq \epsilon; \quad N \leq N_{\max} \quad (5.83)$$

and the required number of trials

$$\frac{\text{SNR}_s}{\text{SNR}_n} = N \geq \frac{9\sigma_s^2}{\epsilon^2} = \frac{9}{\epsilon^2} (\sigma_s^2 + \sigma_n^2); \quad N \leq N_{\max} \quad (5.84)$$

For the second case, where the responses are statistically dependent, Equation 5.79 is used to get

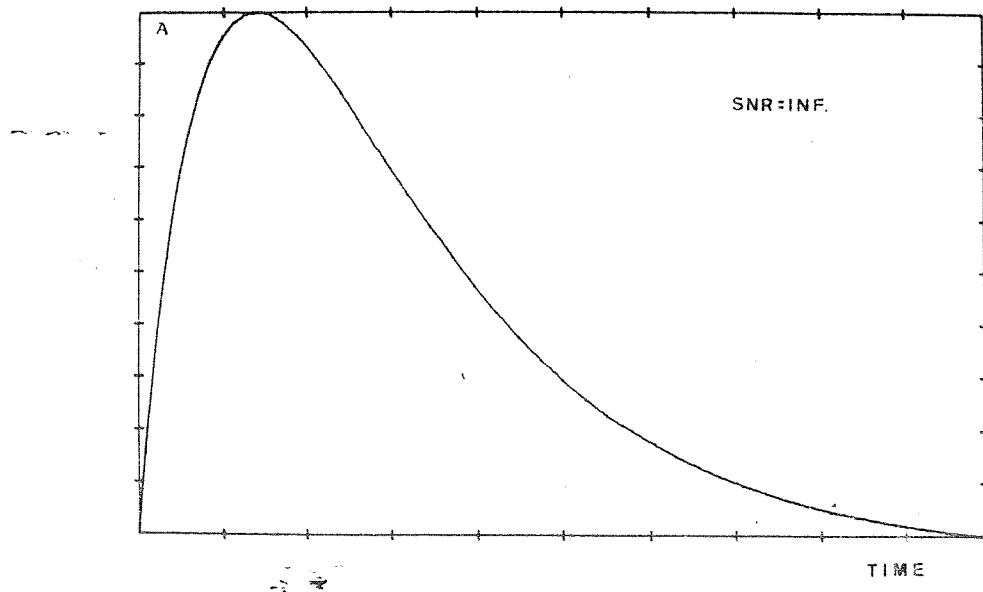


FIGURE 4. Synchronous averaging. (A) The signal ($\text{SNR} = \infty$); (B) raw data, signal with additive noise, $M = 1$; (C) averaging with $M = 200$; (D) averaging with $M = 800$; (E) signal-to-noise ratio vs. M . (See pages 62 and 63.)

$$3\sigma_s = 3|\sigma_s^2 + \frac{1}{N}\sigma_n^2|^{1/2} \leq \epsilon$$

$$\frac{\text{SNR}_s}{\text{SNR}_z} = N \geq \frac{\sigma_s^2}{\frac{1}{9}\epsilon^2 - \sigma_n^2}; \quad N \leq N_{\max} \quad (5.85)$$

Signal-to-noise ratio enhancement by means of synchronized averaging is demonstrated by simulation. A repetitive signal $S_i(t) = S(t)$ was generated by

$$S(t - t_i) = A \exp(-\alpha(t - t_i)) + B \exp(-\beta(t - t_i)); \\ 0 \leq (t - t_i) \leq T \quad (5.86)$$

The observation signal $z(t)$ was generated by adding uniformly distributed Pseudo Random noise. Figure 4 shows the signal $s(t)$, ($\text{SNR} = \infty$) and the estimator output for various N . The signals were sampled and the estimator (Equation 5.63) was implemented on the computer.

E. Records Alignment, Estimation of Latencies

Consider again the problem of latencies in the wavelets, described by Equation 5.81. If the unknown latencies, τ_i , have large variance, the averaged "smeared" response may not be sufficient. An adaptive method for estimating the latencies has been suggested.¹³⁻¹⁴ It is sometimes known as Woody filtering.

We shall proceed with the regular averaging method (Equation 5.63). Let us denote the "smeared" averaged response, $\hat{s}_o(N,t)$, hence:

$$\hat{s}_o(N,t) = \frac{1}{N} \sum_{i=1}^N (s_i(t - t_i) + n_i(t - t_i)) \quad (5.87)$$

We shall use $\hat{s}_o(N,t)$ as the first estimate of the average response. To improve on this

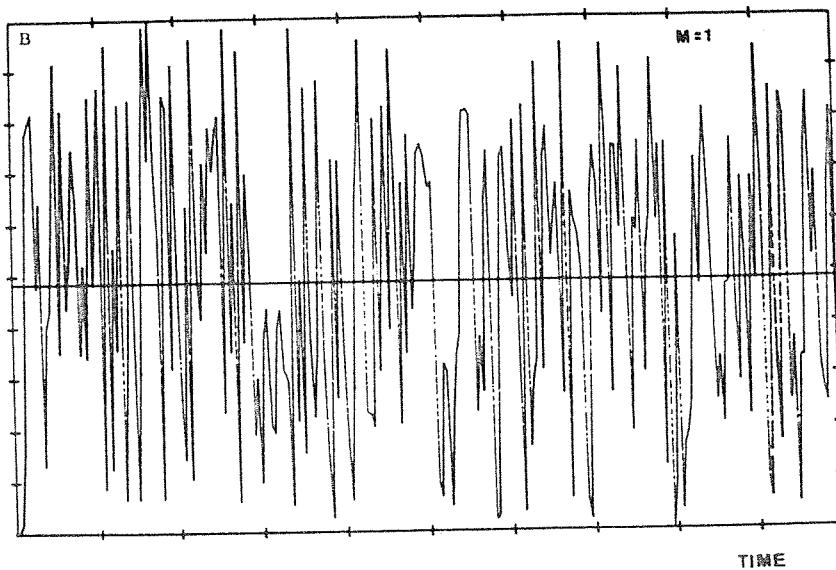


FIGURE 4B.

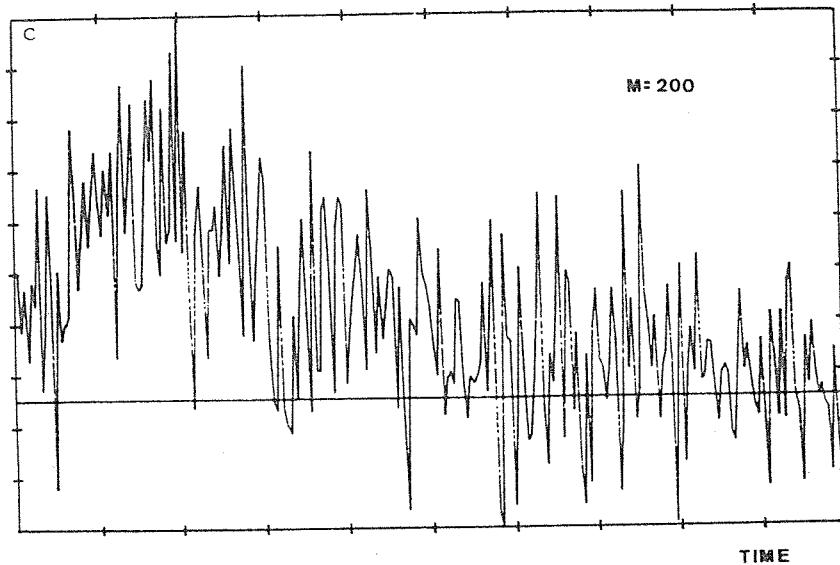


FIGURE 4C.

estimate, each of the N responses will be cross correlated with the estimate. Consider the correlation of the i th response:

$$R_{sz_i}^{(0)}(\lambda) = \frac{1}{T} \int_{t_0}^{T+\Delta t} \hat{s}_0(Nt) \cdot z_i(t - \lambda) dt \quad (5.88)$$

We shall look for the time λ_i for which the cross-correlation is maximum:

$$\text{Max}(R_{sz_i}^{(0)}(\lambda)) = R_{sz_i}^{(0)}(\lambda_i) \quad (5.89)$$

The time λ_i is the time shift required to best align the observation z_i with the estimate of

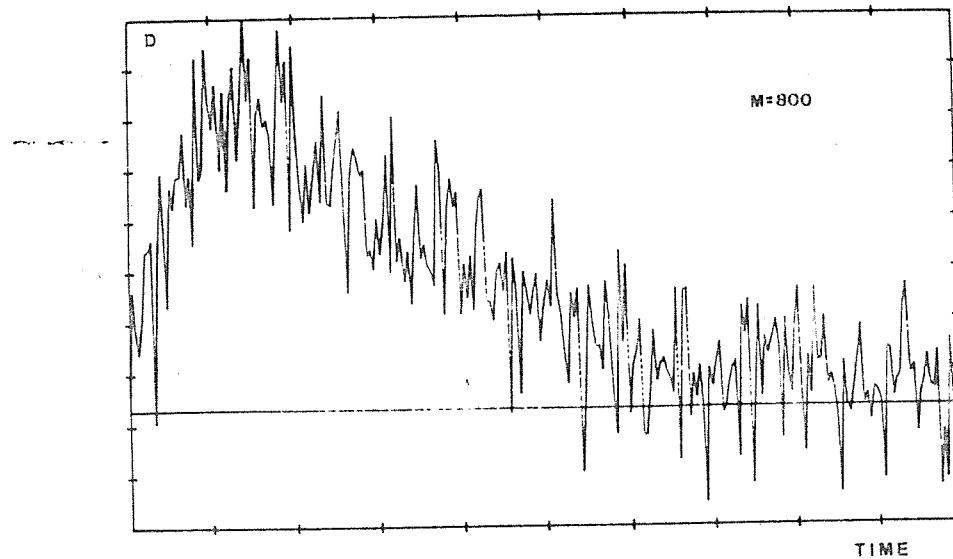


FIGURE 4D.

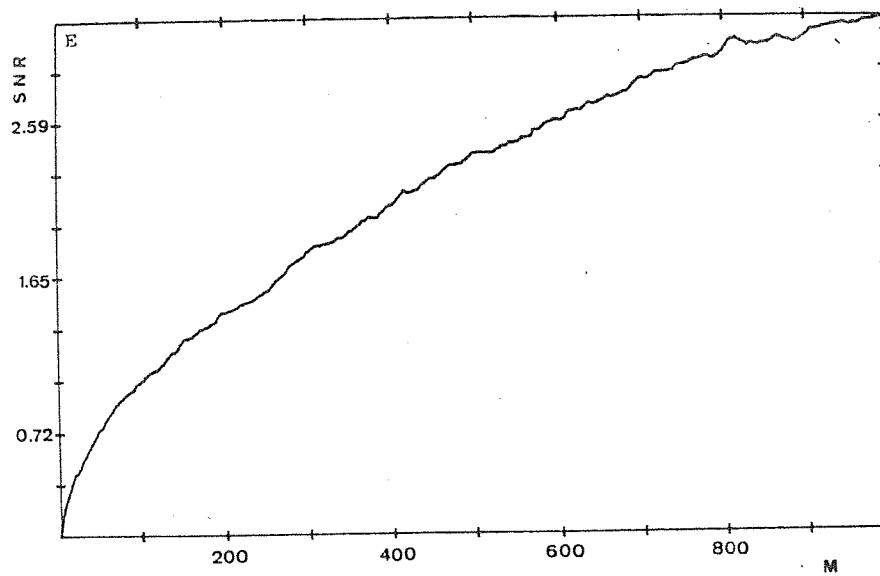


FIGURE 4E.

the average response. Therefore, we shall shift the observation $z_i(t)$ by that amount. The next estimation for the average is given by:

$$\hat{s}_i(N, \bar{t}) = \frac{1}{N} \sum_{i=1}^N (s_i(t - t_i - \lambda_i) + n(t - t_i - \lambda_i)) \quad (5.90)$$

We shall now repeat the correlation process of Equation 5.88 replacing $\hat{s}_o(N, t)$ by the new estimate $\hat{s}_i(N, t)$. The correlation of all N responses with the new estimate, $R_{sz}^{(1)}(\lambda)$, will provide new estimates, λ_i , for realignment. This procedure is repeated until some halting criterion is met. One such criterion can be the difference in the areas under two successive correlations, namely:

$$\left| \int_0^{\infty} R_{sz_i}^{(j+1)}(\lambda) d\lambda - \int_0^{\infty} R_{sz_i}^{(j)}(\lambda) d\lambda \right| \leq \epsilon \quad (5.91)$$

It has been shown that for a correlation function and noise which are well behaved, the above procedure results in an asymptotically stable estimate.¹⁴ In other cases, the procedure must be stopped after some arbitrary number of iterations; visual examination of the data may be recommended.

REFERENCES

1. Bendat, J. S. and Pierson, A. G., *Random Data Analysis and Measurements Procedures*, Wiley-Interscience, New York, 1971, chap. 6.
2. Ernst, R. R., Sensitivity enhancement in magnetic resonance. I. Analysis of the method of time averaging, *Rev. Sci. Instrum.*, 32 (12), 1689, 1965.
3. Bendat, J. S., Interpretation and application of statistical analysis for random physical phenomena, *IRE Trans. Bio-Med. Electron.*, 9, 31, 1962.
4. Davenport, W. B., Johnson, R. A., and Middleton, D., Statistical errors in measurements on random time functions, *J. Appl. Phys.*, 23 (4), 377, 1952.
5. Ruchkin, D. S., An analysis of average response computations based upon a periodic stimuli, *IEEE Trans. Eng.*, 12, 87, 1965.
6. Bendat, J. S., Mathematical analysis of average response values for nonstationary data, *IEEE Trans. Biomed. Eng.*, 11, 72, 1964.
7. Kovacs, Z. L., On the enhancement of the SNR of repetitive signals by digital averaging, *IEEE Trans. Instrum. Meas.*, 28 (2), 152, 1979.
8. Shiavi, R. and Green, N., Ensemble averaging of locomotor EMG patterns using interpolations, *Med. Biol. Eng. Comput.*, 21, 573, 1983.
9. Volkers, A. C. W., Van de Schee, E. J., and Grashuis, J. L., Electrogastrography in the dog: waveform analysis by a coherent averaging technique, *Med. Biol. Eng. Comput.*, 21, 56, 1983.
10. Shvartsman, V., Barnes, G. R., Shvartsman, L., and Flowers, N. G., Multichannel signal processing based on logic averaging, *IEEE Trans. Biomed. Eng.*, 29, 531, 1982.
11. Gethner, J. S., Woodin, R. L., Rabinowitz, P., and Kaldor, A., Multiparameter matrix signal averaging, *Rev. Sci. Instrum.*, 53 (9), 1398, 1982.
12. Thomas, C. W., Rzeszotarski, M. S., and Isenstein, B. S., Signal averaging by parallel digital filters, *IEEE Trans. Acoust. Speech Signal Process.*, 30, 338, 1982.
13. Woody, C. D., Characterization of an adaptive filter for the analysis of variable latency neuroelectric signals, *Med. Biol. Eng. Comput.*, 5, 539, 1967.
14. Senmoto, S. and Childers, D. G., Adaptive decomposition of a composite signal of identical unknown wavelets in noise, *IEEE Trans. Syst. Man Cybern.*, 2, 59, 1972.

Chapter 6

FREQUENCY DOMAIN ANALYSIS

I. INTRODUCTION

A. Frequency Domain Representation

Signals are most often given as a function of time. It is sometimes required to represent the signal in the frequency domain, where the distribution of amplitudes and phase are given with respect to the frequency. Frequency domain representation is advantageous in many applications such as filtering. The two representations are related via the *Fourier transform* (FT) or *Fourier integral* and *inverse Fourier transform* (IFT). Consider the signal $x(t)$ given in the time domain. Its representation in the frequency domain, $X(w)$, is given by the Fourier integral:

$$X(w) = \int_{-\infty}^{\infty} x(t) \exp(-jwt) dt \quad (6.1)$$

The last equation is symbolically written as $X(w) = F\{x(t)\}$.

The inverse transformation, from the frequency domain into the time domain, is given by the inverse Fourier transform:

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(w) \exp(jwt) dw \quad (6.2)$$

The last equation is symbolically written as $x(t) = F^{-1}\{X(w)\}$.

We consider signals in the time domain which are real. In the frequency domain, however, the representation $X(w)$ is, in general, a complex value of w . We write $X(w)$ as:

$$X(w) = |X(w)| \exp(j\theta(w)) \quad (6.3)$$

where $|X(w)|$ is the amplitude and $\Theta(w)$ is the phase of the frequency domain representation.

Not all real functions can be transformed into the frequency domain by Equation 6.1. Sufficient conditions for the existence of $X(w)$ are given by the Dirichlet conditions:

1. $\int_{-\infty}^{\infty} |x(t)| dt < \infty$, namely, $x(t)$ is absolutely integrable.
2. $x(t)$ has a finite number of discontinuities and a finite number of extrema points in any finite interval.

There are some very useful functions such as the impulse (delta) function, step function, or even sine and cosine functions which do not obey the Dirichlet conditions. These functions do not have a Fourier transform; they do, however, have the transform in the limit. We shall talk about the FT of these functions where, in fact, we shall mean the limit that does exist.

B. Some Properties of the Fourier Transform

The FT is a linear operator. If $X_1(w) = F\{x_1(t)\}$ and $X_2(w) = F\{x_2(t)\}$, then for any arbitrary constants, a_1 and a_2 ,

$$F\{a_1 x_1(t) + a_2 x_2(t)\} = a_1 X_1(w) + a_2 X_2(w) \quad (6.4)$$

The FT possesses many interesting properties that help in the calculations of direct and inverse transformations. The reader is referred to references and textbooks dealing in details with the subject.¹⁻² We shall deal here only with some important properties that bear direct relation to the material presented in later sections.

1. The Convolution Theorem

The convolution theorem is an important tool in frequency analysis. The convolution integral of two functions $x_1(t)$ and $x_2(t)$ is defined (for stationary functions) by:

$$x(t) = \int_{-\infty}^{\infty} x_1(\tau)x_2(t - \tau)d\tau \quad (6.5A)$$

The convolution integral is symbolically written as:

$$x(t) = x_1(t) * x_2(t) \quad (6.5B)$$

The importance of Equation 6.5 in systems analysis stems from the fact that the output of a linear system is given by the convolution of the input with the impulse response of the system. It can be easily shown that the FT of $x(t)$ is

$$F\{x(t)\} = F\{x_1(t) * x_2(t)\} = X_1(w) \cdot X_2(w) \quad (6.6)$$

where $X_i(w) = F\{x_i(t)\}$. Hence the convolution operation in the time domain, which is a relatively complex operation, becomes a simple multiplication operation in the frequency domain.

Consider now the convolution of two functions, $X_1(w)$ and $X_2(w)$, in the frequency domain:

$$X(w) = \int_{-\infty}^{\infty} X_1(u)X_2(w - u)du = X_1(w) * X_2(w) \quad (6.7)$$

It can be easily shown that the IFT of $X(w)$ is

$$x(t) = F^{-1}(X(w)) = 2\pi \cdot x_1(t) \cdot x_2(t) \quad (6.8)$$

Hence, convolution of two functions in the frequency domain is given by 2π times the multiplication of the two functions in the time domain.

2. Parseval's Theorem

Consider the energy, E, of the time function $x(t)$:

$$E = \int_{-\infty}^{\infty} x^2(t)dt \quad (6.9)$$

We shall replace $x(t)$ in Equation 6.9 by $F^{-1}\{X(w)\}$, hence:

$$E = \int_{-\infty}^{\infty} x(t) \left(\frac{1}{2\pi} \int_{-\infty}^{\infty} X(w) \exp(jwt) dw \right) dt \quad (6.9A)$$

Interchanging the order of integration, we get

$$E = \int_{-\infty}^{\infty} x^2(t) dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(w) X(-w) dw \quad (6.9B)$$

For a real $x(t)$, we have $X(w) = X^*(-w)$ (where the symbol * denotes complex conjugate), hence

$$E = \int_{-\infty}^{\infty} x^2(t)dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(w)|^2 dw \quad (6.10)$$

Equation 6.10 is known as *Parseval's theorem* which states that the total energy of the signal can be calculated by the integral in the time domain and in the frequency domain. Note that if the energy in a band of frequency $w_1 \leq w \leq w_2$ is required, we have to integrate over this band and the band of negative frequencies $-w_2 \leq w \leq -w_1$. Since $|X(w)|^2$ is an even function, in w , we get:

$$E_{w_1-w_2} = 2 \cdot \frac{1}{2\pi} \int_{w_1}^{w_2} |X(w)|^2 dw \quad (6.11)$$

We can define the *energy spectrum density*, $\tilde{S}(w)$, by:

$$\tilde{S}(w) = \frac{1}{\pi} |X(w)|^2 \quad (6.12)$$

Then Equation 6.11 can be written:

$$E_{w_1-w_2} = \int_{w_1}^{w_2} \tilde{S}(w) dw \quad (6.13)$$

and the total signal energy is

$$E = \int_0^{\infty} \tilde{S}(w) dw \quad (6.14)$$

3. Fourier Transform of Periodic Signals

A periodic signal does not satisfy the Dirichlet conditions and thus does not have a FT. However, if we assume the signal exists only in a finite time interval $(-\tau/2, \tau/2)$, then the FT does exist. We then find the FT in the limit as τ becomes infinite.

We can, however, calculate the FT of a periodic signal in another way. Consider the complex Fourier series of the signal:

$$x(t) = \sum_{n=-\infty}^{\infty} a_n \exp(jnw_o t) \quad (6.15)$$

where $T = 2\pi/w_o$ is the period of the signal, and a_n are the coefficients of the expansion. Applying the FT to Equation 6.15 yields:

$$X(w) = F \left\{ \sum_{n=-\infty}^{\infty} a_n \exp(jnw_o t) \right\} = 2\pi \sum_{n=-\infty}^{\infty} a_n \delta(w - nw_o) \quad (6.16)$$

Hence, the FT of a periodic signal is a train of impulse functions located at the harmonics of the fundamental frequency, w_o , each multiplied by 2π times the corresponding coefficient of the Fourier series.

C. Discrete and Fast Fourier Transforms (DFT, FFT)

Consider a band-limited time signal $x(t)$ that has been sampled at a sampling interval, T_s . Assume that the sampling frequency $w_s = 2\pi/T_s$ obeys the sampling theorem such that the sampled sequence contains all the information present in the signal.

We denote the finite sampled sequence by $\{x(nT_s)\}$, $n = 0, 1, \dots, N - 1$.

Define the *discrete Fourier transform* (DFT), as a linear operator on the sequence $\{x(nT_s)\}$ such that:

$$X\left(k \frac{w_s}{N}\right) = \sum_{n=0}^{N-1} x(nT_s) \exp(-j2\pi k \frac{n}{N}); \quad k = 0, 1, \dots, N - 1 \quad (6.17)$$

The sequence $\left\{X\left(k \frac{w_s}{N}\right)\right\}$ is, in general, a complex sequence. The transformation that maps the complex sequence $\left\{X\left(k \frac{w_s}{N}\right)\right\}$ back into the sequence $\{x(nT_s)\}$ is called the *inverse discrete Fourier transform* (IDFT):

$$x(nT_s) = \frac{1}{N} \sum_{k=0}^{N-1} X\left(k \frac{w_s}{N}\right) \exp(j2\pi n \frac{k}{N}); \quad n = 0, 1, \dots, N - 1 \quad (6.18)$$

Note that T_s and w_s/N are constants, hence we denote $x_n = x(nT_s)$ and $X_k = X\left(k \frac{w_s}{N}\right)$.

The transformations in Equations 6.17 and 6.18 are symbolically written as:

$$X_k = \text{DFT}\{x_n\}; \quad x_n = \text{IDFT}\{X_k\} \quad (6.19)$$

Let us now consider the sampled sequence $\{x(nT_s)\}$ as a time function generated by multiplying the signal $x(t)$ with the ideal sampler operator, $\delta_T(t)$ (see Equation 4.2). We denote the sampled signal by $x^*(t)$ and get:

$$x^*(t) = \sum_{n=0}^{N-1} x(nT_s) \delta(t - nT_s) \quad (6.20)$$

The FT of the signal $x^*(t)$, namely $X^*(w)$, is easily calculated, noting that:

$$F\{\delta(t - nT_s)\} = \exp(-jwnT_s) \quad (6.21)$$

Then, from Equation 6.20 we get:

$$X^*(w) = F\{x^*(t)\} = \sum_{n=0}^{N-1} x(nT_s) \exp(-jwnT_s) \quad (6.22)$$

Equation 6.22 describes the FT of the sampled signal. This FT, $X^*(w)$, is a continuous function of w . Let us sample, in the frequency domain, this continuous function at frequency intervals of $w_s/N = 2\pi/NT_s$. We shall get a sequence of FT samples $\left\{X^*\left(k \frac{w_s}{N}\right)\right\}$, $k = \dots, -1, 0, 1, \dots$

$$X^*\left(k \frac{w_s}{N}\right) = \sum_{n=0}^{N-1} x(nT_s) \exp\left(-j2\pi k \frac{n}{N}\right); \quad k = \dots, -2, -1, 0, 1, 2, \dots \quad (6.23)$$

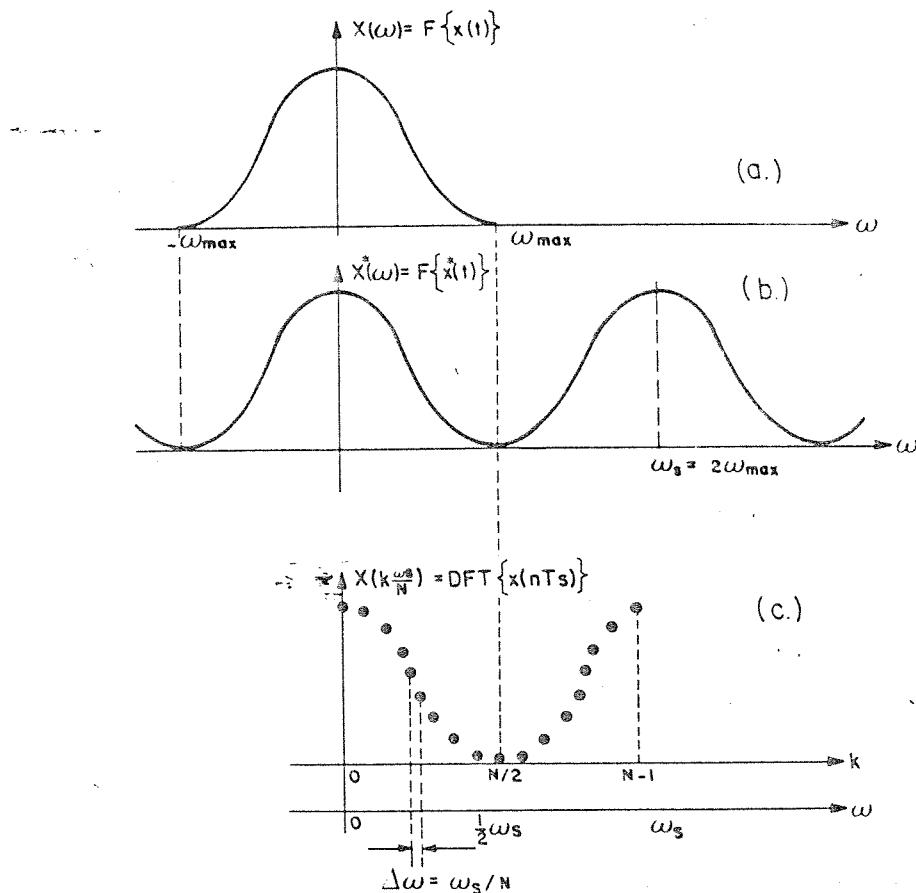


FIGURE 1. The relations between the Fourier transform (FT) of $x(t)$, the FT of $x^*(t)$, and the DFT. (A) The FT of $x(t)$; (B) the FT of $x^*(t)$; (C) the DFT.

Note that the set of N members $k = 0, 1, \dots, N - 1$ of the infinite sequence (Equation 6.23) equals the DFT of Equation 6.17.

We recall also from the discussion in Chapter 4 (Equation 4.4 and Figure 4.2) that the FT, $X^*(\omega)$, of the sampled signal is the repetition of the FT of the continuous signal $X(\omega)$ centered at $\ell\omega_s$. When we sample the FT, the samples $-N/2, \dots, -1, 0, 1, \dots, (N/2 - 1)$ are samples of the FT centered at $w = 0$. The rest of the samples of the sequence convey no new information since they represent the same samples shifted to $(w + \ell\omega_s)$, $\ell = \dots, -1, 1, \dots$. This can also easily be seen from Equation 6.23. The functions $\exp(-j2\pi k \frac{\omega}{N})$ are periodic functions with period N . Hence $X^*\left(k \frac{\omega_s}{N}\right) = X^*\left((k + \ell N) \frac{\omega_s}{N}\right)$ for any integer, ℓ . Since the FT of real signals is symmetric, we can represent the samples of the FT by the sequence $X^*\left(k \frac{\omega_s}{N}\right)$, $k = 0, 1, \dots, N - 1$.

As a conclusion we state that the DFT (Equation 6.17) is a sequence of N uniformly distributed samples of the DF of the signal, $x(t)$. Refer to Figure 1 where an example depicting the relations between the FT and the DFT is given. Here the signal was sampled at the Nyquist rate. Note that the DFT samples $k = 0, 1, \dots, N/2$ are the samples of the positive frequencies of the FT. The rest of the samples $k = (N/2 + 1), \dots, N - 1$ are the

samples of the negative frequencies of the FT centered at w_s . Since, in our cases, the FT is symmetric, these samples contribute no additional information. From the sequence of N samples of the DFT, we require only the $N/2 + 1$ first ones (or last ones); the rest are redundant. In the example given in Figure 1, $w_x = 2w_{\max}$, and hence for $k = N/2$ we get $X\left(k \frac{w_s}{N}\right) = X\left(\frac{w_s}{2}\right) = X(w_{\max})$. If the sampling rate were higher, $w_s > 2w_{\max}$, the frequency corresponding to $k = N/2$ would be higher than w_{\max} , where the FT is zero. In that case, we would get a range of zero DFT samples separating the two symmetric values of the FT.

The DFT has some properties which are similar in nature to those of the FT. The DFT is a linear operator, hence:

$$\text{DFT}\{a_1x_1(nT) + a_2x_2(nT)\} = a_1\text{DFT}\{x_1(nT)\} + a_2\text{DFT}\{x_2(nT)\} \quad (6.24)$$

The convolution of two signals was defined in Equation 6.5. A similar operation can be defined for two sequences. Let us define the *cyclic convolution* of the two sequences $\{x_1(n)\}$ and $\{x_2(n)\}$ by:

$$\begin{aligned} \{x(k)\} &= \{x_1(n)\} \circledast \{x_2(n)\} = \sum_{n=0}^{N-1} x_1(n)x_2(k-n); \\ k; n &= 0, 1, \dots, N-1 \end{aligned} \quad (6.25)$$

where the symbol \circledast denotes cyclic convolution.

It can be easily shown that:

$$\text{DFT}\{x(k)\} = \text{DFT}\{x_1(n) \circledast x_2(n)\} = \text{DFT}\{x_1(n)\} \cdot \text{DFT}\{x_2(n)\} \quad (6.26)$$

which is known as the *cyclic convolution theorem*.

Consider now the *discrete Parseval's theorem*. Let us define the energy of the real sequence $\{x(n)\}$ by:

$$E = \sum_{n=0}^{N-1} x(n)^2 \quad (6.27)$$

It can be easily shown that if $X(k) = \text{DFT}\{x(n)\}$, then:

$$E = \sum_{n=0}^{N-1} x(n)^2 = \frac{1}{N} \sum_{k=0}^{N-1} |X(k)|^2 \quad (6.28)$$

The DFT is an important tool for discrete signal processing for the same reasons the FT was important for continuous signal processing. The direct computation of the DFT requires approximately N^2 complex multiplication and addition operations. In 1965, Cooley and Tukey, in their famous paper, presented an efficient method for calculating the DFT. Their method, known as the *fast Fourier transform* (FFT), requires only $N \log_2 N$ operations (where N is a power of 2). For $N = 1024$, the number of operations required by the FFT is ten times less than the number required for direct computation.

Many different FFT algorithms have been derived for software and hardware implementations. Two commonly used algorithms are known as the *decimation in time* and *decimation in frequency* algorithms. The interested reader is referred to the vast literature on this subject.³⁻⁵

II. SPECTRAL ANALYSIS

A. The Power Spectral Density Function

In many applications, we shall be interested in the distribution of the energy of the signal in the frequency domain, rather than the distributions of amplitude and phase. When dealing with energy and power distributions, we lose information concerning the phase of the signal. In the previous section, the energy spectrum density function was introduced (Equation 6.12). For a random process, however, this definition has to be reconsidered. All the samples of the random process are assumed to exist in the range $-\infty \leq t \leq \infty$ and cannot be described analytically. In addition, each sample function is different from other samples of the same process. Since the random process is given in terms of its statistical properties, it makes sense to define the power distribution also in these terms.

Consider the sample function $x(t)$ of a stationary process in the range $-T/2 \leq t \leq T/2$ and its FT, $X_T(w)$, and define the *power spectral density* (PSD) (or the *spectrum*) by:

$$S_x(w) = \lim_{T \rightarrow \infty} \frac{E\{|X_T(w)|^2\}}{T} \quad (6.29)$$

In Equation 6.29, we use the expected value of the random variable $|X_T(w)|^2$ which is the energy. By introducing the FT we get:

$$|X_T(w)|^2 = X_T(w)X_T^*(w) = \int_{-T/2}^{T/2} \int x(t_1)x(t_2)\exp(-jw(t_2 - t_1))dt_1 dt_2 \quad (6.30)$$

Introducing Equation 6.30 into Equation 6.29 yields:⁶

$$S_x(w) = \int_{-\infty}^{\infty} r_x(\tau)\exp(-jw\tau)d\tau = F\{r_x(\tau)\} \quad (6.31A)$$

and

$$r_x(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_x(w)\exp(jw\tau)dw = F^{-1}\{S_x(w)\} \quad (6.31B)$$

where $r_x(\tau)$ is the autocorrelation function of the process:

$$r_x(\tau) = E\{x(t)x(t + \tau)\} \quad (6.32)$$

Equations 6.31A and B state that the PSD and the autocorrelation are a Fourier pair. Equations 6.31A and B are known as the *Wiener-Khinchin relations*. Note that since the autocorrelation function is an even function, the PSD is real.

Consider a stationary random signal $x(t)$ which has an autocorrelation function:

$$r(\tau) = \alpha\delta(\tau) \quad (6.33A)$$

Namely, the values of the signal $x(t)$ at the time t and at the time $t + \tau$ (for all τ not equal zero) are uncorrelated. The PSD function of such a process is

$$S(w) = \alpha \quad (6.33B)$$

The power is equally distributed along the frequency axis, hence the process is called *white noise*. A random process with power unequally distributed is called *colored noise*.

One principal application of the PSD function is related to the analysis of linear systems. Consider a linear system with an impulse response, $h(t)$, driven at its input by a random sample function, $u(t)$. The output of the system, $x(t)$ is given by the convolution:

$$x(t) = u(t) * h(t) \quad (6.34)$$

Consider an input signal that is stationary in the wide sense. We then calculate the autocorrelation of $x(t)$ and take its FT; the result is:⁷

$$S_x(w) = |H(w)|^2 S_u(w) \quad (6.35)$$

Hence, when the input, $u(t)$, is white, $S_u(w)$ is constant. The output signal $x(t)$ is a nonwhite noise, colored by the frequency response of the system.

B. Cross-Spectral Density and Coherence Functions

We are often interested in the relations between two random processes. As an example, we can consider the EEG signal recorded simultaneously over various locations of the scalp.

The cross-correlation function, r_{xy} , between the two stationary random signals $x(t)$ and $y(t)$ was defined in Equation 3.43. Let us define the *cross-spectral density function* (or *cross spectrum*) as the FT of r_{xy} :

$$S_{xy}(w) = F\{r_{xy}(\tau)\} \quad (6.36)$$

The cross-correlation function is not necessarily even, hence the cross-spectral density is, in general, a complex function:

$$S_{xy}(w) = |S_{xy}(w)| \exp(-j\theta_{xy}(w)) \quad (6.37)$$

It is easily shown that the cross spectrum is bounded by:

$$|S_{xy}(w)|^2 \leq S_x(w)S_y(w) \quad (6.38)$$

An important application of cross spectra is related to the analysis of linear systems. Consider a linear system with impulse response $h(t)$, stationary input signal $u(t)$, and output $x(t)$. It can be shown that the cross spectrum of input and output signals is

$$S_{ux}(w) = H(w)S_u(w) \quad (6.39)$$

Hence, the frequency response of the system can be calculated from the cross spectrum and input spectrum.

A convenient real value bounded quantity is defined, named the *coherence function*:

$$\gamma_{xy}^2(w) = \frac{|S_{xy}(w)|^2}{S_x(w)S_y(w)} \leq 1 \quad (6.40)$$

When $\gamma_{xy}^2(w) = 1$ for all frequencies, $x(t)$ and $y(t)$ are said to be fully coherent; when for some $w = w_o$, $\gamma_{xy}^2(w_o) = 0$, $x(t)$ and $y(t)$ are said to be incoherent at w_o . When $x(t)$ and $y(t)$ are statistically independent, then $\gamma_{xy}^2(w) = 0$ for all w .

Coherence function is useful in investigation of signals which are only slightly correlated.

Hence, it is the low coherence values that are of interest. In practice, the exact values of the various spectra are not known and must be estimated (see Chapter 8); hence, the coherence function is always given in terms of its estimates. Common and novel^{10,11} estimation methods exist for the coherence function. Estimation may cause large inaccuracies in the coherence function, and its application must be carefully considered.⁹

The coherence function has been applied to EEG analysis^{9,12,13} for the investigation of brain asymmetry, localizing epileptic focus, the study of relations between cortical and thalamic activity, and more.

III. LINEAR FILTERING

A. Introduction

In the design of signal acquisition and processing systems we must often alter a given signal such that some parts of it are enhanced, or attenuated, its phase is changed, parts of it are delayed, smoothed, or "predicted". The signal may be deterministic, random, continuous, or discrete. Many of the desired alterations can be achieved by linear transformation. We then design a linear system, or filter, that operates on the signal with the required transformation.

The basic filter is the *time invariant* filter, or *fixed parameters* filter. It is usually designed to meet the required specifications, given some *a priori* information concerning the signals and noise involved. Filters can be designed to meet the required specifications while optimizing some performance criterion; these are called *optimal filters*. One example, the Wiener filter, is discussed in this chapter. Filters in which the values of the parameters are functions of time are called *time varying filters*. An important class of time varying filters is the class of *adaptive filters* which is discussed in Chapter 9.

Consider the signal $u(t)$ which is to be processed. It is desirable to apply a linear transformation such that its outcome will be $x(t)$. We can consider the linear system depicted in Figure 2 to be the filter driven by the input signal $u(t)$ with the output signal $x(t)$.

The relations between the two signals are generally given in terms of the differential equation:

$$\frac{d^n x(t)}{dt^n} + a_{n-1} \frac{d^{n-1}x(t)}{dt^{n-1}} + \dots + a_1 \frac{dx(t)}{dt} + a_0 x(t) = b_m \frac{d^m u(t)}{dt^m} + \dots + b_1 \frac{du(t)}{dt} + b_0 u(t) \quad (6.41)$$

A general solution to Equation 6.41 is given in terms of the *impulse response*, $h(t)$. Since the system is linear, its output is composed of the linear combination of the response to the impulse function.¹⁴ It can be shown that the output, $x(t)$, is given by the convolution of the impulse response with the input:

$$x(t) = h(t) * u(t) \quad (6.42A)$$

or when taking the FT:

$$X(w) = H(w) \cdot U(w) \quad (6.42B)$$

We see that $U(w)$ can be shaped into a desired $X(w)$ by designing the right filter $H(w)$. The advantages of the design in the frequency domain become obvious from Equation 6.42B.

Only the basics of digital and optimal filters design will be discussed here. For detailed discussion of the material, the reader is referred to the literature of these topics.¹⁻⁸ The topic

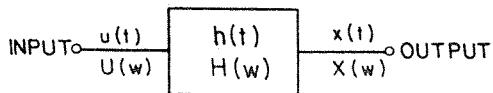


FIGURE 2. Linear system in the time and frequency domains.

of cepstral analysis and homomorphic filtering with applications to biomedical signal processing are discussed in detail at the end of this section.

B. Digital Filters

The availability of low cost and efficient digital computers and dedicated processing circuits have made the implementation of filtering, by digital means, very attractive. Even when dealing with analog environments, where both input and output signals are continuous, it is very often worthwhile to apply analog-to-digital conversion, perform the required filtering digitally, and convert the discrete filtered output back into a continuous signal.

Digital filters are linear discrete systems governed by difference equations (see Chapter 4). Two classes of digital filters are used — *finite impulse response* (FIR) and *infinite impulse response* (IIR).

FIR filters are characterized by finite duration impulse response which, in the Z domain means:

$$H(Z) = \frac{X(Z)}{U(Z)} = b_0 + b_1 Z^{-1} + \dots + b_m Z^{-m} \quad (6.43)$$

where $X(Z)$ and $U(Z)$ are the Z transforms of the input and output sequences. Equation 6.43 states that the FIR filter is a moving average (MA) filter (see Chapter 7), or an all zero filter. FIR filters are always stable.

IIR filters have, in general, infinite duration impulse response, they possess zeroes and poles (ARMA filters — see Chapter 7), and their transfer function in the Z domain is

$$H(Z) = \frac{b_0 + b_1 Z^{-1} + \dots + b_m Z^{-m}}{1 + a_1 Z^{-1} + \dots + a_n Z^{-n}} \quad (6.44)$$

IIR filters are stable if all the poles of $H(Z)$ are within the unit circle in the Z domain.

IIR and FIR filters can be synthesized recursively via the difference equations, or by means of the FFT. Since continuous filter design is well established, one of the approaches for designing digital filters is to find a difference equation, with the associated $H(Z)$, that yields an output sequence close to the samples of the analog output signal. This approach is termed the *impulse invariant method*. Another approach is to transform the analog filter, by means of the bilinear transformation, into the Z domain yielding a digital filter, $H(Z)$. The resultant filter will not possess the same impulse response since the transformation introduces frequency scale distortions. This method is known as the *bilinear transformation method*. A third approach for digital filter design is the *frequency sampling method*. This method is based on the approximation of a function by a sum of sinc functions. Detailed discussion of the design steps can be found for example in Gold and Rader⁴ and Chen.⁸

C. The Wiener Filter

Consider now the problem of optimal filter design. Assume, for example, that a signal $s(t)$ is corrupted with additive noise $n(t)$; it is required to estimate, by linear operations, the value of the signal $s(t + \eta)$, $\eta \geq 0$ from the observation signal $x(t)$:

$$x(t) = s(t) + n(t) \quad (6.45)$$

We assume that both $s(t)$ and $n(t)$ are stationary in the wide sense. Note that for $\eta = 0$ the problem is that of *smoothing*, namely, extracting the current value of $s(t)$ from current and past values of the observation signal. For $\eta > 0$ the problem is that of *prediction*, namely, extracting the future value of $s(t + \eta)$ from current and past values of the observation signal.

Assume we have a linear filter, $h(t)$. We apply the signal $x(t)$ to its input. Let us denote the output of the filter by $\hat{s}(t + \eta)$. We shall look for the optimal filter in the sense of minimization of the mean square error between the output of the filter and the actual desired quantity:

$$\begin{aligned} E\{\epsilon^2(t)\} &= E\{[s(t + \eta) - \hat{s}(t + \eta)]^2\} = \\ &E\{[s(t + \eta) - \int_{-\infty}^{\infty} h(\tau)x(t - \tau)d\tau]^2\} \end{aligned} \quad (6.46)$$

It is required to minimize ϵ^2 over all possible, realizable $h(t)$. Performing the minimization⁴ yields the condition:

$$r_{sx}(\tau + \eta) = \int_0^{\infty} h(\xi)r_x(\tau - \xi)d\xi; \quad \tau \geq 0 \quad (6.47)$$

where r_{sx} and r_x are the cross correlation of the observation signal with the desired signal $s(t)$ and the autocorrelation of $x(t)$, respectively. Condition 6.47 is known as the *Wiener Hopf condition* (see also Equation 9.10).

When the optimal filter, given by Condition 6.47, is used the minimum squared error is⁶

$$E\{\epsilon^2(t)\} = r_s(0) - \int r_{sx}(\tau)h(\tau)d\tau \quad (6.48)$$

and

$$E\{\epsilon(t)x(t - \tau)\} = r_{\epsilon x}(\tau) = 0 \quad (6.49)$$

The last result states that under optimal conditions, the error and observations are uncorrelated; since $E\{\epsilon(t)\} = 0$, the two are also orthogonal. If we remove the realizability constraint, the solution of Equation 6.46 will be similar to Equation 6.47 but with a lower integration boundary including all negative values; namely, integration boundaries will be minus to plus infinity. Note that the right-hand side of Equation 6.47 is the convolution of $r_x(\tau)$ with $h(\tau)$. Taking the FT of the equation yields:

$$S_{sx}(w)\exp(jw\eta) = H(w)S_x(w) \quad (6.50)$$

The exponent in the left side of the last equation is due to the time shift present in r_{sx} of Equation 6.47. The required optimal filter is thus given in the frequency domain by:

$$H(w) = \frac{S_{sx}(w)}{S_x(w)} \exp(jw\eta) \quad (6.51)$$

We have assumed $x(t)$ and $s(t)$ to be stationary; therefore, the filter $h(t)$ given by Equation 6.51 is a stable filter. However, the PSD function $S_x(w)$ is an even function of w , hence the filter will have poles in the right half plane (RHP) of the complex S domain. A stable

system with poles at the RHP must have nonzero impulse response, $h(t)$, at $t < 0$ which means a nonrealizable system.

An optimal realizable Wiener filter can be calculated⁶ from Equation 6.47. Its error will be larger than or equal to that of the optimal filter. Similar arguments can be applied to digital filters. Optimal Wiener FIR and IIR filters can be designed.⁸

IV. CEPSTRAL ANALYSIS AND HOMOMORPHIC FILTERING

A. Introduction

The concept of cepstrum was first introduced in the early sixties in an attempt to analyze a signal containing echoes. The *power cepstrum* was defined as "the power spectrum of the logarithm of the power spectrum". Later the definition was changed to make its connection with the correlation function clearer and to provide it with units of time. The new definition became "the inverse transform of the log of the power spectrum". The term "cepstrum" was derived from the "spectrum" by reversing the order of the first four letters. The domain of the cepstrum was termed *quefrency*, a term derived from frequency. Additional terms have been defined, such as "lifter" (derived from "filter"), but these were not accepted well in the literature.

Cepstral analysis is applied mainly in cases where the signal contains echoes of some fundamental wavelet. By means of the power cepstrum, the times of the wavelet and the echoes can be determined. The *complex cepstrum* is used to determine the shape of the wavelet. These techniques¹⁵⁻²¹ have been discussed in the literature with various applications. It has been applied to the analysis of EEG signals,^{17,21} to ECG signals,²⁰ and to the speech signal.¹⁹

B. The Cepstra

The complex cepstrum, $\tilde{x}(\tau)$, of the real signal $x(t)$ is given by:

$$\tilde{x}(\tau) = F^{-1}\{\log F\{x(t)\}\} \quad (6.52)$$

Since the argument of the logarithm in Equation 6.52 is complex and may be negative, we shall introduce the complex logarithm of a complex function V :

$$\log(V) = \log|V| + j \arg(V) \quad (6.53)$$

We shall also need to perform the inverse operation, namely exponentiation; therefore, let us define complex exponentiation of V :

$$\exp(V) = \exp(\log|V|) \cdot \exp(j \arg(V)) \quad (6.54)$$

The power cepstrum, $\tilde{x}_p(\tau)$, is defined by:

$$\tilde{x}_p = (F^{-1}\{\log|F\{x(t)\}|^2\})^2 \quad (6.55)$$

In the discrete case, when the data is presented in terms of the sequence $\{x(nT)\}$, the cepstra are defined¹⁶ by means of the Z transform.

The power cepstrum of the sequence $\{x(nT)\}$ is the square of the inverse Z transform of the logarithm of the magnitude squared of $X(Z)$. Thus, we write the power cepstrum $\tilde{x}_p(nT)$:

$$\tilde{x}_p(nT) = (Z^{-1}\{\log|X(Z)|^2\})^2 = \left(\frac{1}{2\pi j} \oint_c \log|X(Z)|^2 Z^{n-1} dZ \right)^2 \quad (6.56)$$

The final squaring in Equations 6.55 and 6.56 is frequently omitted. The power cepstrum does not contain information about the phase. If phase is of interest, we must consider the complex cepstrum of the sequence defined as the inverse Z transform of the complex logarithm of $X(Z)$:

$$\tilde{x}(nT) = Z^{-1}\{\log(X(Z))\} = \frac{1}{2\pi j} \oint_c \log X(Z) Z^{n-1} dZ \quad (6.57)$$

If the sequence $x(nT)$ is the convolution of two sequences $u(nT)$ and $h(nT)$, namely, $x(nT) = u(nT)*h(nt)$, then:

$$X(Z) = U(Z) \cdot H(Z) \quad (6.58A)$$

$$\log X(Z) = \log U(Z) + \log H(Z) \quad (6.58B)$$

and since the inverse transform is a linear operation, the complex cepstrum is

$$\tilde{x}(nT) = \tilde{u}(nT) + \tilde{h}(nT) \quad (6.58C)$$

Hence, the complex cepstrum of the convolution of two sequences equals the sum of their cepstra. The complex cepstrum is thus an operator converting convolution into summation. Its application to deconvolution problems becomes apparent. Assume that $x(nT)$, $u(nT)$, and $h(nT)$ are the output, input, and impulse response sequences of a discrete linear system, respectively. If $\tilde{u}(nT)$ and $\tilde{h}(nT)$ occupy different quefrency ranges, then the complex cepstrum can be filtered (filtered) to remove one. In the complex cepstrum, phase information is retained therefore it can be inverted, to yield the deconvolved $h(nT)$ or $u(nT)$.

The computation of the complex cepstrum in Equation 6.57 has to be carefully considered since the complex logarithm is not single valued. The imaginary part of the complex logarithm (Equation 6.53) is the phase. If it is presented in module 2π form (principal value), then discontinuities will appear in the phase term. This will occur due to the jump from 2π to zero, when the phase is being increased over 2π . Phase unwarping algorithms must be employed to overcome this problem. A simple solution is to compute the relative phase between adjacent samples, add them together in order to get a cumulative, unwarped phase.

The complex cepstrum can be implemented²² by means of the DFT replacing the Z transform. This is true since the sequences are of finite length. The region of convergence for the Z transform includes the unit circle allowing the Z transform and its inverse to be evaluated for $Z = \exp(jw)$; therefore:

$$\tilde{x}(nT) = IDFT\{\log(DFT\{x(nT)\})\} \quad (6.59)$$

Equation 6.59 is of great computational importance since the DFT and IDFT can be very effectively calculated by the FFT algorithm.

The upper part of Figure 3 depicts schematically the operations involved in the complex cepstrum computations.

C. Homomorphic Filtering

Let us consider again the example given by Equations 6.58A — C. Here the sequence $\{x(nT)\}$ can be the samples of a speech signal, the sequence $\{h(nT)\}$ the weighting sequence of the vocal tract, and $\{u(nT)\}$ the samples of the pressure wave exciting the vocal tract during voiced utterance, when the vocal cords are vibrating. The pressure $\{u(nT)\}$ can be

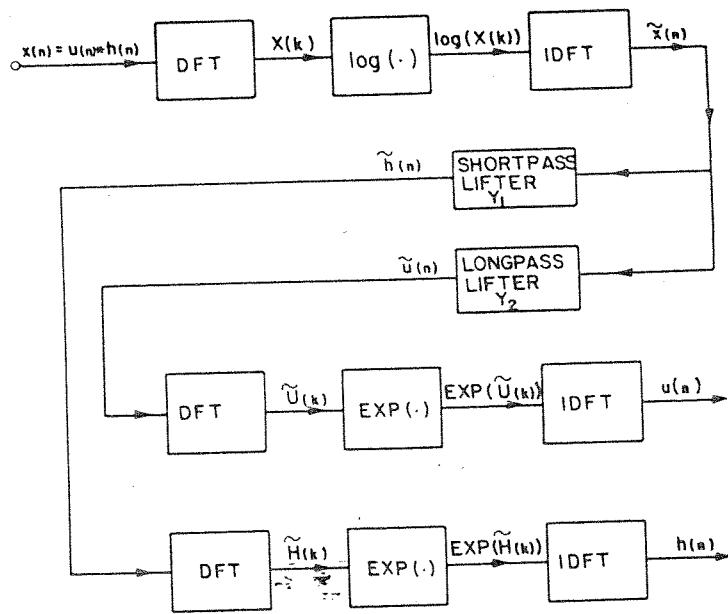


FIGURE 3. Homomorphic filtering.

modeled as a train of very narrow pulses appearing at a frequency known as the fundamental frequency or the pitch. We are interested both in the sequence $\{h(nT)\}$ in order to learn about the vocal tract characteristics, and in the sequence $\{u(nT)\}$ in order to estimate the pitch.

Equation (6.58C) gives the complex cepstrum as the sum of the cepstra of the input and the vocal tract responses. Assume that in the quefrency range we have:

$$\tilde{h}(nT) = 0 \quad \text{for } n \geq n_0 \quad (6.60A)$$

and

$$\tilde{u}(nT) = 0 \quad \text{for } n < n_0 \quad (6.60B)$$

Therefore, these are separable in the quefrency domain. Consider two lifters, a *short pass lifter*, $Y_1(nT)$, given by:

$$Y_1(nT) = \begin{cases} 1; & n < n_0 \\ 0; & \text{otherwise} \end{cases} \quad (6.61A)$$

and a *long pass lifter*, $Y_2(nT)$, given by:

$$Y_2(nT) = \begin{cases} 0; & \text{otherwise} \\ 1; & n \geq n_0 \end{cases} \quad (6.61B)$$

When $\tilde{x}(nT)$ is fed into the input of these two filters, the output of Y_1 will be $\tilde{h}(nT)$ and that of Y_2 will be $\tilde{u}(nT)$. We now want to transfer $\tilde{u}(nT)$ and $\tilde{h}(nT)$ from the quefrency back into the time domain. We have to subject the sequences to the inverse operation. This involves first the DFT followed by complex exponentiation (Equation 6.54) and IDFT. The complete operation of the *homomorphic filtering* is depicted in Figure 3.

Homomorphic filtering has been applied²⁰ to the automatic classification of ECG. Normal inverted T-wave and two types of premature ventricular contractions (PVC) have been considered. It has been found that feature selection for diagnostic purposes could be more efficient using homomorphic filtering than by conventional methods. It has been also demonstrated that the basic wavelet of normal ECG signal evaluated by the homomorphic filtering closely approximates the action potential spike in the cardiac muscle fibers.

Senmoto and Childers²¹ have used homomorphic filtering to decompose visual evoked response (VER) potentials. It has been suggested that the recorded VER signals can be expressed as an aggregate of overlapping signals generated by multiple spatially disparate sources whose basic signal waveforms are unknown and have to be estimated. The assumption, therefore, is that the wavelets are identical in waveshape. We shall consider here the decomposition of two wavelets. The extension to the multiple case can be easily done.

Let $x(t)$ be the composite signal and $s(t)$ the wavelet; then:

$$x(nT) = s(nT) + a\delta((n - n_0)T) \quad (6.62)$$

where the shape of $s(t)$, the delay n_0 , and the echo amplitude $a < 1$ are unknown. $x(nT)$ can be written in terms of the convolution:

$$x(nT) = s(nT) * p(nT) \quad (6.63A)$$

where

$$p(nT) = \delta(nT) + a\delta((n - n_0)T) \quad (6.63B)$$

Taking the Z transform:

$$X(Z) = S(Z) \cdot P(Z) \quad (6.64A)$$

with

$$P(Z) = 1 + aZ^{-n_0} \quad (6.64B)$$

Taking the complex logarithm:

$$\log(X(Z)) = \log(S(Z)) + \log(1 + aZ^{-n_0}) \quad (6.65)$$

The second term in the right side of Equation 6.65 can be expanded in a power series yielding

$$\log(X(Z)) = \log(S(Z)) + aZ^{-n_0} - \frac{a^2}{2} Z^{-2n_0} + \frac{a^3}{3} Z^{-3n_0} \dots \quad (6.66)$$

The complex cepstrum, $\tilde{x}(nT)$, is given by the inverse Z transform:

$$\begin{aligned}\hat{x}(nT) = Z^{-1}\{\log(X(Z))\} &= \hat{s}(nT) + a\delta(nT + n_0T) - \frac{a^2}{2}\delta(nT - 2n_0T) + \\ &\quad \frac{a^3}{3}\delta(nT - 3n_0T) \dots \dots \end{aligned}\tag{6.67}$$

Thus, the complex cepstrum of the composite signal consists of the complex cepstrum of the wavelet plus a train of δ functions located at positive quefrequencies at the echo delay and its multiples. A comb notch lifter can be used to remove the train of delta function. After smoothing, the wavelet is reconstructed by inverting the operations used for the computation of the complex cepstrum, as shown in Figure 3.

A similar procedure can be used for the processing of dye dilution curves (see Appendix A, Volume II).

REFERENCES

1. Bracewell, R. N., *The Fourier Transform and Its Applications*, McGraw-Hill, Kogakusha, Tokyo, 1978.
2. Papoulis, A., *Signal Analysis*, McGraw-Hill Int., Auckland, 1977.
3. Tretter, S. A., *Introduction to Discrete Time Signal Processing*, John Wiley & Sons, New York, 1976.
4. Gold, B. and Rader, C. M., *Digital Processing of Signals*, McGraw-Hill, New York, 1969.
5. Brigham, E. O., *The Fast Fourier Transform*, Prentice-Hall, Englewood Cliffs, N.J., 1974.
6. Lathi, B. P., *An Introduction to Random Signals and Communication Theory*, Int. Textbook Co., Scranton, Pa., 1968.
7. Davenport, W. B. and Root, W. L., *An Introduction to the Theory of Random Signals and Noise*, McGraw-Hill, New York, 1958.
8. Chen, C. T., *One Dimensional Digital Signal Processing*, Marcel Dekker, New York, 1979.
9. Glasser, E. M. and Ruchkin, D. S., *Principles of Neurobiological Signal Analysis*, Academic Press, New York, 1976.
10. Nuttal, A. H., Direct coherence estimation via a constrained least-squares linear predictive fast algorithm, *Proc. ICASSP*, IEEE, Paris, 1982, 1104.
11. Youn, D. H., Ahmed, N., and Carter, G. C., Magnitude squared coherence function estimation: an adaptive approach, *IEEE Trans. Acoust. Speech Signal Process.*, 31, 137, 1983.
12. Shaw, J. C., Brooks, S., Colter, N., and O'Connor, K. P., A comparison of schizophrenic and neurotic patients using EEG power and coherence spectra, in *Hemisphere Asymmetries of Function in Psychopathology*, Gruzelier, J. and Flor-Henry, P., Eds., Elsevier-North Holland, Amsterdam, 1979.
13. Beaumont, J. G., Mayes, A. R., and Rugg, M. D., Asymmetry in EEG alpha coherence and power: effect of task and sex, *Electroencephalogr. Clin. Neurophysiol.*, 45, 393, 1978.
14. Derusso, P. M., Roy, R. Y., and Close, C. M., *State Variables for Engineers*, John Wiley & Sons, New York, 1967.
15. Oppenheim, A. V., Generalized linear filtering, in *Digital Processing of Signals*, Gold, B. and Rader, C. M., Eds., McGraw-Hill, New York, 1969.
16. Childers, D. G., Skinner, D. P., and Kemerait, R. C., The cepstrum: a guide to processing, *Proc. IEEE*, 65, 1428, 1977.
17. Kemerait, R. C. and Childers, D. G., Signal detection and extraction by cepstrum techniques, *IEEE Trans. Inf. Theory*, 18, 745, 1972.
18. Oppenheim, A. V., Kopec, G. E., and Tribolet, J. M., Signal analysis by homomorphic prediction, *IEEE Trans. Acoust. Speech Signal Process.*, 24, 327, 1976.
19. Kopec, G. E., Oppenheim, A. V., and Tribolet, J. M., Speech analysis by homomorphic prediction, *IEEE Trans. Acoust. Speech Signal Process.*, 25, 40, 1977.
20. Murthy, I. S. N., Rangaraj, M. R., Udupa, K. J., and Goyal, A. K., Homomorphic analysis and modeling of ECG signals, *IEEE Trans. Biomed. Eng.*, 26, 330, 1979.
21. Senmoto, S. and Childers, D. G., Adaptive decomposition of a composite signal of identical unknown wavelets in noise, *IEEE Trans. Syst. Man Cybern.*, 2, 59, 1972.
22. Oppenheim, A. V. and Schafer, R. W., *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1975.

Chapter 7

TIME SERIES ANALYSIS-LINEAR PREDICTION

I. INTRODUCTION

Modern signal processing techniques are applied to a variety of fields such as econometrics, speech, seismology, communications, and biomedicine. A major problem in these applications is the need to analyze and process finite time samples of random processes. In general, the processes are nonstationary² and nonlinear. The theoretical basis for modern time series analysis has been developed by mathematicians and statisticians such as Mann and Wald.¹

Recent developments in both theory and computational algorithms of linear stationary signal analysis³⁻¹⁰ provide powerful tools for signal processing. Some of the techniques are well established (with available computer program packages (Reference 11, for example)). When a nonstationary signal is to be processed, it is usually regarded in such a way that each segment can be considered stationary. Stationary signal processing methods then can be applied.

A favorable approach for stationary signal processing is the parametric modeling procedure. The process is modeled by some causal rational parametric model. The signal is then represented by means of the model parameters. Such a procedure is attractive from the point of view of data compression. Rather than handling (for processing, storing, or transmitting) the complete time sample, or sequence, only a reduced number of parameters are used. Consider, for example, the problem of analyzing and storing EEG data¹² in neurological clinics. It would be of great help if these data could be reduced and compressed for storage purposes in such a way that the signal can be regenerated at will. Another example may be the storing of compressed ECG data (or a complete medical file) on a personal credit card carried by the patient in such a way that it can be reproduced at will anywhere.

Signal compression is also attractive from the point of view of classification (diagnosis). Effective algorithms for the automatic classification of signals typically representing various pathological states are available.

Since most modern signal processing is implemented by digital computers, we consider the sampled signal $S^*(t)$ sampled at the frequency of $f_s = 1/T$

$$S^*(t) = \sum_{n=-\infty}^{\infty} S(t) \cdot \delta(t - nT) \quad (7.1)$$

The finite time windowed sampled signal is given by the sequence $\{S(nT)\}$ $n = 0, 1, \dots, N - 1$

$$\{S(nT)\} = S(0), S(T), S(2T), \dots, S((N - 1)T) \quad (7.2)$$

For the sake of brevity, we shall denote the sequence by $\{S_n\}$ without the loss of generality. The sequence in Equation 7.2 is to be modeled by a parametric model.

A very effective parametric model is that of the transfer function (TF). The sampled signal (the sequence), $\{S_n\}$, is assumed to be the output of a linear system driven by an (inaccessible) input sequence $\{U_n\}$ and corrupted by an additive noise. The sequence $\{S_n\}$ is thus given as the solution to the difference equation

$$S_n = -\sum_{i=1}^p a_i S_{n-i} + \sum_{i=0}^q b_i U_{n-i} + \sum_{i=0}^r a_i \xi_{n-i} \quad (7.3)$$

in Equation 7.1; $\{\xi_k\}$ is the additive noise sequence. It is usually convenient to work with noise sequences which are white, with zero mean. Consider, therefore, the sequence $\{\xi_k\}$ to be the output of a noise filter driven by white noise n_k such that

$$\sum_{i=0}^n c_i \xi_{k-i} = \sum_{i=0}^m d_i n_{k-i} \quad (7.4)$$

Defining the operators

$$\begin{aligned} A(z^{-1}) &= \sum_{i=0}^p a_i z^{-i}; a_0 = 1 \\ B(z^{-1}) &= \sum_{i=0}^q b_i z^{-i} \\ C(z^{-1}) &= \sum_{i=0}^n c_i z^{-i} \\ D(z^{-1}) &= \sum_{i=0}^m d_i z^{-i} \end{aligned} \quad (7.5)$$

and transferring Equations 7.3 and 7.4 into the z domain, we get

$$S(z) = \frac{B(z^{-1})}{A(z^{-1})} U(z) + \frac{D(z^{-1})}{C(z^{-1})} N(z); \quad (\text{TF}) \quad (7.6)$$

In Equation 7.6 the sequence $\{S(k)\}$ is modeled by means of the system parameter vector, $\underline{\beta}_s$, and the noise parameter vector, $\underline{\beta}_n$, where

$$\begin{aligned} \underline{\beta}_s^T &= [a_0, a_1, \dots, a_p, b_0, b_1, \dots, b_q] \\ \underline{\beta}_n^T &= [c_0, c_1, \dots, c_n, d_0, d_1, \dots, d_m] \end{aligned} \quad (7.7)$$

The problem of identifying the above parameters when the input is available is well covered in the literature on system identification.¹³ In signal processing modeling the input sequence $\{U_k\}$ is assumed to be a white unaccessible sequence. The parameter vector $\underline{\beta}_s$ is thus describing a linear transformation transferring the white sequence into the (colored) signal sequence. The transfer function model can be decoupled into the deterministic and noise models

$$A(z^{-1})Y(z) = B(z^{-1})U(z); \quad \text{Deterministic system}$$

$$C(z^{-1})\xi(z) = D(z^{-1})N(z); \quad \text{Noise model}$$

$$S(z) = Y(z) + \xi(z); \quad \text{Observation equation}$$

Here the sequence $\{Y_k\}$ is the unaccessible noise-free output of the system.

Figure 1 shows schematically the transfer function model of the sequence $\{S_k\}$. Several time series models have been derived from the TF model of Equation 7.6.

The Autoregressive moving average exogenous variables (ARMAX) model is derived from Equation 7.6 by letting

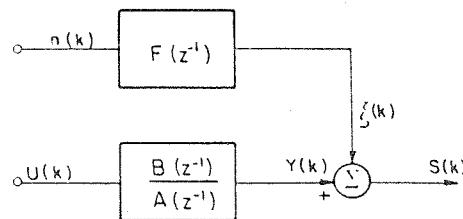


FIGURE 1. Transfer function model.

$$C(z^{-1}) = A(z^{-1}) \quad (7.8)$$

(see, for example, Reference 14). The ARMAX model can thus be written as

$$A(z^{-1})S(z) = B(z^{-1})U(z) + D(z^{-1})N(z); \quad (\text{ARMAX}) \quad (7.9)$$

or in terms of the difference equation

$$S_k = - \sum_{i=1}^p a_i S_{k-i} + \sum_{i=0}^q b_i U_{k-i} + \sum_{i=0}^m d_i n_{k-i} \quad (7.10)$$

The autoregressive moving average (ARMA) model is derived from Equation 7.6 by assuming there is no external noise, hence

$$S(z) = \frac{B(z^{-1})}{A(z^{-1})} U(z); \quad (\text{ARMA}) \quad (7.11)$$

or in terms of the difference equation

$$S_k = - \sum_{i=1}^p a_i S_{k-i} + \sum_{i=0}^q b_i U_{k-i} \quad (7.12)$$

The last model is known as ARMA of order (p,q); in engineering technology, it is known as a pole-zero model.¹⁵⁻¹⁶

In Equation 7.10, the current (k^{th}) sample of the sequence is expressed as linear combination of the past p sequence samples and $(q + 1)$ input samples. Hence, the models discussed here are also known as linear prediction models. Figure 2 depicts the ARMA model in the z and in the time domains.

Two additional models are commonly used: the autoregressive (AR) model and the moving average (MA) model. The AR model¹⁷ is an all pole model given by assuming $B(z^{-1}) = G$

$$S(z) = \frac{G}{A(z^{-1})} U(z); \quad (\text{AR}) \quad (7.13)$$

$$S_k = - \sum_{i=1}^p a_i S_{k-i} + GU(k) \quad (7.14)$$

while the MA, all zero model, is derived from the ARMA model by assuming $A(z^{-1}) = 1$

$$S(z) = B(z^{-1})U(z); \quad (\text{MA}) \quad (7.15)$$

$$S_k = \sum_{i=0}^q b_i U_{k-i} \quad (7.16)$$

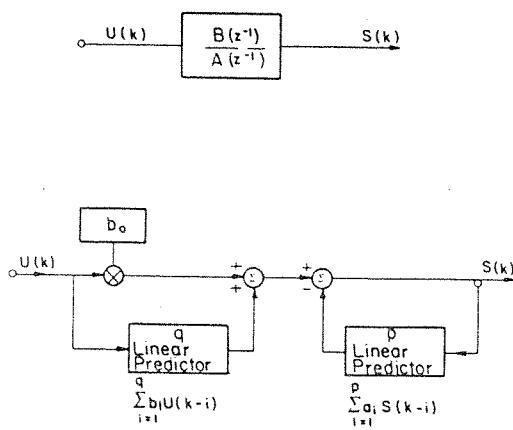


FIGURE 2. Autoregressive moving average (ARMA) model.

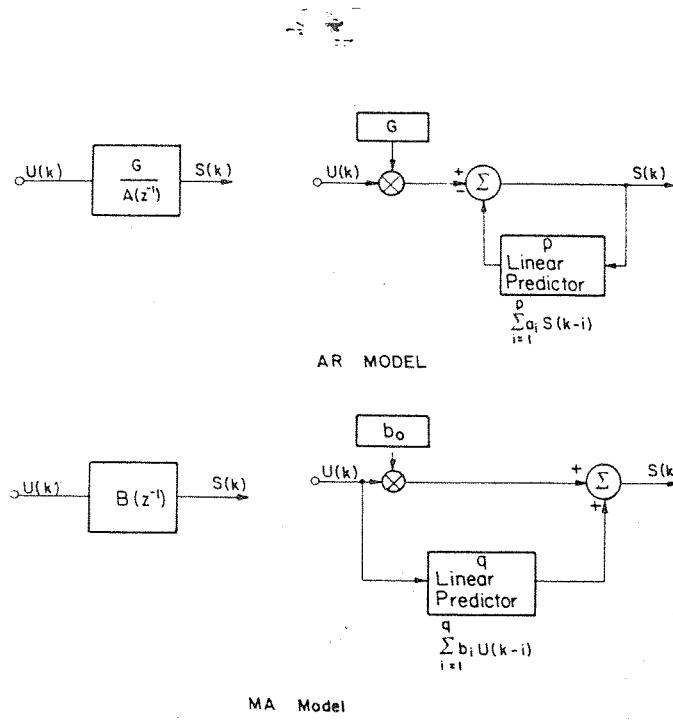


FIGURE 3. Autoregressive (AR) and moving average (MA) models.

$$S_k = \sum_{i=0}^q b_i U_{k-i} \quad (7.16)$$

Figure 3 shows, schematically, the AR and MA models in the frequency and time domains.

Other models such as the IMA or ARIMA³ (autoregressive integrated moving average) are used when homogeneous nonstationary signals are to be modeled (see Section VIII).

The basic idea behind linear predictive modeling is that by assuming the sequence $\{S_k\}$ to be the output of a linear system, one can express the sequence (in a reduced parametric manner) by means of the system parameters. Given a sequence $\{S_k\}$, one has to identify the parameters of the system. Since the input is inaccessible, well-known algorithms for system

identification¹³ cannot be directly applied. The common approach has been to model the sequence by means of a system driven by an input with white spectrum (i.e., white noise, impulse). Algorithms for the estimation of system parameters (without the need to have an access to the input) are available. These will be discussed in the following sections.

The AR and ARMA models are the most commonly used. These will be discussed in more detail in following sections.

II. AUTOREGRESSIVE (AR) MODELS

A. Introduction

The AR model¹⁷ is very often used because of its simplicity and because of the fact that effective algorithms for the estimation of the AR parameters are available. Note that an ARMA model can be approximated by means of an AR model. Assume an ARMA model given by the polynomials $\tilde{B}(z^{-1})$ and $\tilde{A}(z^{-1})$, of order \tilde{q} and \tilde{p} . By long division we can get

$$\frac{\tilde{B}(z^{-1})}{\tilde{A}(z^{-1})} = \frac{1}{A(z^{-1})} \quad (7.17)$$

hence, the ARMA model can be expressed by means of an AR model. In general, however, the polynomial $A(z^{-1})$ will be of infinite dimension. It is possible to approximate the ARMA model by an AR model having the first p coefficients of the $A(z^{-1})$ polynomial.

The choice of the order p of the AR model depends on the accuracy required.

When modeling a stationary process, one must make sure that the model is stationary. In order to ensure weak stationarity, the autocovariance and autocorrelations of the modeled sequence must satisfy a set of conditions.³ For a linear process, these are ensured if the complex roots of the characteristic equation

$$A(z^{-1}) = 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_p z^{-p} = 0 \quad (7.18)$$

are all inside the unit circle in the z plane or outside the unit circle in the z^{-1} plane.

The inverse filter $H^{-1}(z^{-1})$

$$H^{-1}(z^{-1}) = \left(\frac{1}{A(z^{-1})} \right)^{-1} = A(z^{-1}) \quad (7.19)$$

is known as the AR whitening filter. When the sequence $\{S_k\}$ serves as the input to the AR whitening filter, the resultant output will have white spectrum. The simplest AR process is that of the first order. It is known as the Markov process, given by the difference equation

$$S_k = -a_1 S_{k-1} + GU_k \quad (7.20)$$

which is stationary for all $|a_1| < 1$.

B. Estimation of AR Parameters — Least Squares Method

It is required to estimate the order of the process, p , its coefficients a_i , $i = 1, 2, \dots, p$, and the gain factor, G . Since the input is inaccessible, the estimation must be performed without the sequence $\{U_k\}$.

Assume we have at our disposal the samples S_j , $j = 0, 1, \dots, (k - 1)$. We can estimate the incoming sample S_k by the estimator

$$\hat{S}_k = - \sum_{i=1}^{\hat{p}} \hat{a}_i S_{k-i} \quad (7.21)$$

where a circumflex ($\hat{\cdot}$) denotes estimated value. For the time being, we shall assume that \hat{p} is given (for example, by guessing). At time $t = kT$ we can calculate the error $e(k)$ (known as the "residual") between the actual sequence sample and the predicted one:

$$e_k = S_k - \hat{S}_k = S_k + \sum_{i=1}^{\hat{p}} \hat{a}_i S_{k-i} \quad (7.22)$$

Note that the residuals $\{e_k\}$ are the estimates of the inaccessible input $\{G_{U_k}\}$. The least squares method determines the estimated parameters by minimizing the expectation of the squared error.

$$\min_{\hat{a}_i} E\{e_k^2\} = \min_{\hat{a}_i} E\{(S_k + \sum_{i=1}^{\hat{p}} \hat{a}_i S_{k-i})^2\} \quad (7.23)$$

Performing the minimization by

$$\frac{\partial E\{e_k^2\}}{\partial \hat{a}_i}; \quad i = 1, 2, \dots, \hat{p} \quad (7.24)$$

and assuming the sequence to be stationary, we get \hat{p} linear equations

$$\sum_{j=1}^{\hat{p}} \hat{a}_j r_{i-j} = -r_i \quad i = 1, 2, \dots, \hat{p} \quad (7.25)$$

where $r_{i-j} = E\{S_{k-j} S_{k-i}\} = r_{j-i}$.

Equations 7.25 are known as the Yule-Walker equations or the normal equations. They can be solved for the least squares optimal parameters $\hat{a}_j; j = 1, 2, \dots, \hat{p}$, if the $\hat{p} + 1$ correlations $r_j; j = 0, 1, \dots, \hat{p}$ are given.

It can also be shown,¹⁷ using the optimal parameters, that the minimum average error denoted by E_p

$$E_p = r_o + \sum_{j=1}^{\hat{p}} a_j r_j \quad (7.26)$$

The correlation coefficients are not given; hence, they have to be estimated from the given finite sequence $\{S(k)\}$. Assume the sequence $\{S_k\}$ is given. For $k = 0, 1, 2, \dots, (N - 1)$ we can estimate the correlation coefficients by

$$\hat{r}_i = \frac{1}{N} \sum_{k=0}^{N-1-i} S_k S_{k+i} \quad (7.27)$$

In Equation 7.27, we have assumed all samples of $\{S_k\}$ to be zero outside the given range. These estimations (Equation 7.27), known as the autocorrelation method, will be used instead of the correlation coefficients of Equation 7.25. For sake of convenience, we shall continue to use the symbol r_i where indeed \hat{r}_i must be used. Equation 7.23 can be written in a matrix form

or

wh
Eo
foT
go
ma
recuEqu
Tl
7.30LHence
 \hat{p} . A
At
order

One

III. MOVING AVERAGE (MA) MODELS

A. Autocorrelation Function of MA Process

Consider the difference equation for the MA process given in Equation 7.16. The autocorrelation, r_i^s , of the MA sequence is calculated from Equation 7.16 by

$$r_i^s = E\{s_k s_{k+i}\} = \sum_{j=0}^q \sum_{\ell=0}^q b_j b_\ell r_{i+j-\ell}^u \quad (7.36)$$

where r_i^u is the autocorrelation of the white input where

$$r_i^u = \delta(i) \quad (7.37)$$

Hence, terms in Equation 7.36 will be nonzero only for

$$i + \ell - j = 0 \quad (7.38A)$$

$$r_i^s = \sum_{j=0}^{q-i} b_j b_{j+i}; \quad 0 \leq i \leq q \quad (7.38B)$$

Equations 7.38A and B are $q + 1$ nonlinear equations. Assuming the correlations r_i^s ; $i = 1, 2, \dots, q$ are given (or are estimated), Equations 7.38A and B can be solved for the unknown MA coefficients.

B. Iterative Estimate of the MA Parameters

Consider the estimation of the correlation coefficients of the MA process, for example, by

$$\hat{r}_i^s = \frac{1}{M-i} \sum_{k=0}^{M-i} s_k s_{k+i} \quad (7.39)$$

Using Equation 7.39 in Equation 7.38 yields

$$\hat{r}_i^s = \sum_{j=0}^{q-i} \hat{b}_j \hat{b}_{j+i}; \quad 0 \leq i \leq q \quad (7.40)$$

These can be written as

$$\begin{aligned} \hat{b}_0^2 &= \hat{r}_0^s - \sum_{j=1}^q (\hat{b}_j)^2 \\ \hat{b}_i &= \hat{b}_0^{-1} (\hat{r}_i^s - \sum_{j=1}^{q-i} \hat{b}_j \hat{b}_{j+i}); \quad i = 1, 2, \dots, q \end{aligned} \quad (7.41)$$

Equations 7.41 can be solved iteratively by

$$\begin{aligned} b_0^{(m)} &= (\hat{r}_0^s - \sum_{j=1}^q b_j^{(m-1)})^{1/2} \\ m &= 0, 1, 2 \end{aligned}$$

$$b_i^{(m)} = (b_o^{(m)})^{-1} (\hat{r}_i^* - \sum_{j=0}^{q-1} b_j^{(m-1)} b_{j+1}^{(m-1)}) \quad (7.42)$$

where $(\cdot)^{(m)}$ is the value of the m th iteration of (\cdot) and $(\cdot)^{(m)} = 0$ for $m < 0$.

Equations 7.42 are iteratively solved until some convergence criterion is satisfied. Such a criterion may be

$$\sum_{j=0}^q (b_j^{(m)} - b_j^{(m-1)})^2 \leq \epsilon \quad (7.43)$$

where ϵ is a predetermined convergence limit. Equations 7.42 converge linearly. Better procedures which converge quadratically, as well as methods based on maximum likelihood (ML) estimations, have been suggested.³ The estimate of Equation 7.42 is recommended only as a rough estimate of the MA parameters, while the more elaborate ML is recommended for a more precise model identification.

IV. MIXED AUTOREGRESSIVE MOVING AVERAGE (ARMA) MODELS

A. Introduction

In cases where the process does have a number of influential zeroes, the approximation 7.17 will require very large dimension of the AR model. This may become undesirable for signal compression applications. An ARMA model can then be used.

Assume an ARMA (p,q) process given by Equation 7.12. The process is stationary if the roots of the characteristic Equation 7.18 are outside the unit circle in the z^{-1} plane.

The inverse filter

$$H^{-1}(z^{-1}) = \frac{A(z^{-1})}{B(z^{-1})} \quad (7.44)$$

is called the ARMA whitening filter since when driven with the given sequence $\{S_k\}$ will yield white noise as its output.²³

B. Parameter Estimation of ARMA Models — Direct Method

When the input to the model is accessible, input-output identification techniques can be used. Many methods, based on maximum likelihood, least squares, instrumental variables, and others, are available. When modeling a signal, the input is inaccessible so that special methods must be used.

Box and Jenkins³ have suggested an iterative method based on the maximum likelihood functions. Other iterative techniques have been suggested based on the whitening filter approach.²⁴ Durbin¹⁸ has suggested a method where an AR model is first fitted to the sequence. The white input noise is estimated by means of the inverse AR model fitted. Having now both input and output sequences, identification methods are used to estimate the ARMA parameters. An improved version of this algorithm was suggested by Mayne and Firoozan.²⁵

A somewhat simpler method²⁶ will be described here. Consider again the ARMA model of Equation 7.11, where the ARMA (p,q) filter is approximated by long division

$$H(z^{-1}) = \frac{B(z^{-1})}{A(z^{-1})} \approx \frac{1}{\bar{A}(z^{-1})} \quad (7.45)$$

The sequence is now modeled by an AR model

$$S_k = - \sum_{i=1}^r \tilde{a}_i S_{k-i} + GU_k \quad (7.46)$$

In Equation 7.46, r is the order of the $\tilde{A}(z^{-1})$ polynomial. It is chosen to be sufficiently high (see Section II). The AR parameters \tilde{a}_i , $i = 1, 2, \dots, r$ are identified using the Durbin's algorithm (Equation 7.30). By cross multiplication, we get from Equation 7.45

$$(1 + \hat{b}_1 z^{-1} + \hat{b}_2 z^{-2} + \dots + \hat{b}_q z^{-q}) (1 + \tilde{a}_1 z^{-1} + \tilde{a}_2 z^{-2} + \dots + \tilde{a}_r z^{-r}) \\ = (1 + \hat{a}_1 z^{-1} + \hat{a}_2 z^{-2} + \dots + \hat{a}_{\hat{p}} z^{-\hat{p}}) + \epsilon(z^{-r+1}) \quad (7.47)$$

where $\epsilon(z^{-i})$ is a polynomial with $i = r+1, r+2, \dots$. Comparing terms and choosing \hat{p} and \hat{q} such that $r \geq \hat{p} + \hat{q}$ we get $\hat{p} + \hat{q}$ equations

$$\boxed{\begin{array}{lcl} \hat{a}_1 & = & \tilde{a}_1 + \hat{b}_1 \\ \vdots & & \vdots \\ \hat{a}_2 & = & \tilde{a}_2 + \tilde{a}_1 \hat{b}_1 + \hat{b}_2 \\ \hat{a}_3 & = & \tilde{a}_3 + \tilde{a}_2 \hat{b}_1 + \tilde{a}_1 \hat{b}_2 + \hat{b}_3 \\ \vdots & & \vdots \\ \hat{a}_{\hat{q}} & = & \tilde{a}_{\hat{q}-1} \hat{b}_1 + \dots + \hat{b}_{\hat{q}} \\ & & \vdots \\ & & \vdots \\ \hat{a}_{\hat{q}+1} & = & \tilde{a}_{\hat{q}+1} + \tilde{a}_{\hat{q}} \hat{b}_1 + \dots + \tilde{a}_1 \hat{b}_{\hat{q}} \\ & & \vdots \\ & & \vdots \\ \hat{a}_{\hat{p}} & = & \tilde{a}_{\hat{p}} + \tilde{a}_{\hat{p}-1} \hat{b}_1 + \tilde{a}_{\hat{p}-2} \hat{b}_2 + \dots + \tilde{a}_{\hat{p}-\hat{q}} \hat{b}_{\hat{q}} \end{array}} \quad (7.48)$$

$$\boxed{\begin{array}{lcl} 0 & = & \tilde{a}_{\hat{p}+1} + \tilde{a}_{\hat{p}} \hat{b}_1 + \dots + \tilde{a}_{\hat{p}-\hat{q}} \hat{b}_{\hat{q}} \\ & & \vdots \\ & & \vdots \\ \hat{q} & = & \tilde{a}_{\hat{p}+\hat{q}} + \tilde{a}_{\hat{p}+\hat{q}-1} \hat{b}_1 + \dots + \tilde{a}_{\hat{p}} \hat{b}_{\hat{q}} \end{array}} \quad (7.49)$$

In Equations 7.48 and 7.49 \tilde{a}_i , $i = 1, 2, \dots, r$ are known and the ARMA parameters \hat{a}_j , $j = 1, 2, \dots, \hat{p}$ and \hat{b}_k , $k = 1, 2, \dots, \hat{q}$ are to be determined. The order (\hat{p}, \hat{q}) of the ARMA model is estimated by (\hat{p}, \hat{q}) . Equation 7.49 provides a set of \hat{q} equations with \hat{q} MA parameters, \hat{b}_k , as unknowns

$$\begin{bmatrix} \tilde{a}_{\hat{p}} & \tilde{a}_{\hat{p}-1} & \dots & \tilde{a}_{\hat{p}+1-\hat{q}} \\ \tilde{a}_{\hat{p}+1} & \tilde{a}_{\hat{p}} & \dots & \tilde{a}_{\hat{p}+2-\hat{q}} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{a}_{\hat{p}-1+\hat{q}} & \dots & \tilde{a}_{\hat{p}} & \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \vdots \\ \hat{b}_{\hat{q}} \end{bmatrix} = - \begin{bmatrix} \tilde{a}_{\hat{p}+1} \\ \tilde{a}_{\hat{p}+2} \\ \vdots \\ \tilde{a}_{\hat{p}+\hat{q}} \end{bmatrix} \quad (7.50)$$

or

$$\tilde{A}_{(\hat{p})} \underline{\hat{b}} = - \underline{\tilde{a}}_{(\hat{p}+1)} \quad (7.51)$$

where the matrix $\tilde{A}_{(\hat{p})}$, $\underline{\hat{b}}$, and $\underline{\tilde{a}}_{(\hat{p}+1)}$ are defined in Equation 7.50. The solution to the MA parameters $\underline{\hat{b}}$ is thus given by

$$\underline{\hat{b}} = - (\tilde{A}_{(\hat{p})})^{-1} \underline{\tilde{a}}_{(\hat{p}+1)} \quad (7.52)$$

Note that the square $\hat{q} \times \hat{q}$ matrix $\tilde{A}_{(\hat{p})}$ has to be nonsingular for the inversion in Equation 7.52.

The \hat{p} AR parameters, $\underline{\hat{a}}$, can now be calculated using the set of Equations 7.48. Define the augmented $(\hat{p} \times 1)$ dimensional vector, \underline{b}_A , as

$$\begin{aligned} \underline{b}_A^T &= [b_1, b_2, \dots, b_{\hat{q}}, 0, \dots, 0] = \\ &[b^T \ ; \ \underline{0}_{(\hat{p}-\hat{q})}] \end{aligned} \quad (7.53)$$

the $\underline{\tilde{a}}_{(1)}$ vector by

$$\underline{\tilde{a}}_{(1)}^T = [\tilde{a}_1, \tilde{a}_2, \dots, \tilde{a}_{\hat{p}}] \quad (7.54)$$

and the $(\hat{p} \times \hat{p})$ matrix \hat{A}_D by

$$\hat{A}_d = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \tilde{a}_1 & 1 & 0 & \dots & 0 \\ \tilde{a}_2 & \tilde{a}_1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{a}_{\hat{p}-1} & \tilde{a}_{\hat{p}-2} & \dots & \tilde{a}_1 & 1 \end{bmatrix} \quad (7.55)$$

The required AR parameters are thus given by

$$\hat{\underline{a}} = \hat{\underline{a}}_{(p)} + \hat{A}_d \underline{b}_A \quad (7.56)$$

Equations 7.52 and 7.56 constitute the ARMA estimation. The estimation, however, is very sensitive to noise and ill conditioning of $\hat{A}_{(p)}$.

C. Parameter Estimation of ARMA Models — Maximum Likelihood Method

An ARMA parameter identification method, based on maximum likelihood estimation, was suggested by Cadzow.²⁷

Consider the ARMA difference Equation 7.12. Multiplying by S_{k+j} and taking expectation, we get the Yule-Walker equations for the ARMA case

$$r_j + \sum_{i=1}^p a_i r_{j-i} = \left\{ \begin{array}{l} \sum_{i=j}^q b_i h_{i-j}; \quad 0 \leq j \leq q \\ 0; \quad j > q \end{array} \right. \quad (7.57)$$

where the sequence $h(k)$ is the weighting sequence (the response to Kronecker delta) of the ARMA filter. Equations 7.57 show that the AR parameters, a_i , appear in a linear fashion for all $j > q$. Consider the set of t equations; evaluated over the range $q+1 \leq j \leq q+t$, these yield

$$\begin{bmatrix} r_{q+1} & r_q & r_{q-1} & \dots & r_{q+1-p} \\ r_{q+2} & r_{q+1} & \dots & \dots & r_{q+2-p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{q+t} & r_{q+t-1} & \dots & \dots & r_{q+t-p} \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (7.58)$$

or in matrix notation

$$R_t \underline{\alpha} = \underline{0} \quad (7.59)$$

where R_t is the $(t \times (p+1))$ autocorrelation matrix, $\underline{0}$ is the $((p+1) \times 1)$ null vector, and $\underline{\alpha}$ is the AR augmented vector:

$$\underline{\alpha}^T = [1 \cdot \underline{a}^T] \quad (7.60)$$

If the order (p,q) is known and the correct (and not estimated) correlations are used to form R_t , Equation 7.59 will have a unique solution for all values of $t \geq p$. This is true since the rank of R_t is equal to $\min(p,t)$. However, both (p,q) and the exact correlation coefficients are not given; thus, estimation methods must be applied.

The unbiased estimates of the autocorrelation coefficients are adopted here

$$\hat{r}_j = \frac{1}{M-j} \sum_{i=0}^{M-j-1} s_{k+i} s_{k+i+j} \quad (7.61)$$

where M is the number of samples in the finite sequence over which the estimation is performed. Estimating the ARMA order (p, q) by (\hat{p}, \hat{q}) , the elements of the estimated correlation matrix \hat{R}_t are given by

$$\begin{aligned} \hat{f}(i,j) &= \hat{f}_{\hat{q}+1+i-j} \\ 1 \leq i \leq t; \quad 1 \leq j \leq \hat{p} \end{aligned} \quad (7.62)$$

The left side of Equation 7.59 with the estimate of the correlation matrix will no longer equal 0. We can define an error vector $\underline{\epsilon}$ such that

$$\hat{R}_t \underline{\alpha} = \underline{\epsilon} \quad (7.63)$$

Since the error is the result of the sum of many random variable products, its joint density function can be assumed gaussian with zero mean and covariance matrix W .

$$p(\underline{\epsilon}) = (2\pi)^{-\frac{M}{2}} |W|^{-1/2} \exp(-\frac{1}{2} \cdot \underline{\epsilon}^T W^{-1} \underline{\epsilon}) \quad (7.64)$$

where the covariance matrix W is given by

$$W = E\{\underline{\epsilon} \underline{\epsilon}^T\} \quad (7.65)$$

The estimate of the AR parameter vector $\underline{\alpha}$ can be performed by maximizing the joint density function. The maximum likelihood³ estimate of $\underline{\alpha}$ is given by the equation

$$\hat{R}_t^T W^{-1} \hat{R}_t \underline{\alpha} = \underline{\mu} \quad (7.66)$$

where the $(p + 1) \times 1$ vector $\underline{\mu}$ is defined by

$$\underline{\mu}^T = [\mu, 0, \dots, 0] \quad (7.67)$$

μ is a constant selected such that the first component of $\underline{\alpha}$ is one. Cadzow¹⁶ has suggested taking the $(p \times p)$ covariance matrix as the identity matrix $W = I$, hence the estimate of the AR parameters becomes

$$\begin{bmatrix} 1 \\ \dots \\ \hat{a} \end{bmatrix} = \hat{\underline{\alpha}} = (\hat{R}_t^T \hat{R}_t)^{-1} \underline{\mu} \quad (7.68)$$

The dimension t should be selected in the range $\hat{p} \leq t \leq M - \hat{q} - 1$. Selecting the maximal value of t will improve model fidelity at the expense of higher calculation load.

Having identified the AR parameters of the ARMA model by Equation 7.68, its MA parameters must be estimated.

Consider transferring the sequence $\{S(k)\}$ through a filter $\hat{A}(z^{-1})$ whose parameters are the estimated AR parameters (Equation 7.68). Denote the output of the filter by $\{\tilde{S}(k)\}$, hence

$$\begin{aligned} \tilde{S}(z) &= S(z) \cdot \hat{A}(z^{-1}) = \frac{B(z^{-1})}{A(z^{-1})} \hat{A}(z^{-1}) U(z) \\ &= \tilde{B}(z^{-1}) U(z) \end{aligned} \quad (7.69)$$

Assuming $\hat{A}(z^{-1}) \approx A(z^{-1})$, then the MA model of Equation 7.69 is close to the MA part of the ARMA model to be identified. The sequence $\{\tilde{S}(k)\}$ is therefore assumed to be a $MA(\hat{q})$ process. Its parameters \hat{b} are identified using techniques discussed in Section III.

V. PROCESS ORDER ESTIMATION

A. Introduction

One of the important decisions that has to be made when modeling data with parametric models is that of estimating the order. In ARMA models, this means the estimation of (p,q) the AR and MA orders.

Order estimation is usually done by means of some optimization technique. It is desirable to have the minimum values for \hat{p} and \hat{q} to reduce the amount of calculations, storage, and ill conditioning problems. It is required, however, to have large enough values of \hat{p} and \hat{q} to adequately represent the process.

Currently, many methods are available for choosing the order of AR or ARMA models (see, for example, References 3, 17, 19 through 22, and 28 through 37). Several basic ideas are used for order determination. Some check the correlation or spectral flatness of the residuals; others use decision rules based on Bayesian approach, maximum likelihood approach, and amount of information measures.

B. Residuals Flatness

For the estimation of the order of an AR process, a measure of residuals correlation has been generally used.¹⁷ As noted by Equation 7.22, the residuals are the estimates of the inaccessible input sequence $\{GU_k\}$ which is an uncorrelated (white) sequence. Hence, for an adequate model, the residuals must be uncorrelated or, in other words, possess flat power spectrum.

Consider an AR model with order \hat{p} , and define the normalized error \hat{V}_p

$$V_{\hat{p}} = \frac{E_p}{\hat{r}_o} \quad (7.70)$$

where E_p is the minimum average error (Equation 7.34) and \hat{r}_o is the total energy in the sequence (Equation 7.27). From Equations 7.32 and 7.70, it is clear that $V_{\hat{p}}$ is a monotonically decreasing function of \hat{p} . It has been shown¹⁷ that $V_{\hat{p}}$ is bounded by

$$V_o = 1 \leq V_{\hat{p}} \leq V_{\min} \quad (7.71)$$

where

$$V_{\min} = \lim_{\hat{p} \rightarrow \infty} V_{\hat{p}} = \frac{1}{r_o} \exp \left(\frac{1}{2\pi T} \int_{-\pi/T}^{\pi/T} \log S_{(\omega)} d\omega \right) \quad (7.72)$$

$S_{(\omega)}$ being the PSD function of the process.

Theoretically, $V_{\hat{p}} = V_p$ for all $\hat{p} \geq p$. One can plot $V_{\hat{p}}$ as a function \hat{p} and check for the model order for which the curve becomes flat. In practice, because estimations are based on finite samples, the curve will not become completely flat. Therefore, the following test has been suggested.¹⁷ Choose that model order, \hat{p} , for which the following test has been succeeded for several consecutive values:

$$1 - \frac{V_{\hat{p}+1}}{V_{\hat{p}}} < \epsilon \quad (7.73)$$

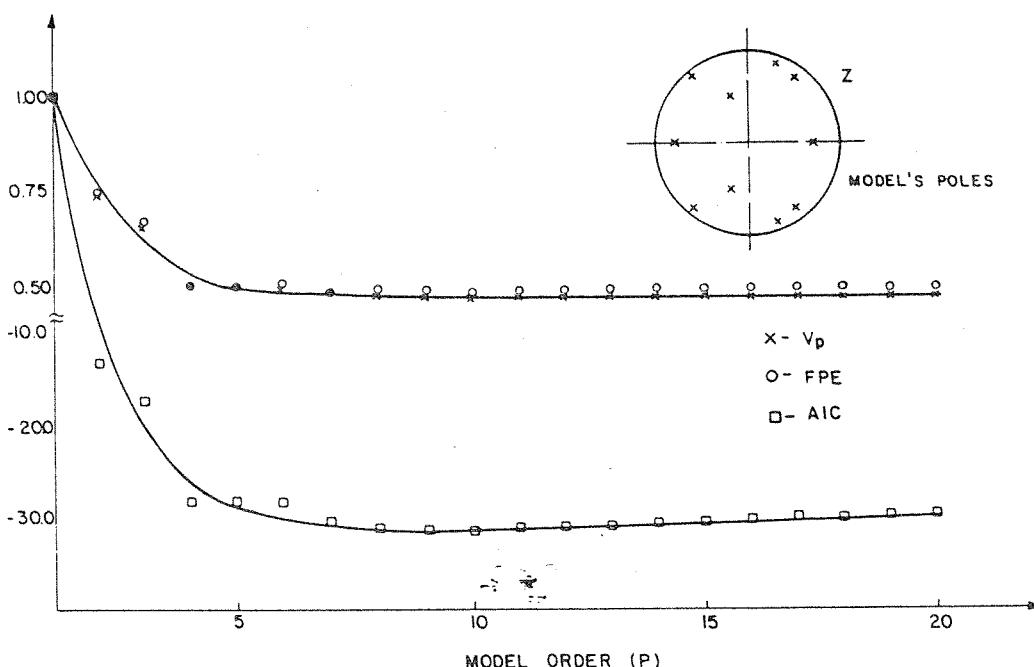


FIGURE 5. Estimation of AR model's order. Data was synthesized from an AR difference equation of order ten.

The test is based on the hypothesis testing methods, rather than on some procedure where multiple decisions are used. This has been a major reason for criticizing the method²⁰ and developing new methods for model order estimation.

In Figure 5, the method for order estimation is demonstrated by means of synthesized data. A sequence $\{S_k\}$ was synthesized from an AR Model of order 10. The curve of $V_{\hat{p}}$ vs. \hat{p} is plotted.

C. Final Prediction Error (FPE)¹⁹

Consider the sequence $\{S_k\}$ $k = 0, 1, \dots, N - 1$ with its linear prediction of order \hat{p} given by Equation 7.21. Its prediction error is $E_{\hat{p}}$ given by

$$E_{\hat{p}} = \frac{1}{N} \sum_{k=0}^{N-1} \left(S_k + \sum_{j=1}^{\hat{p}} \hat{a}_j S_{k-j} \right)^2 \quad (7.74)$$

It can be shown that the expectation of $E_{\hat{p}}$ as N approaches infinity is

$$E\{E_{\hat{p}}\} = \left(1 - \frac{\hat{p} + 1}{N} \right) G^2 \quad (7.75)$$

Consider now another sequence $\{y_k\}$ infinitely long with the same statistical properties as $\{S(k)\}$. The predictor for this sequence will be

$$\hat{y}_k = - \sum_{j=1}^{\hat{p}} \hat{a}_j y_{k-j} \quad (7.76)$$

where a_j , $j = 1, 2, \dots, \hat{p}$ are functions of $\{S_i\}$. The variance of the residuals tends asymptotically to $(1 + (\hat{p} + 1) N^{-1}) G^2$ as N approaches infinity. Hence, it is logical to estimate the final prediction error (FPE) of the predictor (Equation 7.76) by

$$FPE(\bar{p}) = \left(1 + \frac{\bar{p} + 1}{N} \right) G^2 \quad (7.77)$$

where G^2 is the variance of the input sequence $\{GU(k)\}$.

Since G^2 is not known, its estimate from Equation 7.75 can be used such that

$$FPE(\bar{p}) = \frac{1 + \frac{\bar{p} + 1}{N}}{1 - \frac{\bar{p} + 1}{N}} E_{\bar{p}} = \frac{N + \bar{p} + 1}{N - \bar{p} - 1} E_{\bar{p}} \quad (7.78)$$

and the optimal model order \hat{p} is given by minimizing Equation 7.78:

$$\text{Min}_{\bar{p}} FPE(\bar{p}) \quad (7.79)$$

An alternate method which very often gives the same results has been suggested by Parzen.⁶ The method is known as the criterion autoregressive transfer function (CAT). It is based on minimizing the average normalized spectrum error between model of infinite order and of order p .

D. Akaike Information Theoretic Criterion (AIC)

Akaike²⁰ has extended the FPE method to any maximum likelihood situation. He has defined the information theoretic criterion, AIC, as

$$AIC = -2 \ell n (\text{maximized likelihood}) + 2k \quad (7.80)$$

where k is the number of independently adjusted parameters within the model. The AIC is defined as an estimate of twice the negentropy (known also as the Kullback information) of the true structure with respect to the fitted model. The estimates of \hat{a} and $E_{\bar{p}}$ by the Yule-Walker Equation (7.40) are approximately maximum likelihood estimates.¹ Thus the AIC for an AR model is given by

$$AIC(\bar{p}) = \rho_{\omega} N \ell n(E_{\bar{p}}) + 2\bar{p} \quad (7.81)$$

and the optimal model order, \hat{p} , by

$$\text{Min}_{\bar{p}} AIC(\bar{p}) \quad (7.82)$$

In Equation 7.81, ρ_{ω} is a factor that compensates for windowing effects. Its value is taken as the ratio of the energy under the window function used in the data handling to that of rectangular window. For a Hamming window, $\rho_{\omega} = 0.4$.

It has been suggested⁷ that the search for the absolute minimum of Equation 7.81 will be carried out over the range $1 \leq \bar{p} \leq 3 N^{1/2}$.

The AIC is also successfully used⁷ in determining the order of ARMA models. Here the AIC is defined by

$$\text{Min}_{\bar{p}, \bar{q}} (\rho_{\omega} N \ell n E_{\bar{p}, \bar{q}} + 2(\bar{p} + \bar{q})) \quad (7.83)$$

E. III Conditioning of Correlation Matrix

Consider the estimation of the AR parameters by means of Equation 7.66. The estimation is performed by multiplying both sides of Equation 7.66 by the inverse of $\hat{R}_t^T W^{-1} \hat{R}_t$. This $\hat{p} \times \hat{p}$ matrix is well conditioned (i.e., strongly nonsingular) when $\hat{p} = p$. If this matrix becomes ill conditioned, the estimated parameters will be very sensitive to small errors in the estimation of the autocorrelation.

Pao and Lee²⁸ and later Cadzow²⁷ have suggested the estimation of order model \hat{p} as the largest possible value that renders $\hat{p} \times \hat{p}$ matrix which is well conditioned.

Denote the matrix $A(\hat{p})$

$$A(\hat{p}) = \hat{R}_t^T W^{-1} \hat{R}_t = [a_{i,j}]; \quad i, j = 1, 2, \dots, \hat{p} \quad (7.84)$$

and its normalized determinant $|A|_n(\hat{p})$

$$|A|_n(\hat{p}) = \det |a_{i,j}/\sigma_i| \quad (7.85)$$

where

$$\sigma_i^2 = \sum_{j=1}^{\hat{p}} (|a_{i,j}|^2) \quad (7.86)$$

The matrix A becomes ill conditioned when its normalized determinant becomes much smaller than one. The optimal order, \hat{p} , can be defined as the model order for which $|A|_n(\hat{p} + 1)$ exhibits a significant falloff in magnitude.

In the case where the correlations are calculated with no error, the rank of R_t equals the true AR order, p (for $\hat{p} \leq t$, $\hat{p} \geq p$). In the practical case, the matrices will have full rank for all estimated orders, \hat{p} , larger than the correct one. However, it is found that $(\hat{p} - p)$ eigenvalues have values "close" to zero. A procedure for order estimation can thus be formulated as: the optimal order, \hat{p} , is the value for which the estimated correlation matrix \hat{R}_t has $(\hat{p} - \hat{p})$ of its eigenvalues sufficiently close to zero (for all $\hat{p} > \hat{p}$). A method for implementing the above procedure, based on singular values decomposition (SVD), has been suggested by Cadzow.²⁷ Efficient algorithms for calculating the SVD are available.³⁸

VI. LATTICE REPRESENTATION

Time series, modeled by AR model of order \hat{p} , is characterized by the inverse filter $A(z^{-1})$. The coefficients of the filter, \hat{a} , can be estimated from the Yule-Walker Equations 7.25, e.g., by the Durbin algorithm (Equation 7.30). Theoretically, the stability of the all pole filter is guaranteed; however, in practice, due to finite wordlength (FWL) computations, the estimated filter is not always stable.

A lattice formulation has been suggested by Itakura and by Burg (see Makhoul³⁹) for the filter. This formulation guarantees stability with much smaller sensitivity to finite wordlength computations. The estimation of the lattice coefficients, however, require heavy computations. Makhoul³⁹ has suggested an improved lattice formulation (covariance lattice).

The lattice inverse filter is shown in Figure 6. The coefficients of the filter, k_i ; $1 \leq i \leq \hat{p}$, are the reflection, or PARCOR, coefficients (Equation 7.30). Filter stability is guaranteed for $|k_i| < 1$; $1 \leq i \leq \hat{p}$. The PARCORS can be calculated from the LPC (AR) parameters vector \underline{a} by¹⁷ the backward recursion equation

$$k_i = a_i^{(i)} \\ a_j^{(i-1)} = (a_j^{(i)} - a_i^{(i)} a_{i-j}^{(i)}) (1 - k_i^2)^{-1}; \quad 1 \leq j \leq (i - 1) \quad (7.87)$$

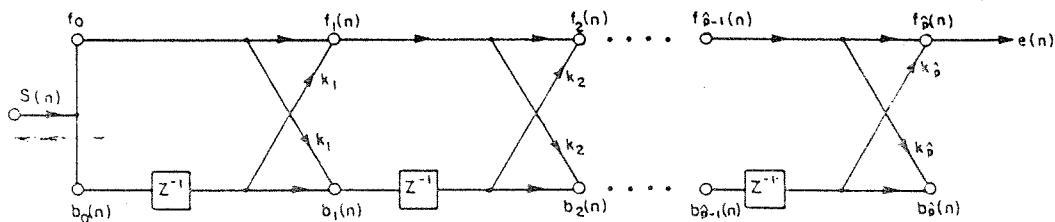


FIGURE 6. Lattice filter.

Equation 7.87 is calculated recursively for $i = \hat{p}, (\hat{p} - 1), \dots, 1$ (in that order) where initially $a_j^{(\hat{p})} = a_j$, $1 \leq j \leq \hat{p}$.

The LPC coefficients can be calculated from the PARCORs by the recursive equation

$$\begin{aligned} a_m^{(m)} &= k_m \\ a_j^{(m)} &= a_j^{(m-1)} + k_m a_{m-j}^{(m-1)}; \quad 1 \leq j \leq (m-1) \end{aligned} \quad (7.88)$$

Equation 7.88 is calculated for $m = 1, 2, \dots, \hat{p}$. After each recursion, the coefficients $a_j^{(m)}$, $1 \leq j \leq m$ are the AR coefficients of the m th order model. The desired AR vector of order \hat{p} is given by $a_j = a_j^{(\hat{p})}$, $1 \leq j \leq \hat{p}$.

Makhoul³⁹ has developed a procedure for the direct estimation of the lattice coefficients with calculation complexity compatible with that of the regular prediction filter. Lattice filters have been suggested also for ARMA models.⁴⁰

VII. NONSTATIONARY PROCESSES

A. Trend Nonstationarity — ARIMA³

Consider a nonstationary continuous time series $y(t)$ such that

$$y(t) = m(t) + x(t) \quad (7.89)$$

where $x(t)$ is zero mean stationary time series and $m(t)$ is continuous trend represented by a polynomial of degree $(d - 1)$.

Signals like the one described in Equation 7.89 frequently occur in biomedical applications. The trend, $m(t)$, in these cases may be, for example, a baseline shift due to undesirable electrode movement in bioelectric signals.

The trend can be removed from the signal $y(t)$ by the repeated difference filter ∇^d . To illustrate this, consider for example a trend polynomial of degree 2:

$$m(t) = m_2 t^2 + m_1 t + m_0 \quad (7.90)$$

Then, applying the difference filter once,

$$\begin{aligned} \nabla y(t) &= y(t) - y(t-1) = m_2 t^2 + m_1 t + m_0 - m_2(t-1)^2 \\ &\quad - m_1(t-1) - m_0 + \nabla x(t) \\ &= m_2 \cdot 2t - m_2 + m_1 + \nabla x(t) \end{aligned}$$

applying the difference operator again,

$$\begin{aligned}\nabla^2 y(t) &= 2tm_2 - m_2 + m_1 - 2(t-1)m_2 + m_2 - m_1 + \nabla^2 x(t) \\ &= 2m_2 + \nabla^2 x(t)\end{aligned}$$

and applying once more the difference operator,

$$\nabla^3 y(t) = \nabla^3 x(t)$$

In the last equation, the trend has been completely removed.

Define the time series $S(t)$ as

$$S(t) = \nabla^d y(t) \quad (7.91)$$

$S(t)$, therefore, is stationary and stationary modeling techniques (such as ARMA modeling) can be applied to its processing. Consider now the discrete sequence $\{S_k\}$ consisting of the samples of $S(t)$, modeled by an ARMA model

$$A(z^{-1}) S(z) = B(z^{-1}) U(z) \quad (7.92)$$

Since $z\{\nabla\} = (1 - z^{-1})$, we get from Equation 7.91

$$S(z) = (1 - z^{-1})^d Y(z) \quad (7.93)$$

Inserting Equation 7.93 into Equation 7.92

$$A(z^{-1})(1 - z^{-1})^d Y(z) = A(z^{-1}) \nabla^d Y(z) = B(z^{-1}) U(z) \quad (7.94)$$

Note that the original process can be written from Equation 7.91 as

$$Y(z) = [(1 - z^{-1})^{-1}]^d S(z) = (\nabla^{-1})^d S(z) \quad (7.95)$$

where

$$\nabla^{-1} = (1 - z^{-1})^{-1} = \sum_{i=0}^{\infty} z^{-i} \quad (7.96)$$

The sequence $\{y_k\}$ can be calculated by transferring back Equation 7.95

$$Y_k = \sum_{i_d=-\infty}^k \dots \sum_{i_2=-\infty}^{i_3} \sum_{i_1=-\infty}^{i_2} S_{i_1} \quad (7.97)$$

The nonstationary sequence $\{Y_k\}$ can be retrieved from the stationary sequence $\{S_k\}$ by d summations (or "integrations"); the process in Equation 7.94 is therefore called an autoregressive integrated moving average or ARIMA (p,d,q) process. The indexes p,d,q denote the order of the AR process, the order of the difference operator, and the order of the MA process, respectively.

The ARIMA (p,d,q) model of Equation 7.94 has q (stable) poles outside the unit circle due to the polynomial $A(z^{-1})$, and d poles on the unit circle at $z = 1$. When solving Equation 7.94 for $\{Y_k\}$ by the inverse transformation, the d th order pole at $z = 1$ is responsible for the trend form present in $\{Y_k\}$, while the stable poles are responsible for the stationary part of $\{Y_k\}$.

B. Seasonal Processes³

Many biological signals exhibit periodic behavior. Some show a period of 12 months, some (like menstrual secretions) have a period of about a month, others have a period of about a day (Circadian rhythms), and others exhibit shorter periods of 30 min, 7 min, and shorter.

Assume a nonstationary time series $\{Y_k\}$ with seasonal behavior of period λ . The analysis of the sequence $Y_k, Y_{k-\lambda}, Y_{k-2\lambda}, \dots$ is of interest. This sequence may possess a trend of order $(d - 1)$. The stationary sequence $\{S_k\}$ is thus given by

$$A(z^{-\lambda}) S(z) = B(z^{-\lambda}) U(z) \quad (7.98)$$

with

$$S(z) = \nabla_{\lambda}^d Y(z) \quad (7.99)$$

where

$$\nabla_{\lambda}^d = (1 - z^{-\lambda})^d \quad (7.100)$$

The nonstationary sequence $\{Y_k\}$ is described by the transfer function equation:

$$Y(z) = \frac{B(z^{-\lambda})}{A(z^{-\lambda})(1 - z^{-\lambda})^d} U(z) \quad (7.101)$$

Note that the term $(1 - z^{-\lambda})$ yields λ equally spaced roots, on the unit circle, at angles $2\pi k/\lambda; k = 0, 1, \dots, \lambda - 1$. These roots describe the λ periodic behavior in $\{Y_k\}$.

VIII. ADAPTIVE SEGMENTATION

A. Introduction

The assumption usually made that the signal under test is stationary does not generally hold for real signals. A more practical assumption is that of taking the signal to be stationary within a certain time window. The length of the window is determined taking into account the nonstationary dynamics. For the speech signal, stationarity can be assumed for windows of about 10 to 20 msec. The short time segments are due to the relatively fast changes in the spectral characteristics of the speech signal. Other, slower changing signals, like the EEG, can be considered stationary for much longer durations.

Thus, it is often convenient to treat the nonstationary signal as piecewise-stationary. That is to say, a signal which is stationary within every given segment. The nonstationarity manifests itself by the changes in spectral characteristics between the various segments. Thus, the continuous changes in the statistical characteristics of the process are modeled by a series of jump changes.

A simple method of segmentation is to divide the signal into a sequence of constant length segments. The length of the segments is determined *a priori* such that it will be short enough to be considered stationary, yet long enough for lowest frequencies of its spectrum to be estimated. Such segmentation is usually used in speech processing.

A better segmentation procedure is the one in which the segments are determined adaptively. Here, the segment length depends upon the dynamics of nonstationarity of the process. Adaptive segmentation procedures are especially important when classification and effective storing and transmission are required.

Adaptive segmentation procedures have been applied to the detection of failures in linear systems.⁴¹⁻⁴³ It has been applied mainly to the detection of abrupt, step-wise changes in the signal statistics due to failure in some systems components. In signal analysis, however, the usual case is the presence of "trend-like" nonstationarity where the statistical characteristics of the process are slowly changing.

Adaptive segmentation procedures have been applied to biomedical signal processing⁴⁴ and, in particular, to the analysis of EEG signals.⁴⁵⁻⁵² In these applications, it is important that the procedure can be implemented on line and will be immune to isolated short-time nonstationarities caused by noise.

B. The Autocorrelation Measure (ACM) Method

This method was suggested by Michael and Houchin⁴⁸ for automatic segmentation of EEG signals. The method uses a constant length reference and sliding windows (Figure 7). The method is based on distance measures defined on the amplitudes and on the spectra of the reference and sliding windows. A linear combination of these two distances is defined as the criterion for segmentation.

Define the normalized amplitude distance measure

$$D_A = \frac{|(r_t(o))^{1/2} - (r_o(o))^{1/2}|}{\text{Min}\{(r_t(o))^{1/2}, (r_o(o))^{1/2}\}} \quad (7.102)$$

where $r_t(o)$ and $r_o(o)$ are the zero lag correlation coefficient of the sliding and reference windows, respectively. Since the zero lag correlation is the energy of the signal, D_A is a normalized distance measure of amplitudes. For the spectral distance measure, we shall consider only the correlation coefficients for lags zero until the lag for which the correlation first becomes negative. Define the modified normalized correlation, $r^o(m)$

$$r^o(m) = \begin{cases} \frac{r(m)}{r(o)}; & m = 0, 1, \dots, m^o - 1 \\ 0; & m^o, m^o + 1, \dots \end{cases} \quad (7.103)$$

where m^o is the lag for which the correlation first becomes negative. The spectral distance

$$D_s = \frac{\sum_{m=1}^q |r_t^o(m) - r_o^o(m)|}{0.5 + \sum_{m=1}^q \text{Min}\{r_t^o(m), r_o^o(m)\}} \quad (7.104)$$

where q is chosen such that the correlations for lag larger than q can be neglected.

The rationale behind Equation 7.104 is that for pure sine waves of different frequencies in the reference and sliding window, the measure yields the difference between the cosine of the frequencies divided by their minimum.

The autocorrelation measure is now defined by a linear combination of the two distances

$$\text{ACM} = \frac{D_A}{R_A} + \frac{D_s}{R_s} \quad (7.105)$$

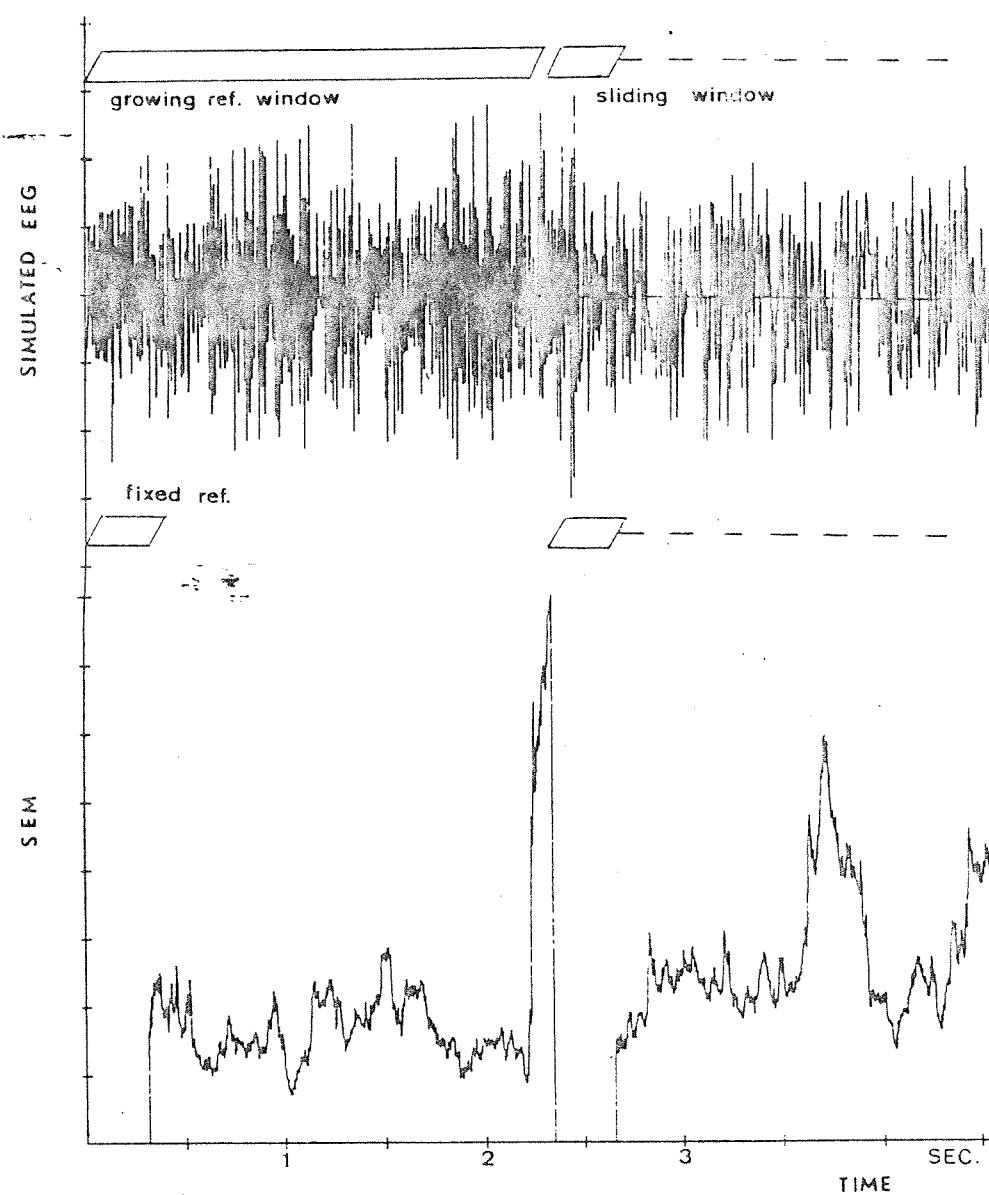


FIGURE 7. Adaptive segmentation. Upper trace: piecewise stationary simulated EEG. First 2.5 sec were simulated from an AR process and last 2.5 sec from a different AR process. Fixed and growing reference windows are demonstrated. Lower trace: the SEM. A new segment has been detected at $t = 2.5$ sec.

where R_A and R_S are constants. To determine a new segment, ACM is compared with a threshold. In Reference 48, the constants were determined such that the threshold was unity.

The boundary of the new segment is located within the sliding window (for which ACM is greater than the threshold) proportionally to the steepness of the autocorrelation measure.

The ACM method is based on somewhat heuristic measures. Its calculations load is, however, relatively low and it is general in the sense that it does not assume a model for the signal under test.

C. Spectral Error Measure (SEM) Method

This method has been suggested by Bodenstein and Praetorius⁴⁵ who have applied it to adaptive segmentation of EEG. The method uses a constant length reference and sliding

windows (Figure 7). The basic idea is to estimate an AR model for the signal in the reference window and to observe the variations of the spectrum in the sliding window. Variations in spectrum are measured by means of the SEM criterion.

Consider the spectrum of the signal estimated through the reference window by means of AR spectral estimation (see Chapter 8). An AR model is fitted to the signal using the reference window data. Denote this by $\hat{A}_o^{-1}(z)$. The estimated signal samples are given by:

$$\hat{S}_o(z) = \frac{GU(z)}{\hat{A}_o(z)} \quad (7.106)$$

The power spectrum of the signal at the reference, $\hat{S}_o(w)$, is given by:

$$\hat{S}_o(w) = \frac{G^2}{|\hat{A}_o(j\omega)|^2} \quad (7.107)$$

since the input signal $\{U_k\}$ is assumed to be white noise with unit variance.

Consider now the inverse filter, $\hat{A}_o(z)$, applied to the samples, $\{s_t(k)\}$, of the sliding window. The output of the filter, the residuals $\{e_t(k)\}$ transformed into the z domain are

$$E_t(z) = \hat{A}_o(z) \cdot S_t(z) \quad (7.108)$$

where $E_t(z)$ and $S_t(z)$ are the Z transform of the residuals and sliding window samples, respectively. The power spectral density function of the residuals can be expressed in terms of the spectrum of the reference $\hat{S}_o(\omega)$ by

$$E_t(\omega) = |\hat{A}_o(z)|^2 S_t(\omega) = G^2 \frac{S_t(\omega)}{\hat{S}_o(\omega)} \quad (7.109)$$

Note that here all functions of ω denote power spectral density functions. The power spectral density of the residuals is the Fourier transform of residuals autocorrelation $\{r_t(k)\}$

$$E_t(\omega) = r_t(0) + 2 \sum_{k=1}^{\infty} r_t(k) \cos(\omega k) \quad (7.110)$$

Equations 7.109 and 7.110 give the relation between the residuals autocorrelations and the ratio of spectra.

Since we are interested in the relative spectrum error, it is logical to define an average error criterion:

$$\epsilon = \int \left(\frac{\hat{S}_o(\omega) - S_t(\omega)}{\hat{S}_o(\omega)} \right)^2 d\omega = \int \left(1 - \frac{S_t(\omega)}{\hat{S}_o(\omega)} \right)^2 d\omega \quad (7.111)$$

Introducing Equations 7.109 and 7.110 into Equation 7.111 and noting the orthogonality of the cosine functions, we get

$$\epsilon = \left(1 - \frac{r_t(0)}{G^2} \right)^2 + \frac{2}{G^4} \sum_{k=1}^{\infty} r_t^2(k) \quad (7.112)$$

When comparing the spectra in the reference and the sliding windows, G is constant. Therefore, Equation 7.112 can be multiplied by G^4 without loss of meaning. Bodensteiner

and Praetorius⁴⁵ also claim that for the EEG application it was shown that normalization to $r_t^2(0)$ was appropriate. Therefore, define the spectral error measure (SEM) as

$$\text{SEM} = \frac{G^4}{r_t^2(0)} \epsilon = \left(\frac{G^2}{r_t(0)} - 1 \right)^2 + 2 \sum_{k=1}^{\infty} \left(\frac{r_t(k)^2}{r_t(0)} \right)^2 \quad (7.113)$$

Since G is not given and only M autocorrelation coefficients are estimated, a more practical measure is

$$\text{SEM} = \left(\frac{\hat{G}^2}{r_t(0)} - 1 \right)^2 + 2 \sum_{k=1}^M \left(\frac{r_t(k)^2}{r_t(0)} \right)^2 \quad (7.114)$$

where \hat{G} is the estimation of the reference AR filter gain given by Equation 7.35. If the signal contains short-term, isolated nonstationarities caused by disturbances, the system may introduce short, meaningless segments. It has been shown⁴⁵ that clipping the residual signal to a predetermined level before calculating the correlation, $r_t(k)$, removes most of the effect of these nonstationarities.

Segmentation is determined by placing a threshold, T_{SEM} , on the SEM. As long as $\text{SEM} \leq T_{\text{SEM}}$, we continue to slide the window and include the data in the current segment. For the first window for which $\text{SEM} > T_{\text{SEM}}$, we declare a new segment. The boundary between segments is placed at the middle of the window. The complete process is then repeated with a reference window initiated at the boundary, serving for the new AR coefficients estimation.

The complete segmentation is given by the following steps:

1. Estimate the AR coefficients from a reference window $k = 1, 2, \dots, M$, yielding the AR filter, $\hat{A}^{-1}(z)$;
2. Estimate the filter gain, G;
3. Set $n = M + 1$;
4. Define the sliding window $k = n, n + 1, \dots, n + M$;
5. Calculate the residuals by passing the data from the sliding window through the inverse filter $\hat{A}(z)$;
6. Clip the residuals to remove the effects of noise. Clipping level has to be determined *a priori*;
7. Calculate the correlations of the sliding window data by $r_t^n(m) = r_t^{n-1}(m) + e_{n+M-m+1}e_{n+M+1} - e_n e_{n+m}$ for $n > M + 1$. The correlations of the first window ($n = M + 1$) have to be calculated in full;
8. Calculate SEM and compare with the threshold T_{SEM} ;
9. If $\text{SEM} \leq T_{\text{SEM}}$, shift the window by one sample ($n \rightarrow n + 1$) and go to step 5;
10. If $\text{SEM} > T_{\text{SEM}}$, a new segment is defined at the middle of the current window;
11. Shift the time base by substituting $k \rightarrow k - n - M/2 + 1$;
12. Return to step 1.

D. Other Segmentation Methods

Several other adaptive segmentation methods have been proposed (see, for example, References 53 to 55). Three of them will be briefly discussed here; the reader is referred to the original papers for a detailed discussion.

Appel and Brandt⁵² have proposed a method based on the generalized likelihood ratio (GLR). In their method, the sliding window is constant in length as is the case in the method discussed previously. However, the reference window is continuously growing and includes all the data up to the first sample of the sliding window (see Figure 7.7). This is definitely

an advantage when abrupt, step-wise changes are expected since the data up to the change does indeed belong to a stationary process. When the nonstationarity is "trend" wise, namely, gradual changes occur, the advantage of this arrangement over the constant length reference is doubtful. The GLR method puts emphasis on the detection of the boundary within the window in which a change has been detected. This is again important in the case of abrupt changes. Its significance is lost when the change is gradual and boundaries of segments are arbitrarily defined by means of some average error allowed.

Basseville and Benveniste⁵⁵ have suggested a procedure based on two AR models with distance measure between them such as the log likelihood ratio and Kullback's divergence between conditional probability laws. They have suggested a continuously growing reference window but with "forgetting" weights that will decrease influence of past samples. This may cause even more severe problems with trend-like nonstationaries than the previous growing reference window.

A method based on parametric families of distributions was suggested by Sclove.⁵³ The assumption here is that each segment is the sample function of a random process with different probability distribution.

A comparison between three segmentation methods — the ACM, SEM, and GLR — has been conducted⁵⁶ on simulated data with parameter jumps and real EEG data. The results show that for the simulated signal (with "jump" nonstationaries), the GLR method is superior. For the EEG data, however, the SEM method may be recommended due to its lower calculations load and satisfactory segmentation properties.

Segmentation approaches based on syntactic methods (see Chapter 13) also have been suggested and applied, for example, to the analysis of speech signals.⁵⁷

REFERENCES

1. Mann, H. B. and Wald, A., On the statistical treatment of linear stochastic difference equations, *Econometrica*, 11, 173, 1943.
2. Kitagawa, G., A nonstationary time series model and its fitting by a recursive filter, *J. Time Series Anal.*, 2, 103, 1981.
3. Box, G. E. P. and Jenkins, G. M., *Time Series Analysis Forcasting and Control*, Holden-Day, San Francisco, 1970.
4. Hannan, E. J., Time series analysis, *IEEE Trans. Autom. Control*, 19, 706, 1974.
5. Robinson, E., *Physical Application of Stationary Time Series*, Macmillan, New York, 1980.
6. Parzen, E., Some recent advances in time series analysis, *IEEE Trans. Autom. Control*, 19, 723, 1974.
7. Sawaragi, Y., Soeda, T., and Nakamizo, T., "Classical" methods and time series analysis, in *Trends and Progress in System Identification*, Eykhoff, P., Ed., Pergamon Press, Oxford, 1981, chap. 3.
8. Anderson, T. W., *The Statistical Analysis of Time Series*, John Wiley & Sons, New York, 1971.
9. Parzen, E., Time series model identification and prediction variance horizon, in *Applied Time Series Analysis*, Vol. 2, Findley, D., Ed., Academic Press, New York, 1981, 415.
10. Granger, C. W. J., Acronyms in time series analysis, *J. Time Series Anal.*, 3, 103, 1982.
11. Wang, D. C. C. and Vagnucci, A. H., TSAN: A package for time series analysis, *Comput. Programs Biomed.*, 11, 132, 1980.
12. Isaksson, A., Wennberg, A., and Zetterberg, L. H., Computer analysis of EEG signals with parametric models, *Proc. IEEE*, 69(4), 451, 1981.
13. Eykhoff, P., *System Identification: Parameter and State Estimation*, John Wiley & Sons, London, 1974.
14. Jakeman, A. J. and Young, P. C., Advanced methods of recursive time series analysis, *Int. J. Control.*, 37, 1291, 1983.
15. Cadzow, J. A., ARMA time series modeling: an effective method, *IEEE Trans. Aerosp. Electron. Syst.*, 19, 49, 1983.
16. Cadzow, J. A., ARMA modeling of time series, *IEEE Trans. Pattern Anal. Mach. Intelligence*, 4, 124, 1982.
17. Makhoul, J., Linear prediction: a tutorial review, *Proc. IEEE*, 63, 561, 1975.

18. Durbin, J., The fitting of time series models, *Rev. Inst. Int. Stat.*, 28(3), 233, 1961.
19. Akaike, H., Statistical prediction identification, *Ann. Inst. Stat. Math.*, 22, 203, 1974.
20. Akaike, H., A new look at the statistical model identification, *IEEE Trans. Autom. Control*, AC-19, 716, 1974.
21. Akaike, H., Fitting autoregressive models for prediction, *Am. Inst. Stat. Math.*, 21, 243, 1969.
22. Akaike, H., Likelihood of a model and information criteria, *J. Econ.*, 16, 3, 1981.
23. Bittani, S., Is the prediction of regression model white?, *J. Franklin Inst.*, 315, 239, 1983.
24. Tretter, S. A. and Steiglitz, K., Power spectrum identification in terms of rational models, *IEEE Trans. Autom. Control*, AC-12, 185, 1967.
25. Mayne, D. Q. and Firoozan, F., Linear identification of ARMA processes, *Automatica*, 18, 461, 1982.
26. Graupe, D., Krause, D. J., and Moore, J. B., Identification of autoregressive moving average parameters of time series, *IEEE Trans. Autom. Control*, AC-20, 104, 1975.
27. Cadzow, J. A., Spectral estimation: an overdetermined rational model equation approach, *Proc. IEEE*, 70(9), 907, 1982.
28. Pao, Y. and Lee, D. T., Performance characteristics of the Cadzow modified direct ARMA method for spectrum estimation, Proc. 1st Acoust. Speech and Signal Process. Workshop on Spectral Estimation, McMaster University, Hamilton, Ontario, Canada, Aug. 1981, 2.51-2.5.10.
29. Akaike, H., Modern development of statistical methods, in *Trends and Progress in System Identification*, Eykhoff, P., Ed., Pergamon Press, Oxford, 1981, chap. 3.
30. Suzumura, N. and Ishii, N., Estimation of the order of autoregressive process, *Int. J. Syst. Sci.*, 8, 905, 1977.
31. Kashyap, R. L., Optimal choice of AR and MA parts in ARMA models, *IEEE Trans. Pattern Anal. Mach. Intelligence*, 4, 99, 1982.
32. Chaure, C. and Benveniste, A., AR and ARMA identification algorithms of Levinson type: an innovation approach, *IEEE Trans. Autom. Control*, 26, 1243, 1981.
33. Hannan, E. J., The estimation of the order of an ARMA process, *Ann. Stat.*, 8, 1071, 1980.
34. Schwarz, G., Estimating the dimension of a model, *Ann. Stat.*, 6, 461, 1978.
35. Rissanen, J., Modelling by shortest data description, *Automatica*, 14, 465, 1978.
36. Ishii, N., Iwata, A., and Suzumura, N., Evaluation of an autoregressive process by information measure, *Int. J. Syst. Sci.*, 9, 743, 1978.
37. Moden, Y., Yamada, M., and Arimoto, S., Fast algorithm for identification of an ARX model and its order determination, *IEEE Trans. Acoust. Speech Signal Process.*, 30, 390, 1982.
38. Haimi Cohen, R. and Cohen, A., A rapid gradient search algorithm for computing partial SVD and principal component decompositions, to be published in *IEEE Trans. Pattern Anal. Mach. Intelligence*, in press.
39. Makhoul, J., Stable and efficient lattice methods for linear prediction *IEEE Trans. Acoust. Speech Signal Process.*, 25, 423, 1977.
40. Friedlander, B., Instrumental variable methods for ARMA spectral estimation, *IEEE Trans. Acoust. Speech Signal Process.*, 31, 404, 1983.
41. Willsky, A. S., A survey of design methods for failure detection in dynamic systems, *Automatica*, 12, 601, 1976.
42. Willsky, A. S. and Jones, H. L., A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems, *IEEE Trans. Autom. Control*, 21, 108, 1976.
43. Rao, T. S., A cumulative sum test for detecting change in time series, *Int. J. Control.*, 34, 285, 1981.
44. Vasseur, C. P. A., Rajagopalan, C. V., Couvreur, M., Toullettes, J. M., and Dubois, O., A microprocessor oriented segmentation technique: an efficient tool for electrophysiological signal analysis, *IEEE Trans. Instrum. Meas.*, 28, 259, 1979.
45. Bodenstein, G. and Praetorius, H. M., Feature extraction from electroencephalogram by adaptive segmentation, *Proc. IEEE*, 65, 642, 1977.
46. Sanderson, A. C., Segen, J., and Richey, E., Hierarchical modeling of EEG signals, *IEEE Trans. Pattern Anal. Mach. Intelligence*, 2, 405, 1980.
47. Jansen, B. H., Hasman, A., and Lenten, R., Piecewise analysis of EEG using AR modeling and clustering, *Comput. Biomed. Res.*, 14, 168, 1981.
48. Michael, D. and Houchin, J., Automatic EEG analysis: a segmentation procedure based on autocorrelation function, *Electroencephalogr. Clin. Neurophysiol.*, 46, 232, 1979.
49. Barlow, J. S., Creutzfeldt, O. D., Michael, D., Honchin, J., and Epelbaum, H., Automatic adaptive segmentation of clinical EEGs, *Electroencephalogr. Clin. Neurophysiol.*, 51, 512, 1981.
50. Bodenstein, G. and Schneider, W., Pattern recognition of clinical electroencephalograms, Proc. Int. Conf. Dig. Signal Process., Florence Italy, 1981, 206.
51. Praetorius, H. M., Bodenstein, G., and Creutzfeldt, O. D., Adaptive segmentation of EEG records: a new approach to automatic EEG analysis, *Electroencephalogr. Clin. Neurophysiol.*, 42, 84, 1977.

52. Appel, U. and Brandt, A. V., Adaptive sequential segmentation of piecewise stationary series, *Inf. Sci. N.Y.*, 29, 27, 1983.
53. Sclove, S. L., Time series segmentation: a model and a method, *Inf. Sci. N.Y.*, 29, 7, 1983.
54. Segen, J. and Sanderson, A. C., Detecting changes in time-series, *IEEE Trans. Inf. Theory*, 26, 249, 1980.
55. Basseville, M. and Benveniste, A., Sequential segmentation of nonstationary digital signals using spectral analysis, *Inf. Sci. N.Y.*, 29, 57, 1983.
56. Appel, U. and Brandt, A. V., A comparative study of three sequential time series segmentation algorithms, *Signal Process.*, 6, 45, 1984.
57. De Mori, R., *Computer Models of Speech Using Fuzzy Algorithms*, Plenum Press, New York, 1983.

Chapter 8

SPECTRAL ESTIMATION

I. INTRODUCTION

Biomedical signals most often are the result of processes that take place in the time domain. The analysis of such signals, however, may be more convenient and effective in the frequency domain. This may be the case both in the deterministic and in the stochastic cases. In most cases, power spectral density function (PSD), known also as "the spectrum", is of interest (see Chapter 6).

Spectral analysis of the EEG¹ has been used both for clinical and research purposes. Detailed spectral analysis can be applied to the automatic classification of sleep states² and of the depth of anaesthesia³ as well as to the classification of a variety of neurological disorders. Spectral analysis of EMG signals serves also in the clinic and in research work. It has been shown, for example, that muscle fatigue⁴ can be characterized and, to some extent, predicted by processing the EMG spectrum. Power spectral analysis of speech signals has been used to assist in the diagnosis of laryngical disorders.⁵ The analysis of hand tremors, pressure and flow waveforms are but a few examples of the application of PSD processing in biomedical signals.

The exact PSD function cannot, in general, be calculated. The given signal is time limited, nonstationary, and corrupted by noise. It is necessary, therefore, to estimate the PSD from the given, short data record. Earlier methods of PSD estimation were based on the estimation of the Fourier transform. An important step in modern spectral analysis was Wiener's work⁶ which established the theoretical framework for the treatment of stochastic processes. Wiener and, independently, Khinchin⁷ have shown the Fourier transform relationship between the autocorrelation function (of a stationary process) and its PSD. This is often referred to as the Wiener-Khinchin relationship.

Prior to the introduction of the FFT algorithm in 1965, the accepted method for estimating the PSD was based on the implementation of the Wiener-Khinchin relationship, suggested by Tukey and Blackman.⁸ According to this method, the discrete autocorrelation coefficients are first estimated using the sequence of windowed data. The windowed correlation is then Fourier transformed to provide the estimated PSD. The Blackman-Tukey estimation procedure suffers from poor resolution, high cost of computation, and poor accuracy mainly due to sidelobes of the window which may even produce negative estimates for the positive PSD.

Since the introduction of the FFT algorithm, a lot of progress has been made in spectral estimation. Methods based on the FFT and on time series analysis have been extensively researched and reported in books⁹⁻¹⁴ and articles.¹⁵⁻²⁵ Attempts to provide a uniform approach to the various estimation methods or to categorize them have been reported.^{22-23,31}

This chapter describes several methods of PSD estimation. The Blackman-Tukey procedure is briefly discussed followed by the more modern approach of the periodograms. Time series analysis techniques based on AR (and the maximum entropy method), MA, and ARMA models are presented in more details. More specialized cases of the Pisarenko harmonic decomposition (PHD) and Prony's methods are discussed as well as the maximum likelihood method (MLM).

Finally, a comparison of the various techniques is made with the goal of providing some help in choosing the right method for a given application.

II. METHODS BASED ON THE FOURIER TRANSFORM

A. Introduction

We shall consider two approaches for PSD estimation based on the Fourier transform. The first is the direct method suggested by Tukey and Blackman⁸ which uses the discrete form of the Wiener-Khinchin relation. This method requires first the estimation of correlation coefficients and second the application of DFT (to the correlation sequence) to get the desired PSD estimation. The second approach is the indirect approach known as the periodogram. Here the estimation is achieved by applying the DFT operator directly to the (windowed) data and then smoothing or averaging the absolute values of the DFT.

In general, these two methods do not yield identical results. However, if a certain biased estimator is used for the correlation estimation, and as many correlation coefficients as data samples are used, then the two methods do yield identical results.²⁶

One major problem of the Fourier transform PSD estimation methods is due to finite time data sequence used. The PSD estimated is not that of the process, but that of a sample function multiplied by a window. In the frequency domain, this yields the PSD of the process (the desired function) convoluted with Fourier transform of the window. When the power of the signal is concentrated in a narrow bandwidth, the convolution operation will "spread" the power into adjacent frequency regions, a phenomenon known as *leakage*. The leakage causes both lack of resolution and inaccuracies. The power due to weak sinusoidal components in the signal can be completely masked by the sidelobes of adjacent stronger sinusoids. For a rectangular window, for example, the resolution (the 3 db of main lobe width) is approximately¹⁵ the inverse of the observation time ($N\Delta t$). Better weighting windows are available²⁷⁻²⁸ (see also Appendix B, Volume II).

One way to reduce the finite observation time problem is to use extrapolation techniques to estimate the data outside the observation window.⁸⁵ Another problem arises when the number of data samples is small. A common procedure, known as *zero padding*, is often used. Zeroes are added to the given data sequence (autocorrelation in the case of Blackman-Tukey algorithm or actual signal in the case of the periodogram) before applying the DFT. The result is the estimation of the PSD with additional interpolated values within the given frequency range. The basic resolution of the estimation is not improved by the zero-padding technique; the results are, however, much smoother and in some cases with less ambiguities in peak spectral determination.

Consider a data sequence $\{x_k\}$, $k = 0, 1, \dots, P - 1$. Its DFT is given by

$$X_m = \Delta t \sum_{k=0}^{P-1} x_k \exp(-j2\pi m k/P)$$

$$m = 0, 1, \dots, P - 1 \quad (8.1)$$

where Δt is the sampling interval. The estimation range in the frequency domain is given by:

$$-\pi/\Delta t \leq w \leq \pi/\Delta t \quad (8.2)$$

If we now pad the data with L zeroes, we get a "new" data sequence $\{x_k\}$, $k = 0, 1, \dots, P - 1 + L$ with

$$x_k = 0; \quad k = P, P + 1, \dots, P - 1 + L \quad (8.3)$$

The DFT of the zero-padded sequence is

$$X_{\tilde{m}} = \Delta t \sum_{k=0}^{P-1} x_k \exp \{-j2\pi \tilde{m}k/(P+L)\}$$

$$\tilde{m} = 0, 1, \dots, P-1+L \quad (8.4)$$

The DFT (Equation 8.4) of the padded sequence has the same frequency region (Equation 8.2) as that of the nonpadded signal (Equation 8.1), but with $P+L$ spectral points rather than P spectral points. If we choose $L = qP$ where q is an integer, then Equations 8.1 and 8.4 are equal for all frequency samples $\tilde{m} = (1+q)m$. An efficient FFT algorithm for zero-padded sequences is available.²⁹

B. The Blackman-Tukey Method

Assume a wide sense stationary process $x(t)$. It is desired to estimate its PSD function. The Wiener-Khinchin theorem relates the PSD with the autocorrelation function

$$S(w) = \int_{-\infty}^{\infty} r_x(\tau) \exp(-jw\tau) d\tau \quad (8.5)$$

Consider the discrete form of Equation 8.5

$$\hat{S}(w) = \Delta t \sum_{m=-M}^{M} \hat{r}_x(m) \exp(-jwm\Delta t) \quad (8.6)$$

where Δt is the sampling interval and $\hat{r}_x(m)$, $m = -M, \dots, M$ are the discrete estimates of the correlation function.

The correlation coefficients are estimated from the sampled data sequence $\{x_k\}$, $k = 0, 1, \dots, N-1$. Rather than using the unbiased estimator

$$\frac{1}{N-m} \sum_{n=0}^{N-m-1} x_{n+m} x_n; \quad m = 0, 1, \dots, P \leq N-1 \quad (8.7)$$

a biased estimator has been suggested:¹⁴

$$\begin{aligned} \hat{r}_x(m) &= \frac{1}{N} \sum_{n=0}^{N-m-1} x_{n+m} x_n \\ \hat{r}_x(-m) &= \hat{r}_x(m); \quad m = 0, 1, \dots, P \leq N-1 \end{aligned} \quad (8.8)$$

This biased estimator of the correlation function has an expectation

$$E\{\hat{r}_x(m)\} = \frac{N-m}{N} r_x(m) \quad (8.9)$$

which is the true autocorrelation function weighted by a triangular weighting window (Bartlett window). Estimator 8.8 tends to have lower mean square error than the unbiased one for many finite data sequences. Note that this type of estimator has been used for the estimation of the AR coefficients (see Equation 7.27).

The Blackman-Tukey⁸ PSD estimation is given by using the estimates of Equation 8.8 in Equation 8.6. Equation 8.6 can be solved by means of the FFT. The sequence $\hat{r}_x(m)$ used in Equation 8.6 can be zero-padded as discussed in the previous section.

Figure 1 shows the Blackman-Tukey estimated PSD of synthesized sine waves corrupted with PRBN. Three sine waves were used and the estimation was performed with various

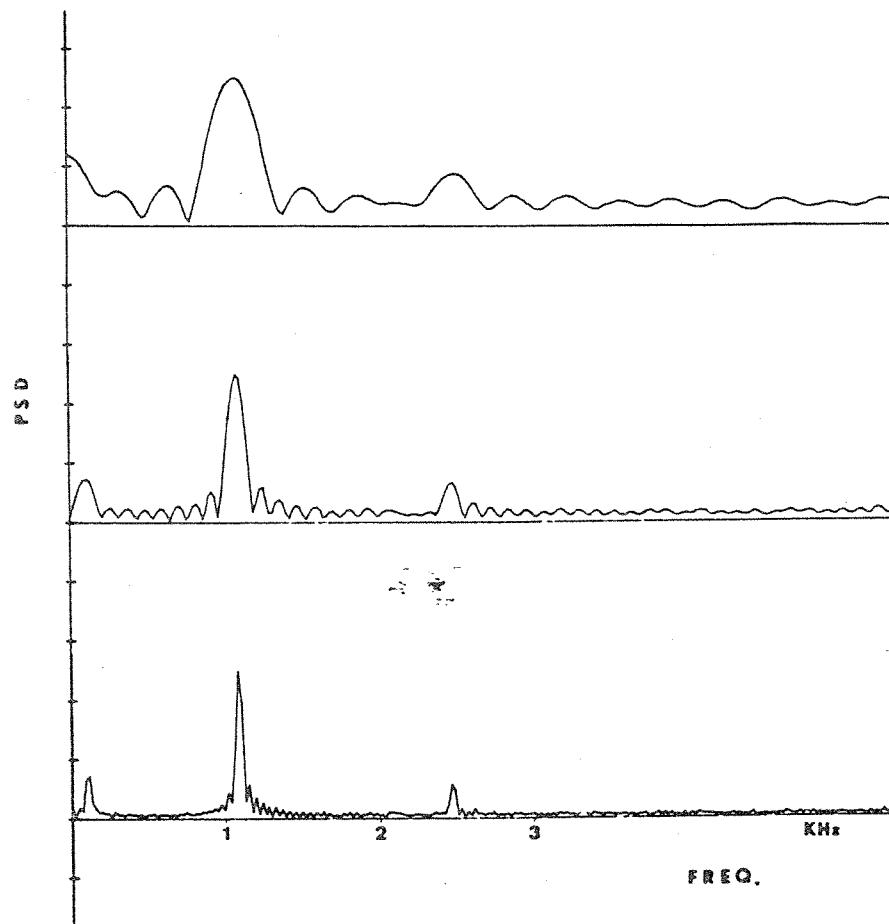


FIGURE 1. Power spectral density function estimation by the Blackman-Tukey method. Synthesized signal consisting of three sinusoids with additive white noise. Upper trace: 32 correlation coefficients and 480 padding zeroes. Middle trace: 256 correlation coefficients and 416 padding zeroes. Lower trace: 256 correlation coefficients and 256 padding zeroes.

numbers of data points, observation times, and zero padding. Figure 2 shows the estimation of an EMG signal by the Blackman-Tukey method.

C. The Periodogram

1. Introduction

The periodogram is a method for PSD estimation using the data sequence without the need to first estimate the correlation coefficient.

It can be shown⁹ that the Wiener-Khinchin relation (Equation 8.5) can be rewritten as:

$$S(w) = \lim_{T \rightarrow \infty} E \left\{ \frac{1}{2T} \left| \int_{-\tau}^{\tau} x(t) \exp(-jwt) dt \right|^2 \right\} \quad (8.10)$$

Implementation of Equation 8.10 is impractical since it requires both infinite time integration and statistical expectation. The periodogram is thus an estimator of Equation 8.10, which can be practically implemented.

Consider the wide sense stationary process $x(t)$ with the windowed data sequence $\{x_k\}$ such that

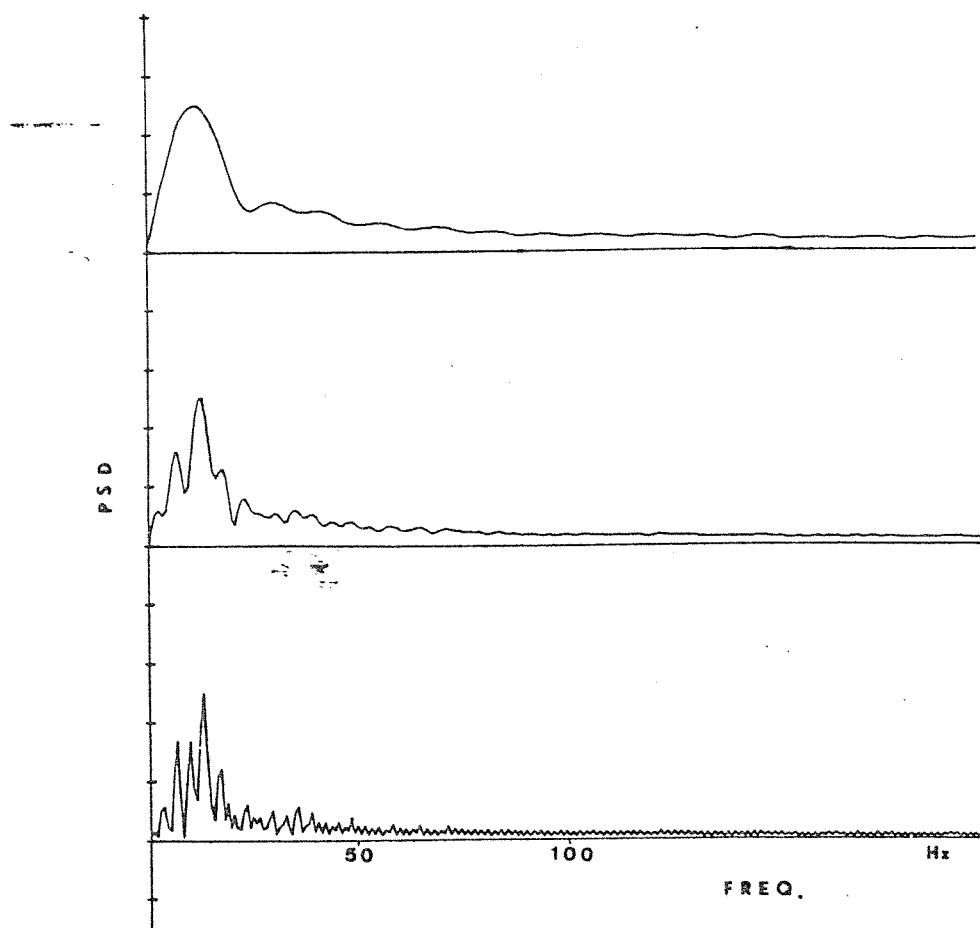


FIGURE 2. Power spectral density function estimation by the Blackman-Tukey method. Surface EMG recorded over the respiratory diaphragmatic muscle, sampled at 400 Hz. Traces are same as in Figure 1.

$$x_k = \begin{cases} w(k) x(k \Delta t); & k = 0, 1, \dots, N - 1 \\ 0; & \text{otherwise} \end{cases} \quad (8.11)$$

where $w(k)$ is the window function and N is the number of samples in the sequence. Use the following estimator for the autocorrelation:

$$\hat{r}_x(m) = \frac{1}{N} \sum_{k=-\infty}^{\infty} x_k x_{k+m}; \quad -\infty < m < \infty \quad (8.12)$$

Note that since we have defined an infinite sequence in Equation 8.11 we can talk about infinite dimensions for the correlation function estimates. Introducing the estimates of Equation 8.12 into the PSD estimation Equation 8.6 yields

$$\hat{S}(w) = \frac{\Delta t}{N} \sum_{m=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} x_k x_{k+m} \exp(-jwm\Delta t) \quad (8.13)$$

which can be rewritten as

$$\hat{S}(w) = \frac{\Delta t}{N} \sum_{k=-\infty}^{\infty} x_k \text{EXP}(jwk\Delta t) \cdot \sum_{m=-\infty}^{\infty} x_{k+m} \exp(-jw(k+m)\Delta t) \quad (8.14)$$

Changing the variable of the inner summation $n = k + m$ we get

$$\hat{S}(w) = \frac{1}{N\Delta t} X(w) \cdot X^*(w) = \frac{1}{N\Delta t} |X(w)|^2 \quad (8.15)$$

where $X(w)$ is the DFT of x_k given by Equation 8.1 and an asterisk (*) denotes the conjugate. Note that $|X(w)|^2$ is the energy distribution function. The division by Δt was required in order to get the PSD.

The major advantage of the periodogram method is the fact that one can use the efficient FFT algorithm to compute the DFT. When calculating the square absolute value of the DFT of the sequence $\{x_k\}$ by means of the FFT we get:

$$|X^F(w)|^2 = \left| \sum_{k=0}^{n-1} x_k \exp(-j2\pi mk/N) \right|^2 \quad (8.16)$$

where X^F is the FFT result which is scaled differently than the quantity calculated from Equation 8.1. When using the FFT, a scaling factor must be used such that

$$\hat{S}(w) = \frac{\Delta t}{N} |X^F(w)|^2 \quad (8.17)$$

The periodogram (Equation 8.17), though an efficient estimator from calculations point of view, has been shown to have large variance. Example of periodograms of white noise process in which the variance does not decrease even when longer sequences are used have been demonstrated.²⁶ The large variance comes as no surprise since in the estimation 8.15 the expectation operator present in Equation 8.10 has been ignored.

The expectation and variance of the periodogram will be discussed followed by several methods of smoothing and averaging.

2. The Expected Value of the Periodogram

Consider the expectation of the periodogram. For Equation 8.13 and using Equation 8.12 we get

$$E\{\hat{S}(w)\} = \Delta t \sum_{m=-\infty}^{\infty} E\{\hat{r}_x(m)\} \exp(-jwm\Delta t) \quad (8.18)$$

The expected value of the correlation depends on the correlation estimation, namely, the window used in Equation 8.11. Hence the expectation of the correlation

$$\begin{aligned} E\{\hat{r}_x(m)\} &= \frac{1}{N} \sum_{k=-\infty}^{\infty} E\{x(k\Delta t) x(k+w\Delta t)\} w(k) w(k+m) = \\ r_x(m) \frac{1}{N} \sum_{k=-\infty}^{\infty} w_k w_{k+m} &= r_x(m) r_w(m) \end{aligned} \quad (8.19)$$

where $r_x(m)$ are the true process correlation coefficients and $r_w(m)$ are the correlation coefficients of the window. Introducing Equation 8.19 into Equation 8.18 we get

$$E\{\hat{S}(w)\} = \Delta t \sum_{m=-\infty}^{\infty} r_x(m) r_w(m) \exp(-jwm\Delta t) \quad (8.20)$$

The periodogram is thus a biased estimator, with bias resulting from the window $r_w(m)$. In the windows used, the bias vanishes as N approaches infinity.

When a rectangular window is used,

$$w_{(k)} = \begin{cases} 1; & 0 \leq k \leq N - 1 \\ 0; & \text{otherwise} \end{cases} \quad (8.21)$$

the window correlation coefficients are given by

$$r_w(m) = \begin{cases} 1 - \frac{|m|}{N}; & |m| \leq N \\ 0; & \text{otherwise} \end{cases} \quad (8.22)$$

and the expected value of the periodogram (given rectangular window) is

$$E\{\hat{S}(w)\} = \Delta t \sum_{m=-N}^N r_x(m) \left(1 - \frac{|m|}{N}\right) \exp(-jwm\Delta t) \quad (8.23)$$

Note that Equation 8.20 gives the expected value of the PSD estimator in terms of the DFT of the multiplication of the two sequences $r_x(m)$ and $r_w(m)$. Applying the complex convolution theorem we get

$$E\{\hat{S}(w)\} = \frac{\Delta t}{2\pi} \int_{-\pi/\Delta t}^{\pi/\Delta t} S(\eta) R_w(w - \eta) d\eta \quad (8.24)$$

where $S(w)$ is the true PSD and $R_w(w)$ is the DFT of $r_w(m)$. Equation 8.24 states that the expected value of the periodogram is the true PSD "viewed" through the filter $R_w(w)$.

Consider, for example, the use of a rectangular data window (Equation 8.21). The correlation window is then given by Equation 8.22, and $R_w(w)$ is

$$R_w(w) = \frac{1}{N} \left(\frac{\sin WN/2}{\sin WN} \right)^2 \quad (8.25)$$

Figure 3 shows the data window, the correlation window, lag window in the time and in the frequency domain, and a schematic description of the frequency convolution (Equation 8.24). Here the process under investigation, whose true PSD is $S(w)$, was chosen as a random process with strong spectral peak at $w = w_1$. Figure 3D depicts the layout for the estimation of $S(w_0)$ by Equation 8.24. The expected value of the estimation at this frequency is given by the area under the multiplication of the two curves. In this particular example, the estimation of the PSD at $w = w_0$ may be largely due to the effect of the first sidelobe of R_w which coincides with the peak of $S(w)$. Hence, the leakage due to the peak and first sidelobe will cause $\hat{S}(w_0)$ to be large even though $S(w_0)$ is very small.

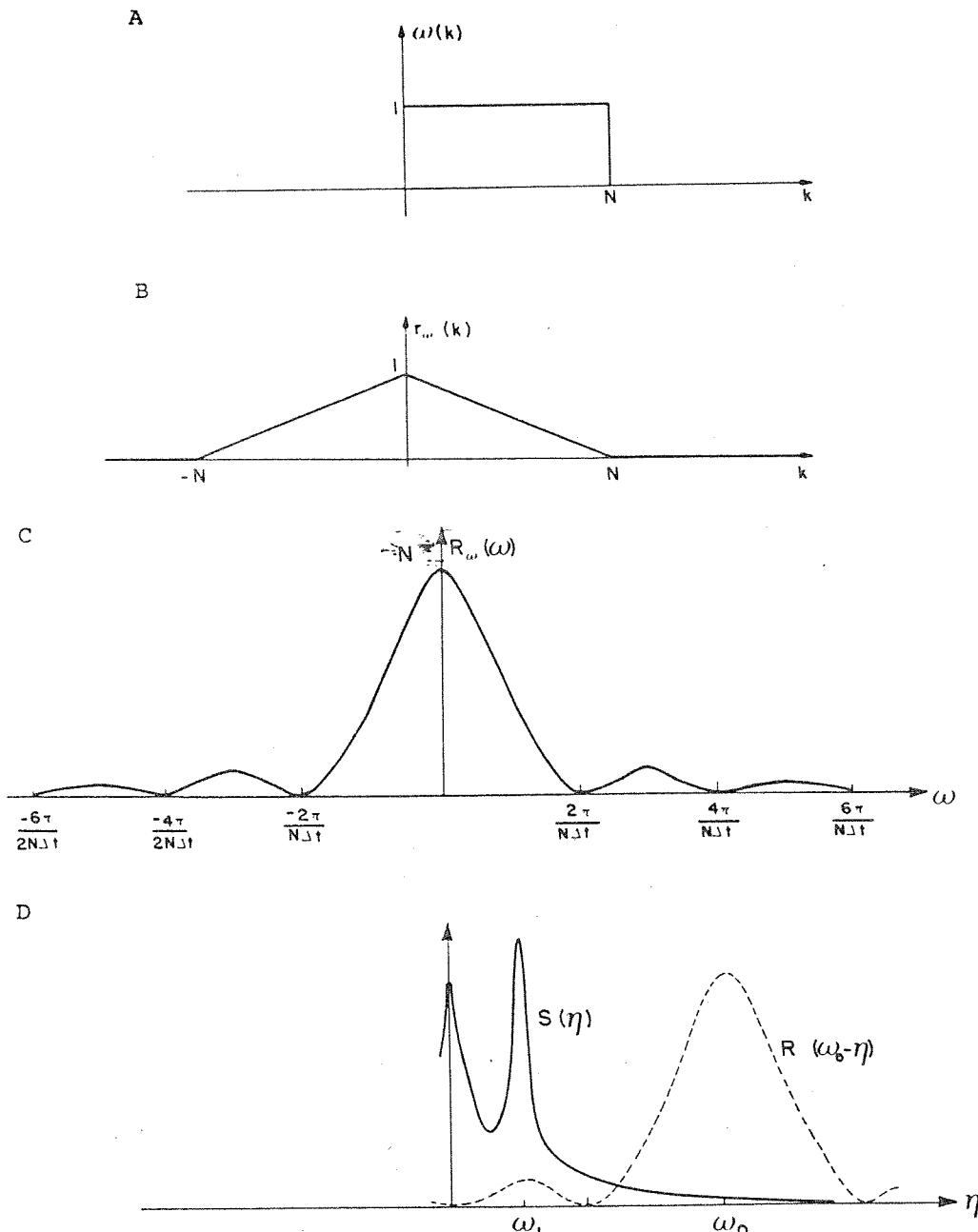


FIGURE 3. Power spectral density estimation with rectangular data window. (A) Data window; (B) correlation (lag) window in time domain; (C) correlation window in frequency domain; (D) frequency convolution.

The main sidelobe for this window has the width of $4\pi/N\Delta t$. As N increases, the window becomes narrower reducing the leakage. At the limit, as N approaches infinity, the window becomes a delta function having no leakage at all. A detailed discussion on the various windows used for signal analysis is presented in the appendix.

3. Variance of the Periodogram

In addition to being a biased estimator, the periodogram suffers from large variance. Unfortunately, in many cases, the variance does not become smaller even when N is increased.

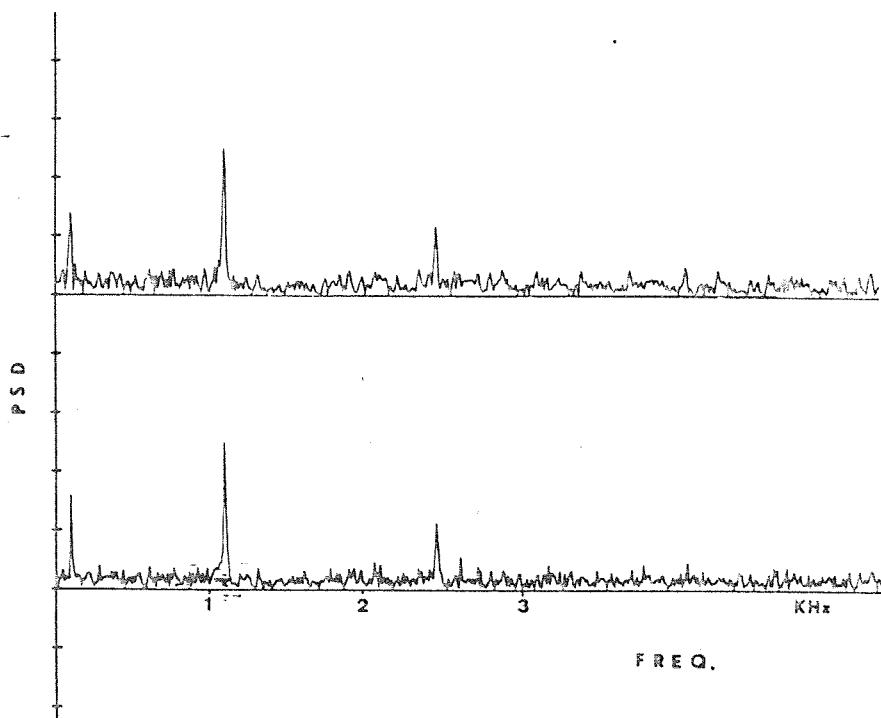


FIGURE 4. Power spectral density function estimation by means of the periodogram. Synthesized noisy sinusoids as in Figure 1. Upper trace: 512 samples and 512 padding zeroes. Lower trace: 1024 samples, no padding zeroes.

Jenkins and Watts¹⁴ have shown that for a gaussian process

$$\lim_{N \rightarrow \infty} [\hat{S}(w)] = S^2(w) \quad (8.26)$$

If we define the standard deviation of the periodogram as the measure for noise, then the signal-to-noise ratio for the periodogram of a gaussian process with an infinitely large data window is only 1. Nongaussian processes yield approximately the same results.

The main reason for the large variance of the periodogram is the fact that the estimator (Equation 8.15) does not take into account averaging as indicated by the expectation operation in Equation 8.10.

A detailed discussion on the variance of the periodogram and on confidence limits of the PSD estimation is given by Koopmans.⁹

An estimation of the PSD with periodograms is demonstrated in Figures 4 and 5. The same signals used for the Blackman-Tukey PSD estimation (Figures 1 and 2) are used here.

4. Weighted Overlapped Segment Averaging (WOSA)

An effective method for the reduction of the large variance is to average several periodograms calculated from finite segments of the stationary time sequence. An improved method has been suggested by Welch.³⁰

Consider the sequence $\{x_k\}$, $k = 0, 1, \dots, N - 1$. We define segments of length L , such that the i th segment is given by the sequence

$$x_k^i = x(k + (i - 1)D) \quad k = 0, 1, \dots, L - 1 \\ i = 1, 2, \dots, I \quad (8.27)$$

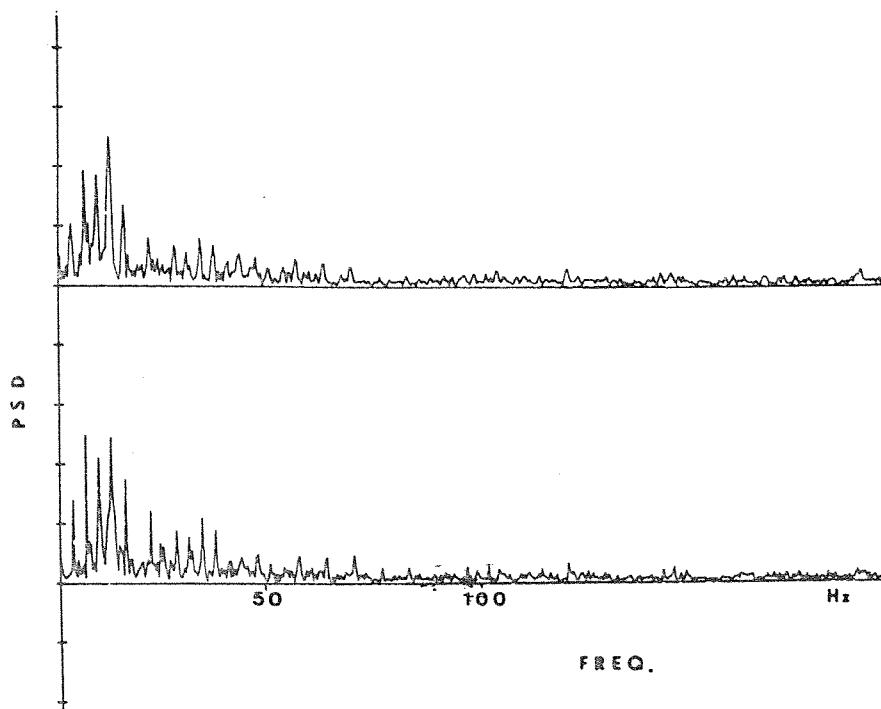


FIGURE 5. Power spectral density function estimation by means of the periodogram. Surface EMG as in Figure 2. Traces as in Figure 4.

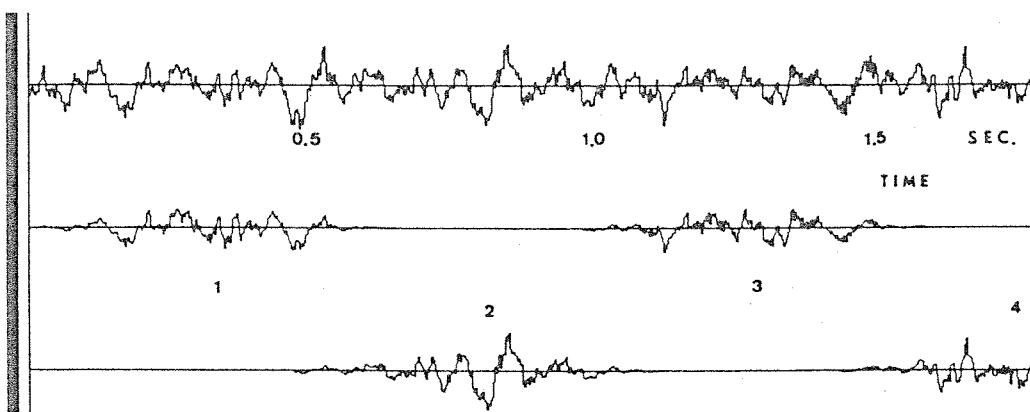


FIGURE 6. EMG signal segmented with 25% overlapping for WOSA spectral estimation. Upper trace: EMG. Lower trace: hamming windowed overlapping segments.

Each two adjacent segments overlap with D -samples. The I segments cover the given data sequence $\{x_k\}$ such that the last sample of the I 's segment obeys $L + (I - 1)D = N$. An EMG signal, such segmented, is described in Figure 6.

We shall now calculate the periodograms of the I overlapped segments, each multiplied by a data window $w(k)$. Denote the normalized periodogram of the i th segment by $\hat{S}^i(w)$, hence:

$$\hat{S}^i(m) = \frac{\Delta t}{E_w L} \left| \sum_{k=0}^{L-1} x_k^i w(k) \exp(-j2\pi mk/L) \right|^2$$

$$i = 1, 2, \dots, I \quad (8.28)$$

Where the normalization factor E_w is the average power of the window

$$E_w = \frac{1}{L} \sum_{k=0}^{L-1} w^2(k) \quad (8.29)$$

The WOSA spectral estimate is the average of all I normalized periodograms

$$\hat{S}(m) = \frac{1}{I} \sum_{i=1}^I \hat{S}^i(m) \quad (8.30)$$

The expectation of the estimator is similar in nature to the one calculated for the periodogram (Equation 8.24). The variance, however, is improved.

Define the normalized covariance, $\rho(j)$, between two normalized periodograms such that

$$\rho(j) = \text{Covar} \{ \hat{S}^i(w), \hat{S}^{i+j}(w) \} / \rho(0) \quad (8.31)$$

Then it can be shown³⁰ that the variance of the WOSA estimate is

$$\text{Var} \{ \hat{S}(m) \} = \frac{\text{Var} \{ \hat{S}^i(m) \}}{I} \left(1 + 2 \sum_{i=1}^{I-1} \frac{I-i}{I} \rho(i) \right) \quad (8.32)$$

Note that for $D > L$ such that the covariance (Equation 8.31) approaches zero we get $\rho(i) \approx 0$ for all $i > 0$. Hence the improvement in the variance (over a periodogram with L samples) is I times. Nonoverlapping segmentation, therefore, should be employed if N is large enough. If the total number of samples, N , is not large, it is recommended³⁰ to overlap the segments by one half of their length ($D = L/2$).

Several attempts have been made recently^{22-23,31} to provide a general framework for PSD estimation. It has been argued that both the Blackman-Tukey and the WOSA estimators are special cases of a general estimator. Figures 7 and 8 show an example of the PSD estimation by means of WOSA. The reader is referred to Figures 1, 2, 4, and 5 for comparison.

5. Smoothing the Periodogram

An alternative method for reducing the variance of the periodogram is by smoothing.¹⁰ Given a single periodogram, $\hat{S}(w)$ (Equation 8.15) calculated from all available data, we can smooth it by passing it through an appropriate spectral filter (window) $H(w)$. The smoothed estimate $\langle \hat{S}(w) \rangle$ is thus given by the spectral convolution

$$\langle \hat{S}(w) \rangle = \Delta t \int_{-\pi/\Delta t}^{\pi/\Delta t} \hat{S}(\eta) H(w - \eta) d\eta \quad (8.33)$$

Assume, for example, a rectangular window of frequency width of $2B$

$$H(w - \eta) = \begin{cases} \frac{1}{2B}; & |w - \eta| < B \\ 0; & \text{otherwise} \end{cases} \quad (8.34)$$

and the smoothed periodogram

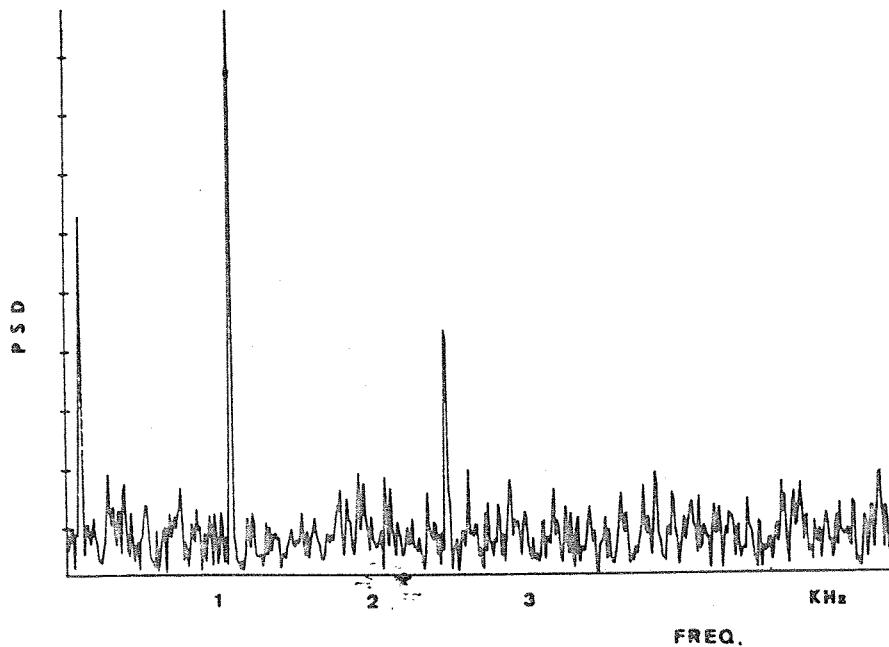


FIGURE 7A. Power spectral density function estimation by means of WOSA with 25% overlapping.
Synthesized sinusoidals signal as in Figure 1. Averaging over one segment of 1024 samples.

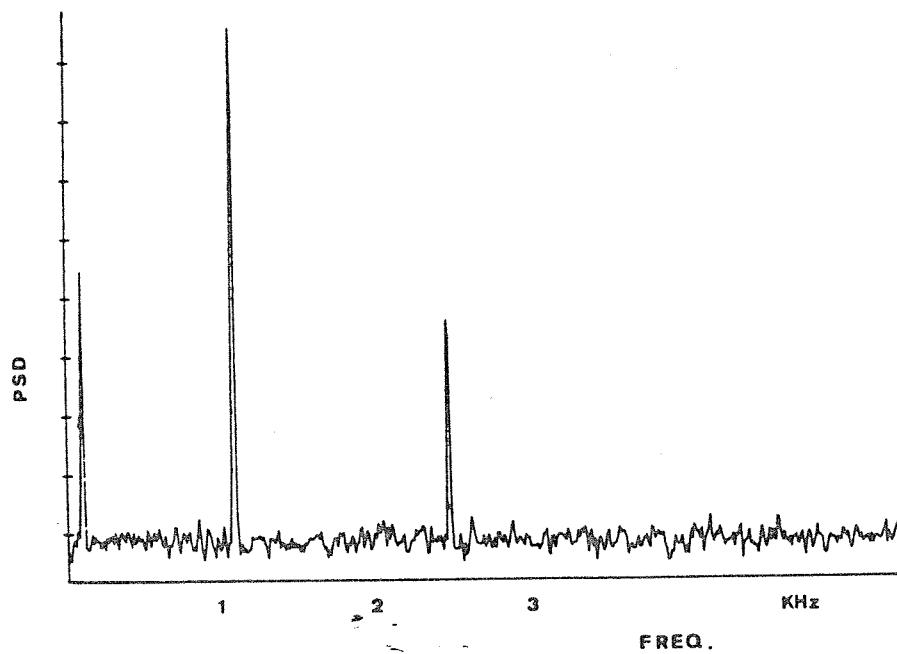


FIGURE 7B. Averaging over seven segments of 1024 samples each.

$$\langle \hat{S}(w_i) \rangle = \frac{1}{2B} \int_{w_i-B}^{w_i+B} \hat{S}(\eta) d\eta \quad (8.35)$$

In the discrete case where $B = L\Delta f$ we get

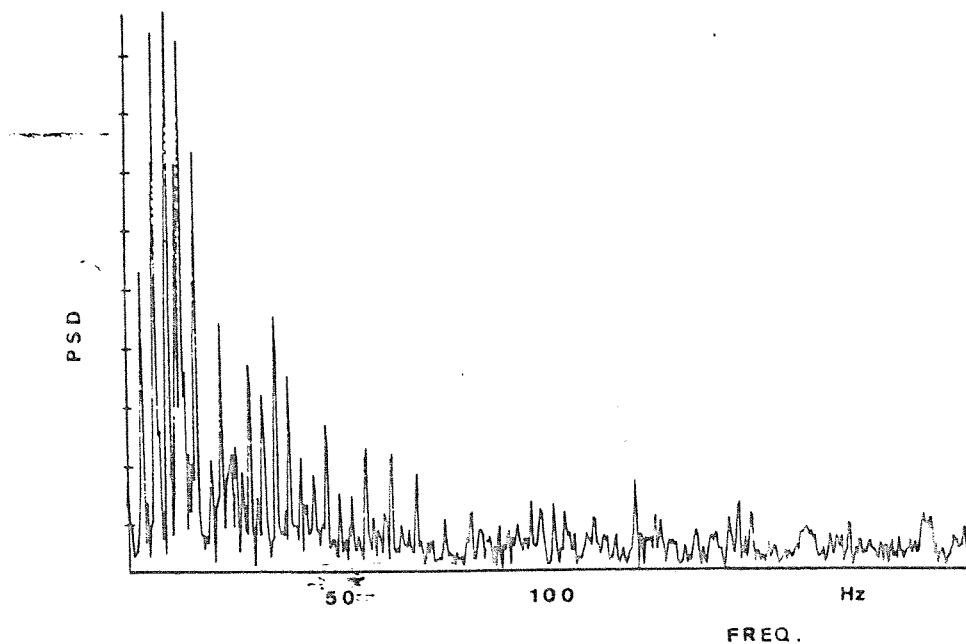


FIGURE 8A. Power spectral density function estimation by means of WOSA with 25% overlapping. EMG signal as in Figure 2. Averaging over one segment of 1024 samples.

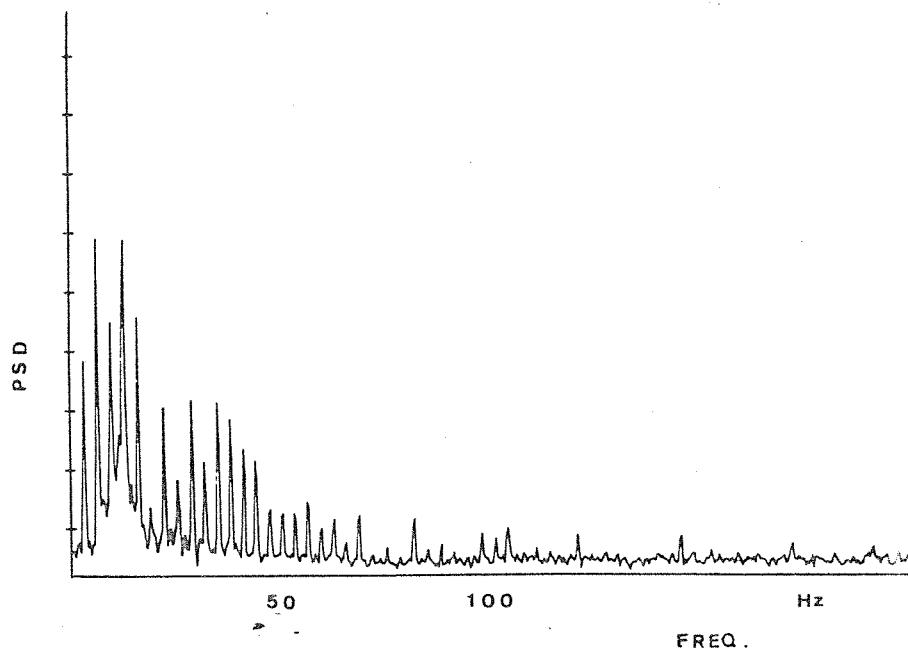


FIGURE 8B. Averaging over seven segments of 1024 samples each.

$$\langle \hat{S}(w_i) \rangle = \frac{1}{2L + 1} \sum_{j=i-L}^{i+L} \hat{S}(w_j) \quad (8.36)$$

which is the average value of $2L + 1$ adjacent values.

The bias of the smoothed periodogram is seen from (Equation 8.33) to be

$$E\{\langle \hat{S}(w) \rangle\} = \int_{-\pi/\Delta t}^{\pi/\Delta t} E\{\hat{S}(\eta)\} H(w - \eta) d\eta \quad (8.37)$$

where the expectation term in the integrant is given by Equation 8.20. For very large N we get

$$E\{\langle \hat{S}(w) \rangle\} \approx \Delta t \int_{-\pi/\Delta t}^{\pi/\Delta t} S(\eta) H(w - \eta) d\eta \quad (8.38)$$

Hence, the smoothing operation has introduced a bias, even for large N.

Triangular smoothing window has often been used. This window results in less bias than the rectangular one since it gives less weight to remote periodogram samples. However, the variance due to the triangular is larger than achieved by the rectangular one.

III. MAXIMUM ENTROPY METHOD (MEM) AND THE AR METHOD

The MEM is also known as the Maximum Entropy Spectral Estimation (MESE)¹⁵ and the Maximum Entropy Spectrum Analysis, MESA.³³ The method was originally proposed by Burg³⁴ for geophysical applications and was since dealt with by many researchers.^{20,32-37} Several special problems have been investigated.³⁸⁻⁴⁰ More recently, attention has been placed on the application of the method to multidimensional PSD estimation.⁴¹⁻⁴³ Modifications and extensions to the original method have also been suggested.⁴⁴⁻⁴⁵

In the one-dimensional case, when consecutive correlation values are available, the MEM method is identical to the AR PSD estimation method.⁴⁶ These two methods will be discussed jointly in this section.

The MEM PSD estimation can be posed as follows: Given $p + 1$ consecutive estimates of the correlation coefficients of the process $\{x(t)\}$, $r_x(i)$, $i = 0, 1, \dots, p$, estimate the PSD of the process. Clearly what is needed for the estimation are the unknown correlation coefficients $r_x(i)$; $i > p$. Burg³⁴ has suggested these autocorrelation coefficients be extrapolated so that the time series characterized by the correlation has maximum entropy. Out of all time series having the $p + 1$ given first autocorrelation coefficient, the time series that yields maximum entropy will be the most random one or, in other words, the estimated PSD will be the flattest among all PSDs having the given $p + 1$ correlation coefficients.

The entropy is a measure of the amount of information (or "ignorance") we have on a process. Consider a discrete process with J states each having the probability of occurrence p_j , $j = 1, 2, \dots, J$. Assume that initially no knowledge is available on the probabilities p_j . The information we possess, on the system, is limited. If now we are given the value of a certain p_i , we have gained a certain amount of information on the system or our state of ignorance has been reduced. The amount of information added, ΔI , is defined (in "bits") by

$$\Delta I = \log_2 \frac{1}{p_i} \quad (8.39)$$

The average information per time interval, H , is known as the entropy, and is given from Equation 8.39 by

$$H = \sum_{i=1}^J p_i \log_2 \frac{1}{p_i} = - \sum_{i=1}^J p_i \log_2 p_i \quad (8.40)$$

In the deterministic case, where we know that event i has happened with probability one,

all other p_j 's are zero and the entropy is zero. In the random case, the entropy is always positive. The entropy is thus a measure for the randomness of the system.

In the continuous case the relative entropy is given by

$$H = - \int_{-\infty}^{\infty} p(s) \log_2 p(s) ds \quad (8.41)$$

where $p(s)$ is the amplitude distribution of the signal $s(t)$. We would like to write down a relation between the estimated spectrum of the signal $\hat{S}(w)$ and the entropy. Assume the signal $s(t)$ was generated by passing a white signal through a linear filter having a transfer function $S(w)$. It can be shown⁴⁷ that the difference in entropies, ΔH , between that of the signal $s(t)$ and the input is

$$\Delta H = \int_{-\infty}^{\infty} \log_2 \hat{S}(w) dw \quad (8.42)$$

Since there are an infinite number of signals with white spectrum, the exact input is unknown. However, we know that we want to maximize the entropy (Equation 8.42) subject to the constraints that

$$r(n) = \int_{-\infty}^{\infty} \hat{S}(w) \exp(j2\pi w \Delta t n) dw \\ n = 0, 1, \dots, P \quad (8.43)$$

This constraint maximization will ensure that the estimated spectrum is the spectrum of a process having the flatness spectrum of all processes with the given $P + 1$ correlation coefficients.

The maximization of Equation 8.42 with the constraints (Equation 8.43) can be solved using the Lagrange multiplier technique. The result is found to be

$$\hat{S}(w) = \frac{\sigma_p^2 \Delta t}{|1 + \sum_{i=1}^p a_i \exp(-j2\pi i w)|^2} \quad (8.44)$$

where σ_p and a_i $i = 1, \dots, P$ are determined from the data. It has been shown⁴⁸⁻⁴⁹ that the MEM PSD estimation (Equation 8.44) is identical with the estimation of PSD of an AR model.⁴⁶

Recall that the AR model (Equation 7.13) presents a random sampled signal x_k as the output of a filter $H(z) = A^{-1}(z)$ with white sequence as an input. The estimated PSD of x_k is thus

$$\hat{S}(w) = |H(jw)|^2 = \left| \frac{1}{A(jw)} \right|^2 = \frac{G^2}{|1 + \sum_{i=1}^p \hat{a}_i \exp(-j2\pi i w)|^2} \quad (8.45)$$

where G^2 equals the mean square error, E_p , and is given in terms of the correlation coefficients by Equation 7.35. The estimated AR coefficients \hat{a}_i are given by Equation 7.30 and the order, p , by one of the methods discussed in Chapter 7.

Equation 8.45 can be rewritten⁴⁶ as

$$\hat{S}(w) = \frac{r(0) + \sum_{i=1}^p \hat{a}_i r(i)}{p(0) + 2 \sum_{i=1}^p p(i) \cos(iw)} \quad (8.46)$$

with

$$p(i) = \sum_{k=0}^{p-i} \hat{a}_k \hat{a}_{k+i}, \quad \hat{a}_0 = 1$$

$$i = 0, 1, \dots, p \quad (8.47)$$

Note that in the Blackman-Tukey PSD estimator (Equation 8.6) the power spectrum is estimated by means of the $p + 1$ first correlation coefficients, calculated from the data while the correlation coefficients $r(i)$, $i > p$ are assumed to be zero. In the MEM and AR PSD estimation, the first $p + 1$ correlation coefficients are calculated from the data and are identical to the ones used in the Blackman-Tukey algorithm. The coefficients $r(i)$, $i > p$ are not assumed zero but are calculated by the maximization of Equation 8.42 with the constraint (Equation 8.43). This yields the same result as the AR model where

$$r(i) = - \sum_{k=1}^p \hat{a}_k r(i-k), \quad i > p \quad (8.48)$$

The statistical properties of the MEM and AR spectral estimators have been investigated by several authors.⁵⁰⁻⁵³ It has been demonstrated that the variance of the estimators is inversely proportional to the data length.

The MEM estimator has been experimentally compared⁵⁴⁻⁵⁵ with the WOSA estimator (Equation 8.30). The MEM estimator has been found to have superior frequency resolution capability, especially for short records and to have larger dynamic range. Hung and Herring,⁵⁵ however, report that when considering the detection of sinusoid in additive white noise, the DFT based detector consistently provided higher signal detection probability and a more accurate estimate of signal frequency than the MEM. Dyson and Rao⁵³ have concluded that the MEM methods show promise of achieving the detection performance of long observation interval DFT analysis, at a reduced observation time, for useful SNR range.

The MEM PSD estimator is known to yield incorrect results for sinusoidal signals in additive white noise, a phenomenon sometimes called *line splitting*.^{39,54,56} Line splitting is the occurrence of two or more closely spaced peaks in the estimated PSD, where only one should be present. Line splitting is most likely to occur when the SNR is high, the initial phase is some odd multiple of 45° , the time duration is such that sine components have an odd number of quarter cycles, and the number of AR coefficients (the order of the model) is a large percentage of the number of data samples.

The correct model (with white noise as an input) for describing an N-pole complex sinusoidal signal in additive white noise is an N-pole and N-zero system with equal gain weights for its pole and zero parts. When an AR model is forced to describe such a signal, infinite number of poles are required. The use of finite numbers of poles is the source for the line splitting and line shifting inaccuracies. Several methods have been suggested to overcome the line splitting problem.⁵⁶⁻⁵⁸

A second problem associated with the MEM method is that of the bias in the positioning of the spectral peaks with respect to the true frequency of the peaks. This shift is sometimes known as the frequency estimation bias. Swingle⁶⁰ has shown that this bias can be of the

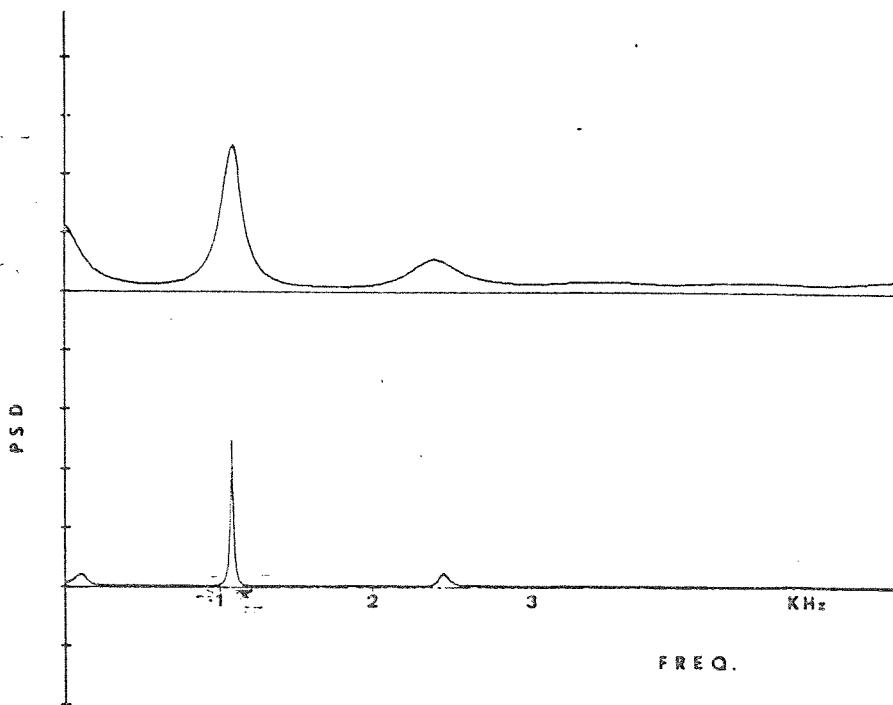


FIGURE 9. Power spectral density function estimation by the MEM method. Synthesized sinusoids as in Figure 1. Upper trace: AR model order is 10. Lower trace: AR model order is 40.

Yet another problem exists when estimating the PSD of sinusoids in noise with MEM. It has been shown⁵⁹ that the peak amplitudes in the MEM are not linearly proportional to the power. In high SNR the peak is proportional to the square of the power.

Recently, many modifications and improvements for the MEM have been suggested for AR and multivariate AR spectral estimation (see, for example, References 61 through 63).

Experimental results with MEM PSD estimation are shown in Figures 9 and 10.

IV. THE MOVING AVERAGE (MA) METHOD

The process, its PSD is to be estimated, can be modeled by an MA(q) model (Equation 7.15). Its estimated PSD is given by

$$\hat{S}(w) = |B(\exp(-jw\Delta t))|^2 S_n(w) \quad (8.49)$$

where $S_n(w)$ is the PSD of the input white noise. The coefficients b_j , $j = 1, 2, \dots, q$ of the MA model can be estimated as discussed in Chapter 7, Section III. In terms of these coefficients, the estimator becomes:

$$\hat{S}(w) = \left| \sum_{i=0}^q \hat{b}_i \exp(-jwi\Delta t) \right|^2 = \sum_{n=0}^q \sum_{m=0}^q \hat{b}_m \hat{b}_n \exp(-j(m-n)w\Delta t) \quad (8.50)$$

The last equation can be rewritten so that

$$\hat{S}(w) = \sum_{n=-q}^q r(n) \exp(-jwn\Delta t)$$

order of 16% of the frequency resolution ($1/N\Delta t$). Methods for overcoming this problem were also suggested.⁵⁸

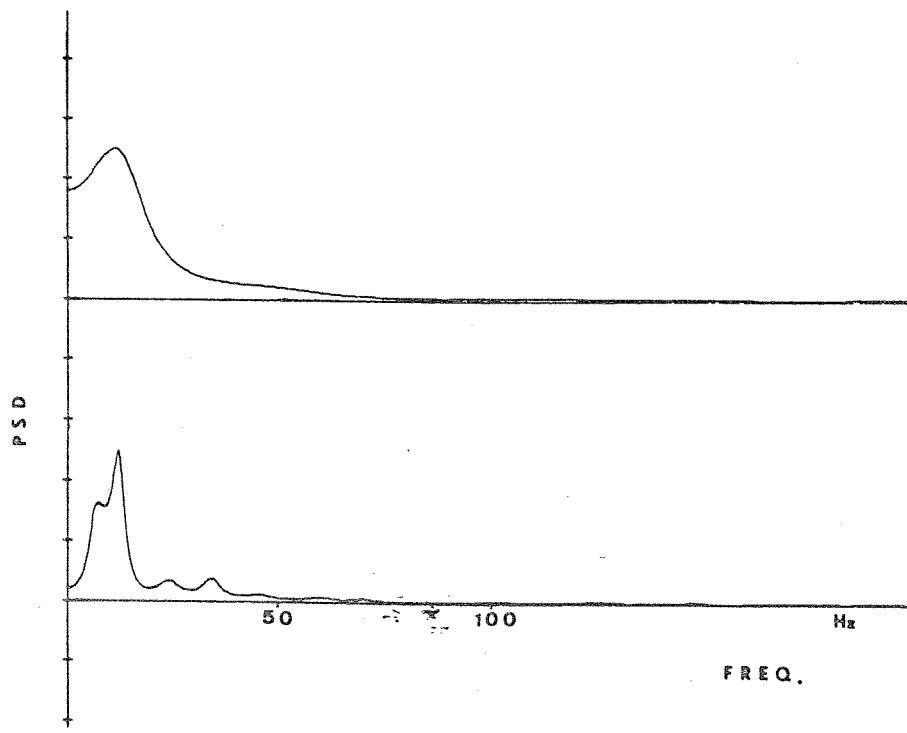


FIGURE 10. Power spectral density function estimation by the MEM method. EMG signal as in Figure 2. Traces are as in Figure 9.

with

$$r(n) = \sum_{k=0}^q \hat{b}_k \hat{b}_{k-n}; \quad -q \leq n \leq q$$

and

$$\hat{b}_k = 0 \text{ for } k > q \text{ and } k < 0 \quad (8.51)$$

It can be easily shown that the quantities $r(n)$ in Equation 8.51 are the autocorrelation coefficients of the MA model.

Hence, the MA spectral estimator (Equation 8.51) is the same as the Blackman-Tukey estimator (Equation 8.6). The Blackman-Tukey spectral estimator can be thus considered a special case of the MA spectral estimator. The MA estimator will be effective when the spectra to be estimated contain sharply defined notches and do not contain sharply defined peaks.

V. AUTOREGRESSIVE MOVING AVERAGE (ARMA) METHODS

A. The General Case

In most applications, the power spectrum to be estimated contains both notches and peaks. The more general ARMA (p,q) rational model has obvious advantages over the AR(p) and MA(q) models discussed in previous sections. Although the estimation of the ARMA coefficients (Chapter 7, Section IV) requires a lot of computation effort, it has become an important method due to its superiority in the PSD function estimation. Several algorithms

had been suggested for ARMA (p,q) spectral estimation (e.g., References 64 to 70). The statistical properties of the ARMA (p,q) estimator have been investigated⁷¹⁻⁷³ and some preprocessing techniques for the improvement of estimation have been suggested.⁷⁴ Cadzow, in a comprehensive paper,⁷⁵ has presented various methods for AR, MA, and ARMA power spectral estimation.

Once the order p,q and the coefficients of the ARMA model had been estimated, the PSD estimate is obtained by

$$\hat{S}(w) = |H(\exp(-jw\Delta t))|^2 S_n(w) = \frac{\sigma_n^2 \Delta t |1 + \sum_{i=1}^q \hat{b}_i \exp(-jwi\Delta t)|^2}{|1 + \sum_{i=1}^p \hat{a}_i \exp(-jwi\Delta t)|^2} \quad (8.52)$$

where $H(jw)$ is given in Equation 7.44 and $S_n(w)$ is the PSD function of the input white noise having a variance σ_n^2 . The order of the ARMA model p,q and its estimation are discussed in Chapter 7, Section IV.

Cadzow⁷⁵ has shown that when the ARMA coefficients are evaluated by the overdetermined rational model equation approach, the resultant PSD estimate is less sensitive to the coefficients estimates. Cadzow's approach calls for the determination of the coefficients from an overdetermined set of the Yule-Walker equation (see Chapter 7, Section II), for example by means of the singular valued decomposition (SVD) technique (for SVD analysis, see Chapter 3, Volume II).

B. Pisarenko's Harmonic Decomposition (PHD)

In many applications, the signal (its PSD is to be estimated) can be considered as several sinusoids in additive noise. This may be the case for special types of EEG, gastrointestinal signals, breathing signals, infants cry, and many other biomedical signals. This type of a signal is also widely used as a test signal for system identification. Pisarenko⁷⁶⁻⁷⁷ has suggested an estimator for the PSD function of a signal consisting of p/2 sinusoids in additive white noise

$$x_n = \sum_{i=1}^{p/2} A_i \sin(W_i n \Delta t + \phi_i) + n_n \quad (8.53)$$

where A_i and ϕ_i , $i = 1, \dots, p/2$ are the amplitudes and phases of the sinusoids and $\{n_n\}$ is a sequence from the white noise process having zero mean and a variance of σ_n^2 . The noise is uncorrelated with the sinusoids.

Using the trigonometric identity:

$$\text{Sin}(wn\Delta t) = 2 \cos(w\Delta t) \sin(w\Delta t(n-1)) - \sin(w\Delta t(n-2)) \quad (8.54)$$

and letting $\tilde{x}_n = \text{Sin}(wn\Delta t)$, we get the difference equation:

$$\tilde{x}_n = (2 \cos(w\Delta t)) \tilde{x}_{n-1} - \tilde{x}_{n-2} \quad (8.55)$$

Hence, the samples of a deterministic sinusoid can be described by means of a second order AR(2) equation.

In general, for the deterministic summation of p/2 sinusoids,¹⁵ the resultant difference equation is an AR(p) equation

$$\tilde{x}_n = - \sum_{m=1}^p a_m \tilde{x}_{n-m} \quad (8.56)$$

Transferring Equation 8.56 into the Z domain yields the characteristic equation

$$\left(1 + \sum_{m=1}^p a_m Z^{-m} \right) = 0 = \sum_{j=1}^{p/2} (Z - Z_j)(Z - Z_j^*) \quad (8.57)$$

There are p roots to the characteristic Equation 8.57 arranged in conjugate pairs

$$z_k = \exp(jw_k \Delta t) \text{ and } z_k^* = \exp(-jw_k \Delta t)$$

$k = 1, 2, \dots, p/2$. The roots are all located on the unit circle where the frequencies w_k are the frequencies of the sinusoids present in the signal.

Returning now to the noisy case (Equation 8.53), we get

$$\tilde{x}_n \rightarrow \tilde{x}_n + n_n = - \sum_{m=1}^p a_m \tilde{x}_{n-m} + n_n \quad (8.58)$$

Defining $a_0 = 1$, the last equation can be rewritten as

$$x_n = - \sum_{m=1}^p a_m x_{n-m} + \sum_{m=0}^p a_m n_{n-m} \quad (8.59)$$

Equation 8.59 states that the signal represented by Equation 8.53 is a special ARMA(p,p) process. In this process, the AR(p) and MA(p) coefficients are identical. Due to this property, the identification of this special ARMA(p,p) process is less complicated than the general case. Techniques, simpler than the ones discussed in Chapter 7, Section IV can be applied; one method is presented here.

The ARMA (p,p) Equation 8.59 can be written in a matrix form. Define

$$\begin{aligned} \underline{x}^T &= [x_n, x_{n-1}, \dots, x_{n-p}] \\ \underline{a}^T &= [1, a_1, a_2, \dots, a_p] \\ \underline{n}^T &= [n_n, n_{n-1}, \dots, n_{n-p}] \end{aligned} \quad (8.60)$$

Introducing Equation 8.60 into Equation 8.59 yields

$$\underline{x}^T \underline{a} = \underline{n}^T \underline{a} \quad (8.61)$$

Premultiplying both sides of Equation 8.61 by \underline{x} and taking the expectation we get

$$E\{\underline{x} \underline{x}^T\} \underline{a} = E\{\underline{x} \underline{n}^T\} \underline{a} \quad (8.62)$$

noting that (because of the assumptions made on the noise) the cross correlation between the noisy observation \underline{x} and the noise \underline{n} is

$$E\{\underline{x} \underline{n}^T\} = \sigma_n^2 I \quad (8.63)$$

where I is the identity matrix, and the signal's $(p+1) \times (p+1)$ autocorrelation matrix R_x

$$E\{\underline{x} \underline{x}^T\} = R_x = \begin{bmatrix} r_x(0) & r_x(-1), \dots, r_x(-p) \\ \vdots & \vdots \\ \vdots & \vdots \\ r_x(p), \dots, & r_x(0) \end{bmatrix} \quad (8.64)$$

Hence, we get

$$R_x \underline{a} = \sigma_n^2 \underline{a} \quad (8.65)$$

Equation 8.65 states that the coefficient vector \underline{a} is an eigenvector of the correlation matrix R_x and the noise variance is the corresponding eigenvalue. The eigenvector must be scaled such that its first component equals one.

It can be shown¹⁵ that the eigenvalue σ_n^2 is the minimum eigenvalue of the correlation matrix with the correct dimension $(p+1)(p+1)$. In the overdetermined case, where the correlation matrix is generated by more than $(p+1)$ lags, the minimum eigenvalue is repeated.

The autocorrelation of the signal x_n given by Equation 8.53 is

$$r_x(0) = \sigma_n^2 + \frac{1}{2} \sum_{i=1}^{p/2} A_i^2 \quad (8.66A)$$

$$r_x(k) = \frac{1}{2} \sum_{i=1}^{p/2} A_i^2 \cos(w_i k \Delta t), \quad k \neq 0 \quad (8.66B)$$

Assuming the frequencies $w_i, i = 1, 2, \dots, p/2$ and the correlation coefficients $r_x(k), k = 0, 1, \dots$ are known, the sinusoids power $A_i^2/2$ can be determined. Define the power vector

$$\underline{p}^T = [A_1^2/2, A_2^2/2, \dots, A_p^2/2] \quad (8.67)$$

the correlation vector

$$\underline{r}_x^T = [r_x(1), r_x(2), \dots, r_x(p)] \quad (8.68)$$

and the cosine matrix

$$C = \begin{bmatrix} \cos(w_1 \Delta t) & \cos(w_2 \Delta t) & \dots & \cos(w_p \Delta t) \\ \vdots & \vdots & & \vdots \\ \cos(w_1 p \Delta t) & \dots & \dots & \cos(w_p p \Delta t) \end{bmatrix} \quad (8.69)$$

The power coefficients are given by introducing Equations 8.67, 8.68, and 8.69 into Equation 8.66B

$$\underline{p} = C^{-1} \underline{r}_x \quad (8.70)$$

Pisarenko's PSD estimator can now be formulated in an algorithmic form:

1. Estimate the order, p , of the model, which is twice the number of sinusoids present in the signal;

2. Estimate, for the data, $p + 1$ terms of the autocorrelation function using the biased estimator (Equation 7.27);
3. Solve the eigenvector equation (Equation 8.65);
4. Repeat steps 2 and 3 with increasing order until the minimal eigenvalue remains unchanged;
5. The order, p , the noise variance (minimal eigenvalue) are determined. The vector \underline{a} is taken to be the eigenvector corresponding to the minimal eigenvalue;
6. Solve Equation 8.57 to get the roots and the frequencies, w_k ;
7. Solve Equation 8.70 to get the power of the various sinusoids;
8. Solve Equation 8.66A to get the noise power.

An efficient method for solving the eigenvector equation is discussed in Chapter 3, Volume II. When *a priori* knowledge about the signal exists, stating it consists of sinusoids in additive noise, Pisarenko's method has the advantage of providing a PSD estimate with δ functions. Other methods such as the AR spectral estimator will "smear" the spectrum. However the order is not known exactly. It may be estimated too high: then spurious components may be introduced to the PSD estimation; or too low: then the spectral components will usually appear at incorrect frequencies. Another inaccuracy source of the method is the fact that the autocorrelation coefficients are estimated by means of the biased estimator. This is done in order to ensure that the autocorrelation matrix is positive definite. The bias estimation, however, causes inaccuracies in both frequency and power estimation.

The technique had been extended⁷⁸ to include the case of colored additive noise.

C. Prony's Method

Prony's method^{15,79-82} is mainly applied for transients analysis. It has been extended, however, to provide PSD function estimation.

Assume the sequence \tilde{x}_n is the samples of a signal composed of damped (complex) sinusoids

$$\begin{aligned} \tilde{x}_n = & \sum_{m=1}^p A_m \exp(\alpha_m n \Delta t) \exp(j(w_m n \Delta t + \phi_m)) \\ n = & 0, 1, \dots, N - 1 \end{aligned} \quad (8.71)$$

Equation 8.71 describes the signal \tilde{x}_n as a sum of p sinusoids with frequencies w_m , phase ϕ_m , and with amplitudes A_m , $m = 1, 2, \dots, p$, exponentially decaying, with rate α_m ($\alpha_m < 0$). For \tilde{x}_n to be real, it is required that the roots of the characteristic equation be complex conjugate pairs of the type $\exp(j(w_m n \Delta t + \phi_m))$ and $\exp(-j(w_m n \Delta t + \phi_m))$. The energy spectral distribution function of Equation 8.71 is given by

$$\hat{S}(w) = |X(w)|^2 = \left| \sum_{m=1}^p A_m \exp(j\phi_m) \frac{2\alpha_m}{[\alpha_m^2 + (w - w_m)^2]} \right|^2 \quad (8.72)$$

To use Equation 8.72 as "spectral" estimator, the parameters p , A_m , α_m , ϕ_m , and w_m $m = 1, 2, \dots, p$ must be identified. In order to do that, rewrite Equation 8.71 as

$$\begin{aligned} \tilde{x}_n = & \sum_{m=1}^p b_m Z_m^n \quad n = 0, 1, \dots, N - 1 \\ b_m = & A_m \exp(j\phi_m) \\ Z_m = & \exp((\alpha_m + jw_m)\Delta t) \end{aligned} \quad (8.73)$$

The last equation is the homogeneous solution to a constant coefficient linear difference equation¹⁵

$$\tilde{x}_n = - \sum_{m=1}^p a_m \tilde{x}_{n-m} \quad n = p, p+1, \dots, N-1 \quad (8.74)$$

Transferring the last equation into the Z domain yields

$$X_{(z)} \left(1 + \sum_{m=1}^p a_m z^{-m} \right) = 0 \quad (8.75)$$

having the roots Z_k at

$$\prod_{k=1}^p (z - Z_k) = 1 + \sum_{m=1}^p a_m z^{-m} = 0 \quad (8.76)$$

The roots Z_k given by the solution of Equation 8.76 are the exponents of Equation 8.71.

In the more practical case, the signal is noisy, hence the observation x_n can be described by the Prony's model as

$$x_n = \tilde{x}_n + n_n \quad n = 0, 1, \dots, N-1 \quad (8.77)$$

where n_n is a sequence of white noise with zero mean and variance σ_n^2 . Introducing Equation 8.77 into Equation 8.74 yields

$$x_n = - \sum_{m=1}^p a_m x_{n-m} + \sum_{m=0}^p a_m n_{n-m} \quad (8.78)$$

The signal is thus modeled as a ARMA (P,P) model with equal coefficients for the AR(p) and MA(p) parts. Unlike the Pisarenko model, the roots of the characteristic Equation 8.78 are not restricted to the unit circle. The model describes, in general, decaying sinusoids rather than pure sinusoids.

The ARMA coefficients of (Equation 8.78) can be solved by the methods discussed in Chapter 7, Section IV. Note that only the AR (or MA) part has to be identified.

Once the a_i have been identified, Equation 8.76 can be solved to provide the roots Z_k . The coefficients required for the estimator Equation 8.71 are computed from the solution of the set of linear equations (Equation 8.73). Define the model vector

$$\underline{\tilde{x}}^T = [\tilde{x}_0, \tilde{x}_1, \dots, \tilde{x}_{N-1}] \quad (8.79)$$

with the observation vector \underline{x} defined similarly. Define the complex coefficient vector

$$\underline{b}^T = [b_1, b_2, \dots, b_p] \quad (8.80)$$

and the matrix Z

$$Z = \begin{bmatrix} 1 & 1 & \dots & 1 \\ Z_1 & Z_2 & & Z_p \\ \vdots & & & \vdots \\ Z_1^{N-1} & \dots & \dots & Z_p^{N-1} \end{bmatrix} \quad (8.81)$$

The set of N equations (8.73) with the unknown vector \underline{b} can be written in a matrix form

$$\underline{\tilde{x}} = Z\underline{b} \quad (8.82)$$

Recall that $\underline{\tilde{x}}$ is the model of the observations \underline{x} . Our aim is to choose the model parameters such that the model will best fit (in some sense) the observations. Assume we want a least squares minimization

$$\min_{\underline{b}} \sum_{n=0}^{N-1} (x_n - \tilde{x}_n)^2 = \min_{\underline{b}} (\underline{x} - \underline{\tilde{x}})^T (\underline{x} - \underline{\tilde{x}}) \quad (8.83)$$

Introducing Equation 8.82 into Equation 8.83 and performing the minimization yields the well-known least squares estimate of \underline{b}

$$\hat{\underline{b}} = [(Z^*)^T Z]^{-1} (Z^*)^T \underline{x} \quad (8.84)$$

where an asterisk (*) denotes the conjugate.

With the Z_k and b_k $k = 1, 2, \dots, p$ estimated, the values needed for the energy spectral distribution estimator (Equation 8.71) can be solved as follows:

$$\begin{aligned} A_m &= |b_m| \\ \phi_m &= \operatorname{tg}^{-1}(I_m(b_m)/R_e(b_m)) \\ \alpha_m &= (1/\Delta t) \ell n |Z_m| \\ w_m &= (1/\Delta t) \operatorname{tg}^{-1}(I_m(Z_m)/R_e(Z_m)) \\ m &= 1, 2, \dots, p \end{aligned} \quad (8.85)$$

In summary, the Prony's method for energy spectral density estimation is given by the following steps:

1. Estimate order of model;
2. Estimate the AR coefficients of the model (Equation 8.78) by one of the methods discussed in Chapter 7;
3. Determine the roots Z_k from the solution of Equation 8.76;
4. Estimate the b_k coefficients by the least squares Equation 8.84;
5. Compute the parameters A_m , α_m , w_m by means of Equations 8.85;
6. Estimate the energy power density function, Equation 8.72.

VI. MAXIMUM LIKELIHOOD METHOD (MLM) — CAPON'S SPECTRAL ESTIMATION

A method for PSD function estimation^{59,83-84} based on the idea of measuring the power out of a set of narrow band (optimal) filters has been suggested. The method has been termed the maximum likelihood method (MLM) or the maximum likelihood spectral estimation (MLSE). It has been argued¹⁵ that the method is not truly a maximum likelihood estimate of the PSD, thus the name capon spectral estimate⁸³ is more appropriate.

We have seen that the effect of unavoidable windowing of the data (Figure 3) is to distort the PSD estimation. The sidelobes of the window cause "leakage" from neighboring frequencies into the estimate of the frequency of interest.

Suppose that for each frequency of interest, prior to the PSD estimation, we shall filter the data by means of an optimal filter in such a way that contribution from other frequencies be minimized. This can be viewed as a set of narrow band pass filters, each optimally designed for the particular frequency. The PSD function estimation is calculated from the power output of these filters.

Suppose that a sequence $\{x_k\}$ is given. The sequence is a finite sequence of samples of $x(t)$ at intervals Δt ; its PSD function estimation is required. To estimate the PSD at a frequency w , we choose to prefilter the data with an MA(N) filter with coefficients to be optimally adjusted. Ideally, we are interested in the output power of an infinity narrow band pass filter at the frequency w . If we allow complex sinusoids, the output of the filter should be of the form $A \exp(jwk\Delta t)$ where A is the amplitude of the sinusoid component from which the PSD is to be derived. Assume that the input to the filter (the observed sequence) is given by

$$x_k = A \exp(jwk\Delta t) + n_k \quad (8.86)$$

where $\{n_k\}$ is the sequence of "noise" appearing due to the leakage. It is our desire to design the MA(N) filter in such a way that the signal $A \exp(jwk\Delta t)$ will pass with no distortion, while the second term in Equation 8.86 be minimized.

Following the arguments presented in Chapter 7, we "predict" the k th value of the output of the MA filter such that

$$\hat{x}_k = \sum_{n=0}^{N-1} b_n x_{k-n} \quad (8.87)$$

We want to find the optimal value of the filter coefficients b_n $n = 0, 1, \dots, N - 1$ such that the variance of \hat{x}_k be minimal, thus optimally reducing the effects of $\{n_k\}$.

The variance of \hat{x}_n is

$$\text{Var} \{\hat{x}_k\} = E \left\{ \left(\sum_{m=0}^{N-1} b_m n((k-m)\Delta t) \right)^2 \right\} \quad (8.88)$$

Equation 8.88 can be written in a matrix form as

$$\text{Var} \{\hat{x}_k\} = (\underline{b}^*)^T \underline{R}_n \underline{b} \quad (8.89)$$

where R_n is the noise autocorrelation matrix and $\underline{b}^T = [b_0, b_1, \dots, b_{N-1}]$.

Note that we want the sinusoidal component to pass the filter undistorted; hence, for the noise-free case we want

$$A \exp(jw k \Delta t) = \sum_{n=0}^{N-1} b_n A \exp(jw(k-n) \Delta t) \quad (8.90)$$

Dividing both sides of Equation 8.90 by the term on the left, we get the constraint

$$1 = (\underline{e}^*)^T \underline{b} \quad (8.91A)$$

$$\underline{e}^T \triangleq [1, \exp(jw \Delta t), \dots, \exp(jw(N-1) \Delta t)] \quad (8.91B)$$

It can be shown that the minimization of Equation 8.89 subject to the constraint (Equation 8.91) yields the optimal filter

$$\underline{b} = R_x^{-1} \underline{e}^* / (\underline{e}^T R_x^{-1} \underline{e}^*) \quad (8.92)$$

where R_x is the observation sequence correlation matrix.

It can also be shown that the power spectrum of the filter, which is the MLM estimate of the process PSD function, is given by

$$\hat{S}(w) = \frac{\Delta t}{\underline{e}^T R_x^{-1} \underline{e}^*} \quad (8.93)$$

Note that R_x has to be inverted only once. The evaluation of the quadratic form (the denominator of Equation 8.93) for each frequency can be done by means of the FFT algorithm. It is easily shown that the quadratic form can be written as a weighted sum of exponentials such that

$$\underline{e}^T R_x^{-1} \underline{e}^* = \sum_{n=-N+1}^{N-1} g(n) \exp(jwn \Delta t) \quad (8.94)$$

Denote the components of the matrix R_x^{-1} by $q_{m\ell}$, $m, \ell = 0, 1, \dots, N-1$. The terms $g(n)$ in Equation 8.94 are the sum of all $q_{m\ell}$ for which $n = m - \ell$. Equation 8.94 can be evaluated by the FFT. For the estimation of N values (frequency samples) of $\hat{S}(w)$, we require the calculation of N FFTs.

It has been shown⁸⁴ that the MLM estimate $\hat{S}_{MLM}(w)$ of Equation 8.93 is related to the AR PSD function estimator $\hat{S}_{AR}(w)$ (Equation 8.45 with model order, p) by

$$\frac{1}{\hat{S}_{MLM}(w)} = \frac{1}{p} \sum_{m=1}^p \frac{1}{\hat{S}_{AR}(w)} \quad (8.95)$$

In Equation 8.95 both estimators were assumed to have the correct autocorrelation matrix of order p .

The MLM has lower resolution as compared to the AR estimator. This can be explained intuitively by Equation 8.95 where it is seen that the high resolution of the p th order AR estimator is reduced by the "inverse" averaging with lower order estimators. The MLM, however, exhibits less variance than the AR estimator.⁵⁰ This can also intuitively be explained by Equation 8.95.

VII. DISCUSSION AND COMPARISON OF SEVERAL METHODS

Only the most commonly used techniques have been presented in this chapter. Some methods, such as the Walsh spectral estimator⁸⁹⁻⁹⁰ and other specific methods, have not been

presented either because of lack of space or due to the fact that they have not been used in biomedical applications. For the same reasons, many interesting approaches⁸⁵⁻⁸⁶ for PSD estimation have not been discussed.

An important topic not discussed in this chapter, but that must be taken into account when estimating the PSD, is that of preprocessing the signal. Care must be taken to low-pass filter the data before sampling so that (almost) band-limited spectrum is to be processed and the sampling rate should be chosen such that no aliasing problems arise. Prewhitening filters are often used (especially when employing an AR estimator). Prewhitening filters have been reported to improve AR estimator results and reduce the window bias in the Blackman-Tukey method. A more general preprocessing filter was suggested by Lagunas-Hernandez et al.⁷⁴ They have reported that the use of their preprocessing method caused the reduction of estimator order and computation complexity. Preprocessing filters, however, requires some *a priori* knowledge about the signal.

Several adaptive and recursive methods for PSD estimation have been suggested; most adapt the coefficients of the appropriate filter (e.g., References 75, 87-88). Adaptive estimation is important when the signal under test is nonstationary as is most often the case with biomedical signals. The most commonly used method to overcome the nonstationarity problem is that of segmenting the signal. In this approach, the signal is segmented into "almost" stationary segments. The PSD function of each segment is then estimated by means of nonadaptive methods. Segmentation may be done *a priori* into predetermined length segments, as is commonly done in speech processing, or by an adaptive algorithm as is done, for example, in EEG processing (see Chapter 7).

It is sometimes required to estimate the PSD function with unequal resolution, when certain regions in the frequency are of specific interest. Methods based both on the FFT⁹² and on AR models¹⁵ have been developed.

The various methods for PSD function estimation discussed here, and others, differ from one another in the assumptions made on the process under test, in their statistical characteristics, and computational complexity. It is therefore not expected that one can define a "universal" criterion by means of which the various methods can be graded. The best method to use heavily depends on the application, the type of signal, the computation facility or accuracy, and time constraints. It is obvious that the estimation method to choose when the computation facility consists of a mini- or microcomputer may be different than the one chosen if a "number crunching" machine (e.g., array processor) is available.

Table 1 presents, in a concise manner, some of the main characteristics of the various methods discussed in the chapter. This table may thus serve as a guideline for choosing an appropriate estimation method for a given problem (see Kay and Marple¹⁵ for more details).

Table 8.1
SEVERAL PSD ESTIMATION METHODS — COMPARISON AND COMMENTS

		Discrete	Continuous	Advantages	Disadvantages	Comments
Method	Model					
Periodogram	Sum of sinusoids	D (8.15)		Computationally efficient; output directly proportional to power	Resolution inversely proportional to observation; leakage, inaccurate with short data	Requires averaging or smoothing
Pisarenko	ARMA with equal MA and AR	D (8.70), (8.66a)		High resolution when number of sinusoids are known	Sensitive to noise; must determine number of sinusoids; computationally inefficient	Requires eigenvalue solutions
Blackman-Tukey (BT)	MA with correlation window	C (8.6)		Computationally efficient for $p \ll N$	Resolution inversely proportional to observation; leakage; negative PSD	Special MA estimator
AR (MEM)	AR	C (8.44), (8.46)		Better resolution than BT or periodogram; no leakage	Stable filter not always guaranteed; bias in peak frequencies	Model order must be estimated
MA	MA	C (8.51)		Better notch estimation	Low resolution; poor peak estimation	Model order must be estimated
ARMA	ARMA	C (8.52)		High resolution, no leakage	Nonlinear estimation	Model order must be estimated
Prony's Method	ARMA with equal AR and MA	C (8.72)		Resolution as AR, no leakage	Requires polynomial rooting	Model order must be determined
Capon's Method (MLM)	Optimal band pass filters	C (8.93)		Variance less than AR	Resolution less than AR	Requires matrix inversion

REFERENCES

1. Boston, J. R., Spectra of auditory brainstem responses and spontaneous EEG, *IEEE Trans. Biomed. Eng.*, 28, 344, 1981.
2. Williams, R. L., Karacan, I., and Hursch, C. Y., *EEG of Human Sleep: Clinical Applications*, John Wiley & Sons, New York, 1974.
3. Butter, L. A., A real time software system on the PDP-11 for two channel EEG spectral analysis during surgery, *Comput. Programs Biomed.*, 6, 1, 1976.
4. Gross, D., Grassino, A., Ross, W. R. D., and Macklem, P. T., Electromyogram pattern of diaphragmatic fatigue, *J. Appl. Physiol.*, 46, 1, 1979.
5. Mezzalama, M., Prinetto, P., and Morra, B., Experiments in automatic classification of laryngeal pathology, *Med. Biol. Eng. Comput.*, 21, 603, 1983.
6. Wiener, N., Generalized harmonic analysis, *Acta Math.*, 55, 117, 1930.
7. Khinchin, A., Ya., Korrelationstheorie der Stationären Stochastischen Prozesse, *Math. Annalen*, 109, 604, 1934.
8. Tukey, J. W. and Blackman, R. B., *The Measurement of Power Spectra From the Point of View of Communications Engineering*, Dover, New York, 1959.
9. Koopmans, L. H., *The Spectral Analysis of Time Series*, Academic Press, New York, 1974.
10. Schwartz, M. and Shaw, L., *Signal Processing Discrete Spectral Analysis Detection and Estimation*, McGraw-Hill, New York, 1975.
11. Childers, D. G., Ed., *Modern Spectrum Analysis*, IEEE Press, New York, 1978.
12. Box, G. E. P. and Jenkins, G. M., *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1970.
13. Haykin, S. S., Ed., *Nonlinear Methods of Spectral Analysis*, Springer-Verlag, New York, 1979.
14. Jenking, G. M. and Watts, D. G., *Spectral Analysis and its Applications*, Holden-Day, San Francisco, 1968.
15. Kay, S. M. and Marple, S. L., Spectrum analysis — a modern perspective, *Proc. IEEE*, 69, 1380, 1981.
16. Spectral estimation, special issue, *Proc. IEEE*, 70(9), 1982.
17. Robinson, E. A., A historic perspective of spectrum estimation, *Proc. IEEE*, 70(9), 885, 1982.
18. Friedlander, B., Lattice methods for spectral estimation, *Proc. IEEE*, 70(9), 990, 1982.
19. McClellan, J. H., Multidimensional spectral estimation, *Proc. IEEE*, 70(9), 1029, 1982.
20. Papoulis, A., Maximum entropy and spectral estimation: a review, *IEEE Trans. Acoust. Speech Signal Process.*, 29, 1176, 1981.
21. Roberts, J. B. G., Moule, G. L., and Parry, G., Design and application of real time spectrum analyser system, *IEE Proc.*, 127(2), 76, 1980.
22. Carter, G. C. and Nuttal, A. H., Analysis of a generalized framework for spectral estimation, *IEE Proc.*, 130(3), 239, 1983.
23. Allen, J. B. and Rabiner, L. R., A unified approach to short time Fourier analysis and synthesis, *Proc. IEEE*, 65(11), 1558, 1977.
24. Linkens, D. A., Short time series spectral analysis of biomedical data, *IEE Proc.*, 129(9), 663, 1982.
25. Makhoul, J., Spectral linear prediction: properties and applications, *IEEE Trans. Acoust. Speech Signal Process.*, 23, 283, 1975.
26. Otens, R. K. and Enochson, L., *Digital Time Series Analysis*, John Wiley & Sons, New York, 1972.
27. Harris, F. J., On the use of windows for harmonic analysis with DFT, *Proc. IEEE*, 66, 51, 1978.
28. Webster, R. J., Leakage regulation in the DFT spectrum, *Proc. IEEE*, 68, 1339, 1980.
29. Markel, J. D., FFT pruning, *IEEE Trans. Audio Electroacoust.*, 19, 305, 1971.
30. Welch, P. D., The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short modified periodograms, *IEEE Trans. Audio Electroacoust.*, 15, 70, 1967.
31. Carter, G. C. and Nuttal, A. H., A brief summary of generalized framework for power spectral estimation, *Signal Process.*, 2, 387, 1980.
32. Morf, M., Vleria, A., Lee, D. T., and Kailath, T., Multichannel maximum entropy spectral estimation, *IEEE Trans. Geosci. Electron.*, 16, 85, 1978.
33. Dyson, T. and Rao, S. S., Equal observation interval comparison of maximum entropy and weighted overlapped segment averaging spectrum estimation techniques, *IEEE Trans. Acoust. Speech Signal Process.*, 29, 919, 1981.
34. Burg, J. P., Maximum Entropy Spectral Analysis, Proc. 37th Annu. Inst. Meeting Soc. Explor. Geophys., Oklahoma City, 1967.
35. Ulrych, T. Y. and Bishop, T. N., Maximum entropy spectral analysis and autoregressive decomposition, *Rev. Geophys. Space Phys.*, 13, 183, 1975.
36. Jaynes, E. T., On the rationale of maximum entropy methods, *Proc. IEEE*, 70, 939, 1982.
37. Lang, S. W. and McClellan, J. H., Frequency estimation with maximum entropy spectral estimators, *IEEE Trans. Acoust. Speech Signal Process.*, 28, 716, 1980.

38. **Theodoridis, S. and Cooper, D. C.**, Application of the maximum entropy spectrum analysis technique to signals with spectral peaks of finite width, *Signal Process.*, 3, 109, 1981.
39. **Herring, R. W.**, The cause of line splitting in Burg maximum entropy spectral analysis, *IEEE Trans. Acoust. Speech Signal Process.*, 28, 692, 1980.
40. **Wu, N.-L.**, An explicit solution and data extension in the maximum entropy method, *IEEE Trans. Acoust. Speech Signal Process.*, 31, 486, 1983.
41. **Lang, S. W. and McClellan, J. H.**, Multidimensional MEM spectral estimation, *IEEE Trans. Acoust. Speech Signal Process.*, 30, 280, 1982.
42. **Malik, N. A. and Lim, J. S.**, Properties of two dimensional maximum entropy power spectrum estimates, *IEEE Trans. Acoust. Speech Signal Process.*, 30, 788, 1982.
43. **McClellan, J. H. and Lang, S. W.**, Duality for multidimensional MEM spectral analysis, *IEE Proc.*, 130(F), 230, 1983.
44. **Newman, W. I.**, Extension to the maximum entropy method, *IEEE Trans. Inf. Theory*, 23, 89, 1977.
45. **Johnson, R. W. and Shore, J. E.**, Minimum cross entropy spectral analysis of multiple signals, *IEEE Trans. Acoust. Speech Signal Process.*, 31, 574, 1983.
46. **Makhoul, J.**, Linear prediction: a tutorial review, *Proc. IEEE*, 63, 561, 1975.
47. **Bartlett, M. S.**, *An Introduction to Stochastic Processes*, 2nd ed., Cambridge University Press, New York, 1966.
48. **Van den Bos**, Alternative interpretation of maximum entropy spectral analysis, *IEEE Trans. Inf. Theory*, 17, 493, 1971.
49. **Grande, J., II, Hamrud, M., and Toll, P.**, A remark on the correspondence between the maximum entropy method and the AR model, *IEEE Trans. Inf. Theory*, 26, 750, 1980.
50. **Baggeroer, A. B.**, Confidence intervals for regression (MEM) spectral estimates, *IEEE Trans. Inf. Theory*, 22, 534, 1976.
51. **Huzii, M.**, On spectral estimate obtained by an AR Model fitting, *Am. Inst. Statist. Math.*, 29, 415, 1977.
52. **Sakai, H.**, Statistical properties of AR spectral analysis, *IEEE Trans. Acoust. Speech Signal Process.*, 27, 402, 1979.
53. **Toomey, J. P.**, High resolution frequency measurement by linear prediction, *IEEE Trans. Aerosp. Electron. Syst.*, 16, 517, 1980.
54. **Fougerre, P. F., Zawalick, E. J., and Radoski, H. R.**, Spontaneous line splitting in maximum entropy power spectrum analysis, *Phys. Earth Planet. Inter.*, 12, 201, 1976.
55. **Hung, E. K. L. and Herring, R. W.**, Simulation experiments to compare the signal detection properties of DFT and MEM spectra, *IEEE Trans. Acoust. Speech Signal Process.*, 29, 1084, 1981.
56. **Kay, S. M. and Marple, S. L., Jr.**, Sources of and remedies for spectral line splitting in AR spectrum analysis, *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 151, 1979.
57. **Fougerre, P. F.**, A solution to the problem of spontaneous line splitting in maximum entropy power spectrum analysis, *J. Geophys. Res.*, 82, 1051, 1977.
58. **Marple, L.**, A new AR spectrum analysis algorithm, *IEEE Trans. Acoust. Speech Signal Process.*, 28, 441, 1980.
59. **Lacoss, R. T.**, Data adaptive spectral analysis method., *Geophysics*, 36, 661, 1971.
60. **Swingler, D. N.**, A comparison between Burg's maximum entropy method and a nonrecursive technique for the spectral analysis of deterministic signals, *J. Geophys. Res.*, 84, 679, 1979.
61. **Scott, P. D. and Nikias, C. L.**, Energy-weighted linear predictive spectral estimation: a new method combining robustness and high resolution, *IEEE Trans. Acoust. Speech Signal Process.*, 30, 287, 1983.
62. **Quirk, M. P. and Liu, B.**, Improving resolution for AR spectral estimation by decimation, *IEEE Trans. Acoust. Speech Signal Process.*, 31, 630, 1983.
63. **Lee, T. S.**, Large sample identification and spectral estimation of noisy multivariate AR processes, *IEEE Trans. Acoust. Speech Signal Process.*, 31, 76, 1983.
64. **Cadzow, J. A.**, High performance spectral estimation — a new ARMA method, *IEEE Trans. Acoust. Speech Signal Process.*, 28, 524, 1980.
65. **Cadzow, J. A.**, ARMA spectral estimation; a model equation error procedure, *IEEE Trans. Geosci. Rem. Sens.*, 19, 24, 1981.
66. **Kay, S. M.**, A new ARMA spectral estimator, *IEEE Trans. Acoust. Speech Signal Process.*, 28, 585, 1980.
67. **Friedlander, B.**, Efficient algorithm for ARMA spectral estimation, *IEE Proc.*, 130 (F.3), 195, 1983.
68. **Friedlander, B.**, Instrumental variable methods for ARMA spectral estimation, *IEEE Trans. Acoust. Speech Signal Process.*, 31, 404, 1983.
69. **Bruzzone, S. and Kaveh, M.**, On some suboptimal ARMA spectral estimators, *IEEE Trans. Acoust. Speech Signal Process.*, 28, 753, 1980.
70. **Porat, B.**, ARMA spectral estimation based on partial autocorrelations, *Circuits. Syst. Signal Process.*, 2(3), 341, 1983.

71. Kaveh, M. and Bruzzone, S. P., Statistical efficiency of correlation based methods for ARMA spectral estimation, *IEE Proc.*, 130 (F-3), 211, 1983.
72. Fischer, J. and Wilfert, H. H., Determination of the statistical errors in the estimation of the power spectrum by means of univariate time series analysis, Proc. 5th IFAC Symp. Ident. and Syst. Param. Estimation, 1235, 1979.
73. Takeuchi, M., Decision of order in ARMA model on power spectrum estimation and accuracy of estimated power spectrum — in fitting ARMA model to time series, *Syst. Comput. Controls*, 12(1), 18, 1981.
74. Lagunas-Hernandez, M. A., Figueiras-Vidal, A. R., Marino-Acebal, J. B., and Vilanova, A. C., A linear transform for spectral estimation, *IEEE Trans. Acoust. Speech Signal Process.*, 29, 989, 1981.
75. Cadzow, J. A., Spectral estimation: an overdetermined rational model equation approach, *Proc. IEEE*, 70, 907, 1982.
76. Pisarenko, V. F., The retrieval of harmonics from a covariance function, *Geophys. J. R. Astron. Soc.*, 33, 247, 1973.
77. Pisarenko, V. F., On the estimation of spectra by means of non linear functions of the covariance matrix, *Geophys. J. R. Astron. Soc.*, 28, 511, 1972.
78. Satorius, E. H. and Alexander, J. T., High Resolution Spectral Analysis of Sinusoids in Correlated Noise, Rec. 1978 ICASSP, Tulsa, Ok., 1978, 349.
79. Bucker, H. P., Comparison of FFT and Prony algorithms for bearing estimation of narrow band signals in a realistic ocean environment, *J. Acoust. Soc. Am.*, 61, 756, 1977.
80. Seahubert, D. H., Application of Prony's method to time domain reflectometer data and equivalent circuits synthesis, *IEEE Trans. Antennas Propag.*, 27, 180, 1979.
81. Weiss, L. and McDonough, R. N., Prony's method, Z. transforms and pade approximation, *SIAM Rev.*, 5, 145, 1963.
82. Kumaresan, R. and Tufts, D. W., Improved spectral resolution III: efficient realization, *Proc. IEEE*, 68, 1354, 1980.
83. Capon, J., High resolution frequency wavenumber spectrum analysis, *Proc. IEEE*, 57, 1408, 1969.
84. Burg, J. P., The relationship between maximum entropy spectra and maximum likelihood spectra, *Geophysics*, 37, 375, 1972.
85. Jain, A. K. and Ranganath, S., Extrapolation algorithms for discrete signals with applications in spectral estimation, *IEEE Trans. Acoust. Speech Signal Process.*, 29, 830, 1981.
86. Beex, A. A. and Scharf, L. L., Covariance sequence approximation for parametric spectrum modeling, *IEEE Trans. Acoust. Speech Signal Process.*, 29, 1042, 1981.
87. Andrews, M., An adaptive NR filter for spectrum analysis, *Comp. Electron. Eng.*, 6, 99, 1979.
88. Friedlander, B., Recursive lattice forms for spectral estimation, *IEEE Trans. Acoust. Speech Signal Process.*, 30, 920, 1982.
89. Larsen, R. D., Crawford, E. F., and Howard, G. K., Walsh analysis of signals, *Math. Biosci.*, 31, 237, 1976.
90. Smith, W. D., Walsh versus Fourier estimators of the EEG power spectrum, *IEEE Trans. Biomed. Eng.*, 28, 790, 1981.
91. Lai, D. C. and Larsen, H., Walsh spectral estimates with applications to the classification of EEG signals, *IEEE Trans. Biomed. Eng.*, 28, 790, 1981.
92. Oppenheim, A., Johnson, D., and Steiglitz, K., Computation of spectra with unequal resolution using FFT, *Proc. IEEE*, 64, 299, 1971.

0 / 8

\ E₉

new / new]

Chapter 9

ADAPTIVE FILTERING

I. INTRODUCTION

Filtering is used to process a signal in such a way that the signal-to-noise ratio is enhanced, noise of a certain type is eliminated, the signal is smoothed, or "predicted", or classification of the signal is achieved. When the signal and noise are stationary and their characteristics are approximately known or can be assumed, an optimal filter can be designed *a priori*. Such is the Wiener filter discussed in Chapter 6 and the matched filter that is presented in Chapter 1, Volume II.

When no *a priori* information on the signal or noise is available, or when the signal or noise is nonstationary, *a priori* optimal filter design is not possible. Adaptive optimal filters are filters that can automatically adjust their own parameters, based on the incoming signal. The adaptation process is conducted such that the filter uses incoming signal information in order to adapt its own parameters so that a given performance index is optimized. Adaptive filters thus require little or no *a priori* knowledge of the signal and noise characteristics.

The applications of adaptive filtering to signal processing in general, and to biomedical signal processing in specific, has been preceded by the development and use of adaptive algorithms¹⁻³ in control theory. Although adaptive filters and algorithms, used in signal processing, are basically similar to those used in control systems, some differences do exist, which demand new design approaches.⁴

Since no (or almost no) *a priori* information is available, the adaptive filter requires an initial period for learning or adaptation. During this period, its performance is unsatisfactory. The time of adaptation is clearly an important characteristic of the filter. Signals, where fast changes are expected, require filters which adapt rapidly. Care should be placed when designing such filters since the filter may track rapid artifacts. After initial adaptation, the filter is supposed to act optimally, while tracking the nonstationary changes in signal and noise. The nonperfect ability of the filter to estimate signal and noise statistics prevents it from being truly optimal. In practical design, however, this loss of performance can be made quite small.⁵

The adaptive filter is required to perform calculations to satisfy the performance index and must have provision for changing its own parameters. Digital techniques, with or without computing device, have clear advantages here over analog techniques. It is mainly because of this reason that most adaptive filter implementations are performed by discrete systems. We shall only consider here discrete adaptive filters operating on sampled signals.

The next section in this chapter will present the general structure of an adaptive filter. This will follow by a detailed discussion of least mean square (LMS) adaptive filter.⁵⁻⁸ The use of the LMS adaptive filter for line enhancement⁹⁻¹⁴ and noise cancellation^{7,15-19} will be discussed with some biomedical applications.^{7,15,17} Finally, the multichannel²⁰ and the time-sequenced adaptive filter²¹⁻²² will be introduced. The discussions are based mainly on Widrow's papers.

The LMS filter discussed here is by no means the only type of adaptive filter available. Other types are discussed in the literature,²³⁻²⁸ where various performance criteria or structures are used (e.g., the lattice structure²⁵⁻²⁷), or in which the structure as well as the weight is adaptable.²⁸

II. GENERAL STRUCTURE OF ADAPTIVE FILTERS

A. Introduction

The adaptive filter consists of three main parts: the performance index which is to be optimized, the algorithm that recomputes the parameters of the filter, and the structure of the filter which actually performs the required operations on the signal. Claasen⁴ has suggested the classification of adaptive filters according to their major parts and to their goals. We follow this approach here.

The performance index is best determined by the application. When using adaptive filtering for elimination of maternal ECG in automatic fetal ECG monitoring, the performance index may be the minimization of false detection. This, however, may be a difficult criterion to implement since we do not know when a false detection has happened. Therefore, we look for performance criteria that will be easily implemented. In most applications, the minimization of the square of an output error is found to be a satisfactory criterion.

The algorithm is the mechanism by means of which the parameters, optimizing the criterion, are calculated. Two basic types of algorithms are to be considered. The first is a nonrecursive algorithm. It requires the collection of all data in a given time window and solving the necessary equations. The exact least square method is such an algorithm. The algorithm usually requires the solution of a set of linear equations by the inversion of a matrix, and the results are not available in real time. The second type of algorithm is the recursive algorithm which updates itself with every incoming signal sample, or a small group of samples. This algorithm usually requires gradient methods and convergence must be checked. Results are available immediately and tracking of signal nonstationarities is possible.

The structure of the filter depends to some extent on the algorithm and the application. Most often a transversal filter is used because of its straightforward hardware structure and its robustness in combination with iterative algorithms. The lattice structure, though somewhat more complex, has been found to possess better convergence and sensitivity.²⁶

We shall proceed by considering various goals of adaptive filtering.

B. Adaptive System Parameter Identification

System parameter identification²⁹ is an important problem in control systems. In order to (optimally) control a system, one must know its dynamic behavior. This is usually given in terms of the differential equations relating the outputs and inputs of the system. If only the structure of the equations (order, in linear systems) is known, but not the parameters, some kind of identification algorithm must be applied in order to estimate it. Identification algorithms are widely applied in the control of sophisticated systems such as missile guidance. It has also been applied to biomedical systems.³⁰⁻³¹ Figure 1 depicts an adaptive scheme for systems parameter identification. The noisy inputs and outputs of the system are measured, either during normal operation or during special identification test. These are fed into a filter with variable coefficients. The coefficients of the filters are adjusted by the algorithm in such a way as to optimize³ the given criterion. In Figure 1, a minimum square output error criterion was chosen. After adaptation, the filter represents the best (in the sense of the given criterion) model for the system. Its coefficients are the systems identified parameters.

C. Adaptive Signal Estimation

The need often arises to estimate the output of a system, given a noisy or distorted output. The noisy input may or may not be given. Note that here, in contrast to the system identification problem, we do not require information on the system nor do we wish to model the system. Our goal here is to best estimate the output of the system, according to the given performance index. Figure 2 depicts this type of adaptive filter.

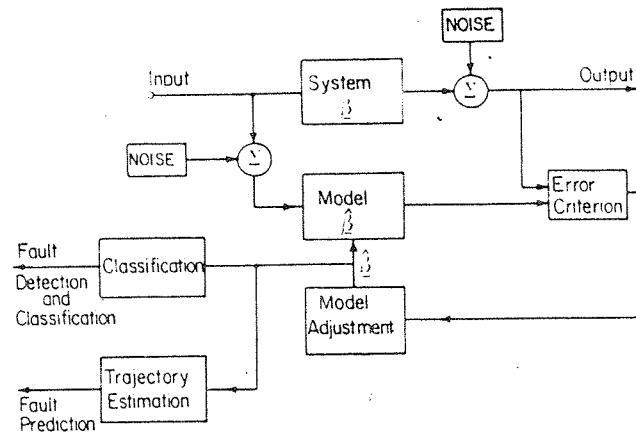


FIGURE 1. System parameter identification.

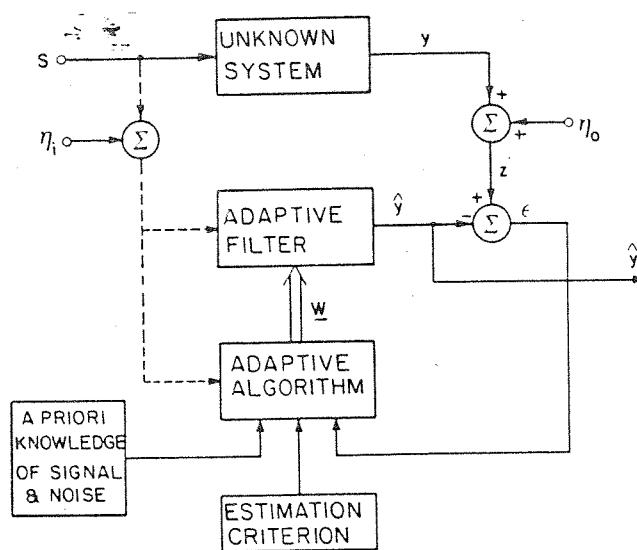


FIGURE 2. Adaptive signal estimation.

D. Adaptive Signal Correction

Assume we are given a signal, produced by the same system, the input to which is inaccessible. We require performance of some correction to the signal. This may be, for example, the elimination of power line interferences from EEG records or the elimination of maternal ECG from fetal ECG recordings. Information on both the signal and the required correction are of extreme importance here. We may want to use this filter, for example, to enhance the alpha rhythm in an EEG signal (adaptive line enhancer, ALE). The filter must be carefully designed so that it will not enhance noise components and introduce it as alpha waves. Figure 3 depicts the adaptive signal correction filter.

III. LEAST MEAN SQUARES (LMS) ADAPTIVE FILTER

A. Introduction

We shall discuss here adaptive filters using the least mean square (LMS) algorithm developed⁵ by Widrow and Hoff in 1960. The filter consists of reference inputs, variable gains multipliers (weights), an adaptation algorithm, and an additional input signal denoted

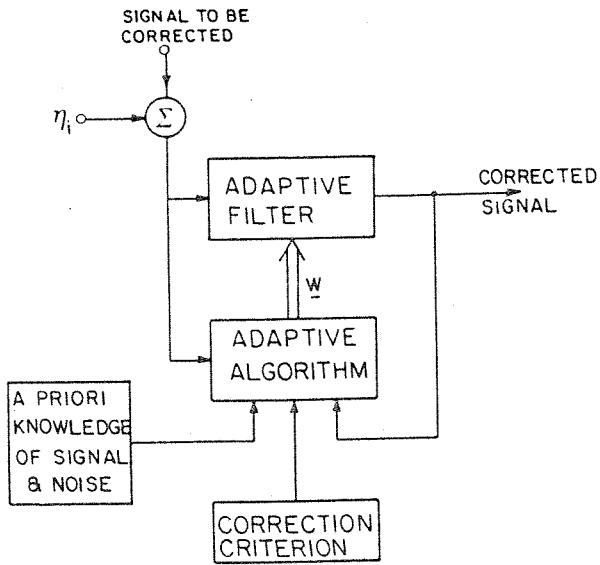


FIGURE 3. Adaptive signal correction.

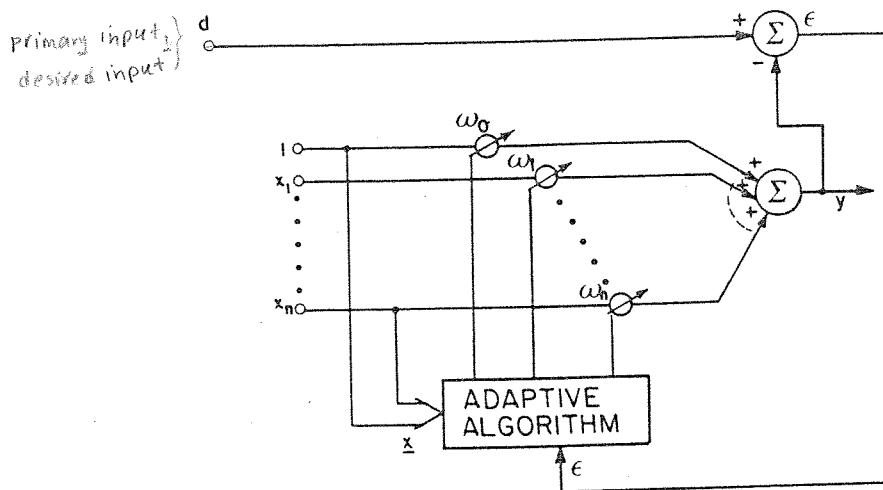


FIGURE 4. Adaptive linear combiner.

the "primary input". We shall present the principal component of the adaptive filter, namely, the adaptive linear combiner, and then discuss various structures of the adaptive filter for adaptive noise cancelling and line enhancement.

B. Adaptive Linear Combiner

The adaptive liner combiner⁷⁻⁸ will be discussed here as a separate unit. It will be later attached to various configurations to meet various applications. The line combiner is shown in Figure 4.

The inputs, x_{ij} , are reference inputs. We shall see later that these are given, or are derived from the signal itself. The primary input, d_j , is sometimes called the "desired" input. This name, however, should be carefully considered. It is definitely not the desired signal since if it were, there was no need to process it by the filter. We shall later see that from the point of view of the adaptation algorithm, it can indeed be considered the "desired" input.

Assume we have a set of n reference (discrete) inputs x_{kj} , $k = 1, 2, \dots, n$ where j denotes the time. Define the $n + 1$ dimensional reference input vector \underline{x}_j :

$$\underline{x}_j^T = [x_{0j}, x_{1j}, \dots, x_{nj}] \quad (9.1)$$

where x_{0j} is usually a constant set to the value 1. Its role is to take care of biases in the inputs. We also define a vector of the variable gains (weights):

$$\underline{W}^T = [w_0, w_1, \dots, w_n] \quad (9.2)$$

where w_0 is the bias weight.

The summer output, at time j , is \hat{s}_j , given by:

$$\hat{s}_j = \sum_{i=0}^n w_i x_{ij} = \underline{W}^T \underline{x}_j = \underline{x}_j^T \underline{W} \quad (9.3)$$

We consider \hat{s}_j to be the estimation of a signal. This signal will depend on the problem to which the combiner is applied. We shall define the error signal ϵ_j by:

$$\epsilon_j = d_j - \hat{s}_j = d_j - \underline{W}^T \underline{x}_j = d_j - \underline{x}_j^T \underline{W} \quad (9.4)$$

ϵ_j is therefore the error between the desired and estimated signals.

C. The LMS Adaptive Algorithm

The performance index for our algorithm is the mean square error. The LMS adaptive algorithm's task is to adjust the weights, \underline{W} , in such a way as to minimize the mean square error. The mean square error is calculated by squaring Equation 9.4 and taking the expectation. Assuming the reference and primary inputs to be stationary, and the weights fixed, we get:

$$E\{\epsilon_j^2\} = E\{d_j^2\} - 2E\{d_j \underline{x}_j^T\} \underline{W} + \underline{W}^T E\{\underline{x}_j \underline{x}_j^T\} \underline{W} \quad (9.5)$$

Define the cross correlation between d_j and \underline{x}_j as the vector \underline{p} :

$$\underline{p}^T = E\{d_j \underline{x}_j^T\} = E\{[d_j x_{0j}, d_j x_{1j}, \dots, d_j x_{nj}]\} \quad (9.6)$$

and denote the symmetric, positive definite input correlation matrix, R ,

$$R = E\{\underline{x}_j \underline{x}_j^T\} \quad (9.7)$$

so that the mean square error can be expressed as the quadratic function of the weights:

$$E\{\epsilon_j^2\} = E\{d_j^2\} - 2\underline{p}^T \underline{W} + \underline{W}^T R \underline{W} \quad (9.8)$$

In the stationary case, the minimization of Equation 9.8 means the adjustment of the weights, descending along the surface (Equation 9.8) until the minimum is reached. In the nonstationary case, the minimum is drifting and the algorithm has to adapt the weights such that they track the minimum.

To find the minimum of Equation 9.8, we have to calculate the gradient of the squared error. The weighting vector, \underline{W}_{opt} , is the vector that zeroes the gradient.

$$\underline{\nabla}_j = \left[\frac{\partial E\{\epsilon_j^2\}}{\partial w_0}, \frac{\partial E\{\epsilon_j^2\}}{\partial w_1}, \dots, \frac{\partial E\{\epsilon_j^2\}}{\partial w_n} \right]^T = -2\underline{p} + 2R\underline{W}_{opt} = \underline{0} \quad (9.9)$$

hence:

$$\underline{W}_{opt} = R^{-1}\underline{p} \quad (9.10)$$

which is the matrix form of the Wiener-Hopf equation (see Chapter 6).

The LMS algorithm does not use directly Equation 9.10 for the optimal solution. Rather, it uses the method of steepest descent. We calculate the optimal vector iteratively, where in each step we change the vector proportionally to the negative of the gradient vector. Hence:

$$\underline{W}_{j+1} = \underline{W}_j - \mu \underline{\nabla}_j \quad (9.11)$$

where μ is a scalar that controls the stability and rate of convergence of the algorithm. We have added a subscript to the weighting vector to denote the number of iteration. Note that using Equation 9.11 does not require the calculations of the correlations and not the inversion of the correlation matrix. The gradient with subscript j in Equation 9.11 is given by Equation 9.9 where the derivatives are taken at $\underline{W} = \underline{W}_j$.

In practice, it is impossible to implement Equation 9.11 since the gradient involves expectations. For practical implementation we have to replace the gradient, $\underline{\nabla}_j$, with some kind of estimation, $\hat{\underline{\nabla}}_j$. Widrow has suggested the crude estimate:

$$\hat{\underline{\nabla}}_j = \left[\frac{\partial \epsilon_j^2}{\partial w_0}, \frac{\partial \epsilon_j^2}{\partial w_1}, \dots, \frac{\partial \epsilon_j^2}{\partial w_n} \right]^T \Big|_{\underline{w}=\underline{w}_j} \quad (9.12)$$

namely, to estimate the expectation of ϵ_j^2 by the value of ϵ_j^2 itself. This means that we estimate the mean by a very short finite time. The derivatives of Equation 9.12 become:

$$\hat{\underline{\nabla}}_j = 2\epsilon_j \left[\frac{\partial \epsilon_j}{\partial w_0}, \frac{\partial \epsilon_j}{\partial w_1}, \dots, \frac{\partial \epsilon_j}{\partial w_n} \right]^T \Big|_{\underline{w}=\underline{w}_j} = -2\epsilon_j \underline{x}_j \quad (9.13)$$

The right side of Equation 9.13 is calculated by taking the derivative of Equation 9.4 with respect to \underline{W} . Introducing the estimate of the gradient of Equation 9.13 into Equation 9.11 yields:

$$\underline{W}_{j+1} = \underline{W}_j + 2\mu \epsilon_j \underline{x}_j \quad (9.14)$$

The last equation is known as the Widrow-Hoff LMS algorithm. It has been shown that the expected value of the weight vector (Equation 9.14) converges to the Wiener weight vector (Equation 9.10) if the reference inputs are uncorrelated over time.

A necessary and sufficient condition for convergence⁸ is

$$\lambda_{max}^{-1} > \mu > 0 \quad (9.15A)$$

where λ_{max} is the largest eigenvalue of the correlation matrix R . The eigenvalues, however, are usually not known. It has been suggested,⁸ therefore, to use a sufficient condition for convergence:

$$1/\text{tr}(R) > \mu > 0 \quad (9.15B)$$

Since R is positive definite, $\text{tr}(R) > \lambda_{\max}$. The trace is easy to estimate since it is the total power in the reference signals. Widrow has shown that the learning curve (the curve describing the convergence of the weights \underline{W}_j to the Wiener weight) can be approximated by a decaying single exponential curve with time constant, τ

$$\tau = \frac{(n + 1)}{4\mu \text{tr } R} \quad (9.16)$$

The LMS adaptive algorithm (Equation 9.14) is easy to implement, and does not require differentiations or matrix inversion. For each iteration, it requires $n + 2$ multiplications and n additions.

D. The LMS Adaptive Filter

Consider now the case where the reference vector, \underline{x}_j , is received by a tapped delay line as in Figure 5. Here the reference vector is

$$\underline{x}_j^T = [1, x_j, x_{j-1}, \dots, x_{j-n+1}] \quad (9.17)$$

and the output of the summer, the estimated signal \hat{s}_j (Equation 9.3):

$$\hat{x}_j = \hat{s}_j = \underline{W}^T \underline{x}_j = w_0 + \sum_{i=1}^n w_i x_{j-i+1} \quad (9.18)$$

Note that \hat{s}_j is the autoregressive (AR) estimation (see Chapter 7). The LMS filter with reference Equation 9.17 is an adaptive AR filter. The AR coefficients (LPC) are optimally adapted in such a way that the output of the filter and the desired input will have minimum mean square error.

If we set $d_j = x_j$, $w_0 = w_1 = 0$, $\bar{w}_i = -w_{i+1}$, $i = 1, 2, \dots, n$, denote $\hat{s}_j = \hat{x}_j$, we get from Equation 9.18:

$$\hat{x}_j = - \sum_{i=1}^{n-1} \bar{w}_i x_{j-i} \quad (9.19)$$

which is the AR equation. The filter, under these conditions, can be used to estimate and track the LPC (AR) of a nonstationary signal.

Adaptive LMS filters have been successfully implemented on small machines. The errors due to finite word length have been analyzed.⁴⁰⁻⁴¹ Adaptive filtering can be implemented also in the frequency domain^{16,32-33} with some advantages over the time domain.¹⁶

IV. ADAPTIVE NOISE CANCELLING

A. Introduction

Consider the following problem (depicted in Figure 6). A signal $s(t)$ is contaminated with an additive noise $n_o(t)$, and with another noise, $\eta(t)$; we assume that s and n_o and η are uncorrelated. The noise, n_o , is generated by a white noise process $n(t)$ that has passed an unknown linear filter, H_1 . The additive noise $n_o(t)$ is therefore a colored noise. Assume also that we have a reference signal $x(t)$ consisting of a white noise $\xi(t)$ and another noise $n_r(t)$. The second noise is the result of the noise process $n(t)$, contributing to the primary noise, but after another unknown linear filter, H_2 . Note that here we have:

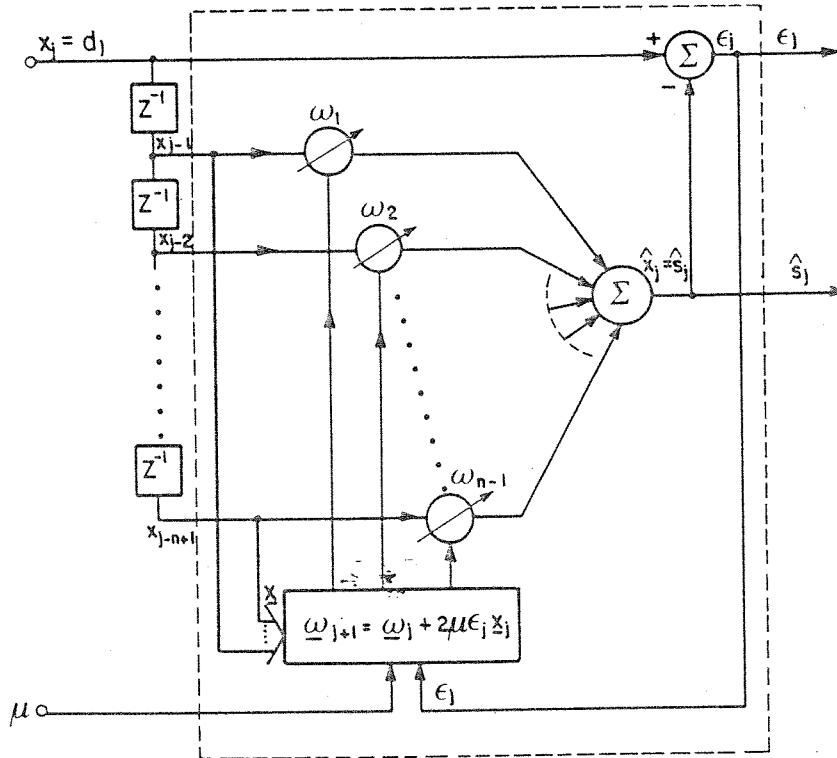


FIGURE 5. The LMS adaptive filter.

$$N_o(z) = H_2^{-1}(z^{-1})H_1(z^{-1})N_r(z) \quad (9.20)$$

where N_o , N_r , $H_1(z^{-1})$, and $H_2(z^{-1})$ are the z transforms of $n_o(t)$, $n_r(t)$, $h_1(t)$, and $h_2(t)$, respectively. We assume that the auxiliary noises $\eta(t)$ and $\xi(t)$ are white and uncorrelated with one another, with $n(t)$, and with the signal $s(t)$. The concept of the adaptive noise canceller is as follows. An adaptive estimate of $n_o(j)$, denoted by $\hat{n}_o(j)$, is calculated by the adaptive LMS filter. As shown before this filter is an adaptive AR filter estimating the unknown filter $H_2^{-1}(z^{-1})H_1(z^{-1})$ by means of the reference input $n_r(j)$ and the error. Note also that the adaptive filter does not operate as the Durbin's algorithm described in Chapter 5. There, the AR coefficients estimated were optimized in such a way as to whiten the output of the filter (the residuals). The estimated AR coefficients were the coefficients of the filter, $\hat{H}_2^{-1}(z^{-1})$. Here the criterion is the minimization of $E\{\epsilon_j^2\}$ so that the estimated AR coefficients are the coefficients of $\hat{H}_2^{-1}(z^{-1})\hat{H}_1(z^{-1})$.

Adaptive noise cancelling filters⁷ have been extensively used in biomedical signal processing and many other applications.

B. Noise Canceller With Reference Input

Assume that s , n_o , n_r and \hat{n}_o are stationary with zero mean, and recall that s and n_o are uncorrelated. The filter's output, ϵ , is given by:

$$\epsilon = \hat{s} = s + \eta + n_o - (\hat{n}_o + \hat{\eta}) \quad (9.21)$$

The expectation of the square error is

$$E\{\epsilon^2\} = E\{s^2\} + E\{(n_o + \eta) - (\hat{n}_o + \hat{\eta})\}^2 \quad (9.22)$$

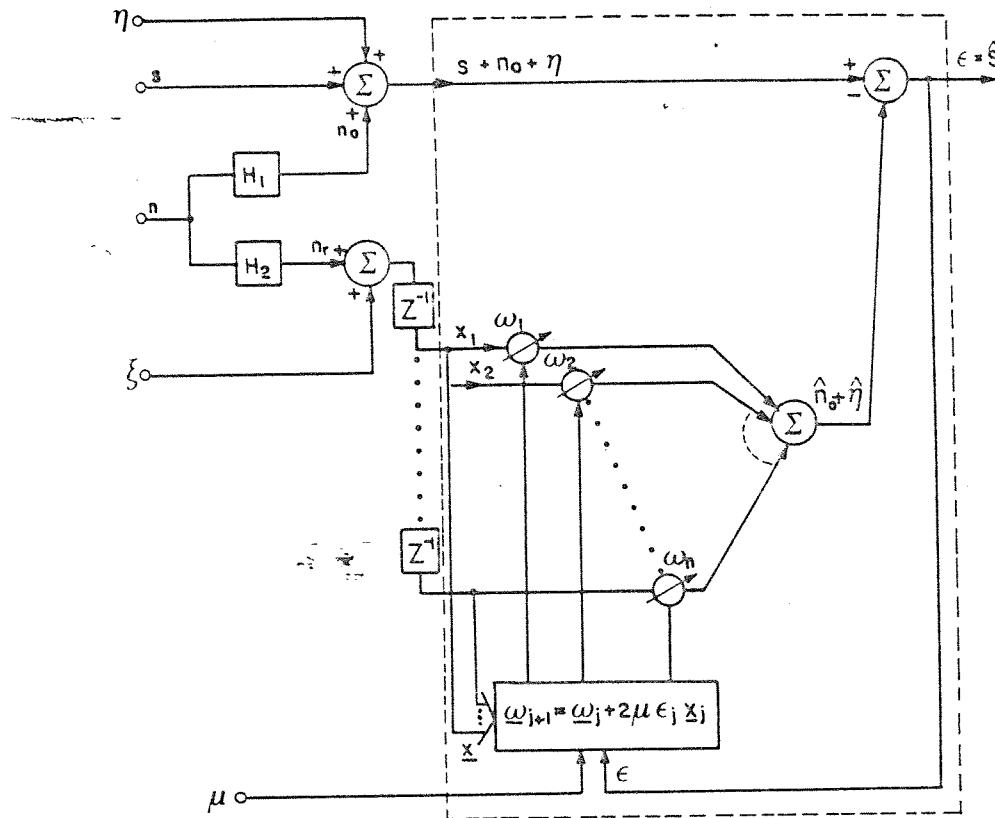


FIGURE 6. Adaptive noise canceller.

The adaptive algorithm will change the weights so that Equation 9.22 is minimized. However, changing the filter weights effects only \hat{n}_o and does not effect the term $E\{s^2\}$. Therefore, minimization of Equation 9.22 is equivalent to the minimization of $E\{(n_o + \eta) - (\hat{n}_o + \hat{\eta})\}^2$. From Equation 9.21 we get:

$$(\hat{s} - s) = (n_o + \eta) - (\hat{n}_o + \hat{\eta})$$

hence the minimization of $E\{(n_o + \eta) - (\hat{n}_o + \hat{\eta})\}^2$ also minimizes $E\{(\hat{s} - s)^2\}$. The adaptive noise canceller provides the estimate \hat{s} , which is the best least squares estimate of s .

The reference input must be correlated with the primary noise, n_o . It is this correlation that allows the LMS noise canceller to function effectively. To demonstrate this, assume that n_r is not correlated with n_o (Figure 6 does not hold for this example). The minimization of Equation 9.22 yields:

$$\begin{aligned} \text{Min } E\{\epsilon^2\} &= \text{Min } E\{s^2\} + \text{Min } E\{(n_o - \hat{n}_o)^2\} = \\ &E\{s^2\} + E\{n_o^2\} + E\{\eta^2\} + \text{Min } E\{-2\eta\hat{\eta} + (\hat{n}_o + \hat{\eta})^2\} \end{aligned} \quad (9.23)$$

The algorithm will minimize Equation 9.23 by adjusting all weights W to zero thus bringing the last term to its minimum, namely zero. The adaption noise canceller has been applied to a variety of biomedical^{17,15} and many other applications, such as echo cancellation in communication network.³⁴⁻³⁵

Example 9.1 — Elimination of Power Line Interferences

Consider an example where the noise contaminating the signal is due to power line interferences. The primary input is

$$d = s + A \cos(w_o t + \phi) \quad (9.24)$$

where the amplitude, A , and the phase, ϕ , are unknown. The frequency of the line voltage, w_o , can vary around its nominal value. Its exact value, at any moment, is unknown *a priori*. This is a common problem in biomedical signal processing. A fixed band stop filter can be designed with a notch at the nominal value of w_o , and with sufficient width to cover the expected variations in the frequency. In cases where meaningful portions of the PSD function of the signal are in the vicinity of w_o , this will cause distortions to the processed signal. Typical examples are ECG, EMG, and EEG signals, all having meaningful information in the region of 50 to 60 Hz.

We shall see now that the adaptive LMS noise canceller can operate as an adaptive narrow notch filter with its central frequency automatically tracking the variations of w_o .

For the reference signal, we take a signal which is directly proportional to the power line voltage; we choose:

$$x_1(t) = B \cos(w_o t + \psi) \quad (9.25A)$$

Here B , w_o , and ψ are known. This can simply be the voltage taken directly from the wall outlet. The second reference is derived from $x_1(t)$ by shifting it 90°. Hence:

$$x_2(t) = B \sin(w_o t + \psi) \quad (9.25B)$$

and the samples of the reference vector

$$\underline{x}^T = [B \cos(w_o j \Delta T + \psi), B \sin(w_o j \Delta T + \psi)] \quad (9.25C)$$

where ΔT is the sampling interval. Note that the output of the LMS combiner is a linear combination of the two normal phasors (Equation 9.25C). It is clear that we can represent the cosine primary noise as a linear combination of the phasors, given the right weights. Any change in w_o will appear both in the primary and reference signals. The LMS will thus track the variations in w_o . Widrow⁷ has shown that the adaptive noise canceller described in Figure 7 is equivalent to a notch filter, with notch frequency always at w_o , and Q (ratio of center frequency to bandwidth) given by:

$$Q \approx \frac{1}{2} \frac{w_o \Delta T}{\mu B^2} \quad (9.26)$$

This configuration was used⁷ to cancel 60-Hz interferences in ECG. Figure 8 shows the cancellation effects. Note the adaptation of the filter. Other types of learning filters for power line interferences removal are available.¹⁷

Example 9.2 — Elimination of ECG Interferences

In some heart transplantation procedures,³⁶ the new donor heart is attached to a small remnant of the old heart. The remaining portion of the old heart contains its SA node, which remains connected to the vagus and sympathetic nerves. The old SA node continues to beat at a rate controlled by the central nervous system (CNS). The new heart, containing both SA and AV nodes, is not connected to the CNS. It thus generates spontaneous pulses. When

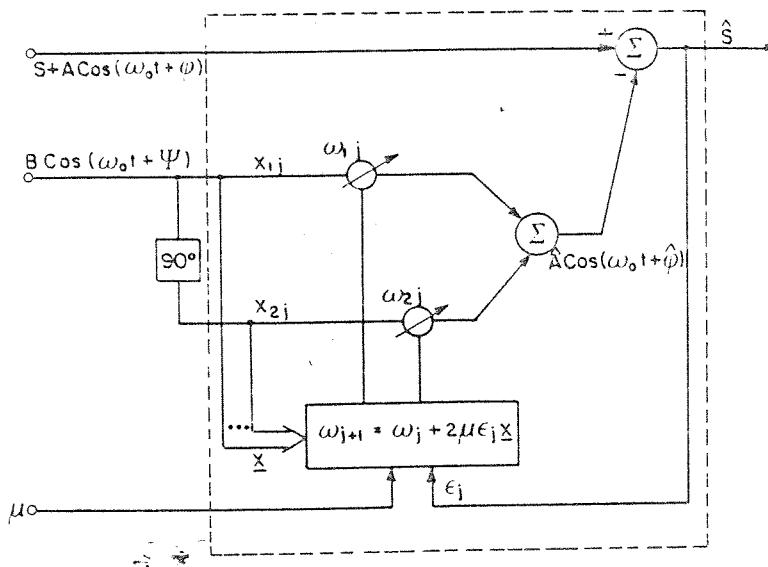


FIGURE 7. Adaptive noise canceller for sinusoidal noise cancelling.

monitoring the ECG of a patient after such heart transplantation, both "old" and "new" QRS complexes are present. It is of interest to the physician to be able to separate between the two and to be able to analyze the "old" ECG without the interferences of the "new" ECG. Adaptive noise cancelling has been applied to this problem by Widrow⁷ and his co-workers. The primary input was supplied by a catheter, inserted through the left brachial vein and the vena cava to the atrium of the "old" heart. The reference input was supplied by ordinary chest electrodes, which carried mainly the "new" heart's signals. Figure 8B shows the improvement in the signal after the application of adaptive noise cancelling.

Example 9.3 — Elimination of Electrosurgical Interferences

During surgery, the electrosurgical unit (ESU) is often used. It supplies a high frequency signal, modulated at twice the power line frequency, to the tip of the surgeon's knife. The power delivered, in the range of 100 to 200 watts, is used to aid in cutting tissue and coagulate severed blood vessels. While the ESU is in operation, strong interferences are introduced into the ECG recorded during the operation. These interferences are due to ESU currents passing through the patient's tissues. Conventional grounding, screening, and other instrumental methods are unable to significantly reduce these interferences. Yelderman, Widrow, and their co-workers¹⁵ have applied the adaptive noise canceller to this problem. Primary signal was taken from normal chest electrodes. Reference signal was taken from a pair of electrodes placed on the arm in such a way that they pick an interfering ESU signal with minimum ECG. Special care had to be taken to isolate the electrodes from the monitoring instrumentation. Passive filtering applied to both primary and reference signals removed most of the spectral components above 60 Hz. Strong interferences from the first three harmonics of the power line and random low frequency signals still remained, these interferences were highly nonstationary and changed considerably as the surgeon moved the knife. The LMS adaptive noise canceller was used to cancel the nonstationary interferences. Figure 8C shows some of the results.

Example 9.4 — Noise Reduction for the Hearing Impaired

Hearing impairment is a major chronic disability affecting people all over the world. The reduced ability to understand speech in everyday communication, in a noisy environment, is a common phenomena for the hearing impaired.³⁷ Widrow⁷ has applied the LMS noise

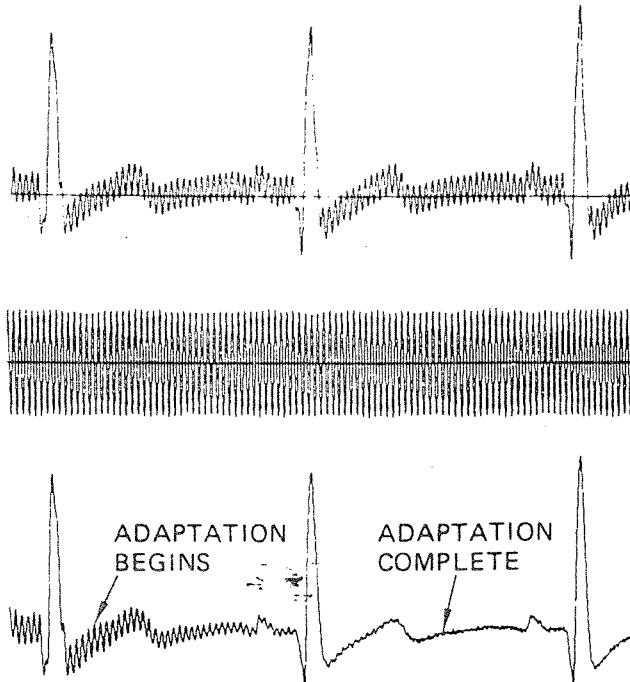


FIGURE 8. Noise cancellation in ECG. (A) Adaptive cancellation of power line interferences; (B) removal of "new" ECG in heart transplant patient; (C) cancellation of electrosurgical noise. (From Widrow, B., Glover, J. R., Jr., McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., Zeidler, J. R., Dong, E., Jr., and Goodlin, R. C., *Proc. IEEE*, 63, 1962, 1965. With permission.)

cancelling filter to the problem of noise reduction in pilot radio communications. The noise in the cockpit is highly nonstationary due to variations in engine speed and load. A second microphone was placed in a suitable location in the cockpit to produce the reference signal. In Widrow's experiments, simulated cockpit noise made the unprocessed speech unintelligible. After LMS processing, the output power of the interferences was reduced by 20 to 25 dB, rendering the interferences barely perceptible to the listener. No noticeable distortion was introduced to the speech signal itself. Other algorithms, which use only one microphone, have been suggested.³⁸ These, however, were ineffective for low signal-to-noise ratios.

The LMS noise canceller was applied to the problem of improving speech communication for the hearing impaired.³⁷ Noise cancellation is important, for example, in cases wherein a hearing impaired child must function in an educational setting. A special amplification system can be used to amplify the teacher's voice for the child. The teacher's microphone, however, picks not only the teacher's voice but also the environmental noise of the classroom. A reference microphone can be placed apart from the teacher to pick the reference noise (which will also, in this case, contain some of the primary signal).

Applying the LMS noise canceller in a controlled environment using very noisy speech improved intelligibility of speech from near zero to about 30 to 50%.

C. Noise Canceller Without Reference Input

Consider the case where periodic interferences contaminate the signal, but this time no reference signal is available. If the signal is broad band, such that its correlation function drops relatively fast, we can "generate" a reference signal out of the primary signal. We shall form the reference, x_j , by placing a delay, τ_R , on the primary signal (see Figure 9). Hence we have:

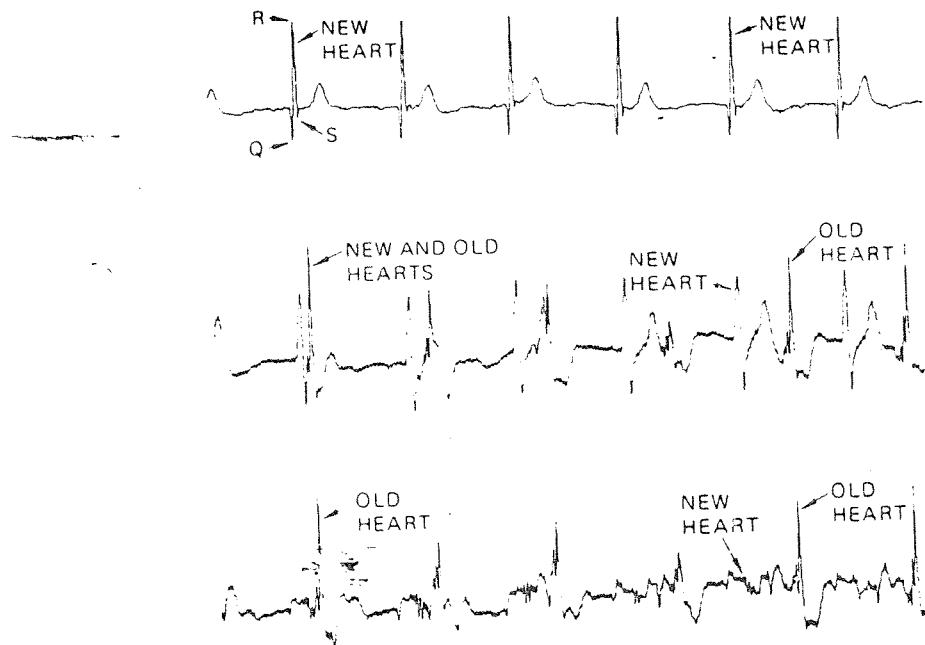


FIGURE 8B.

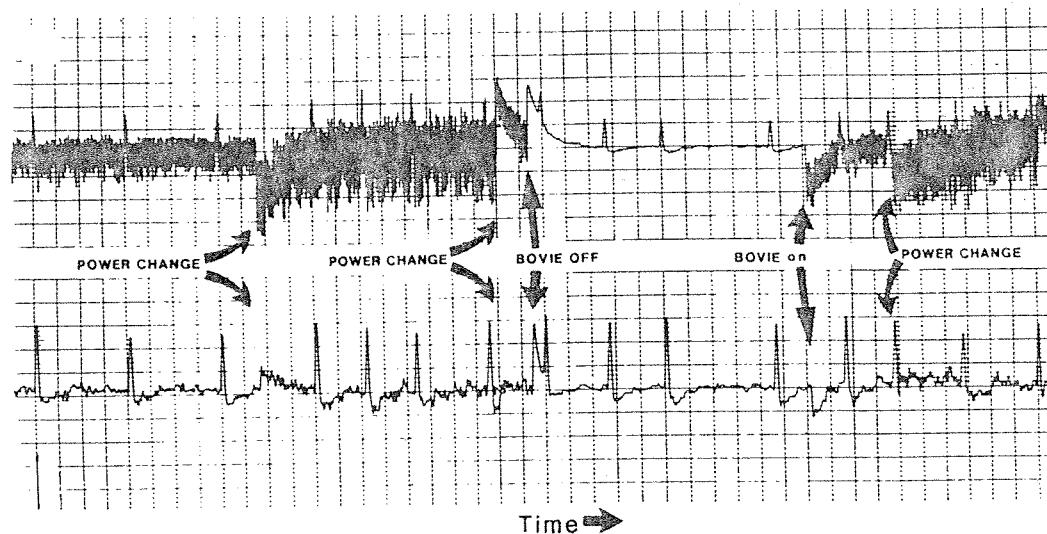


FIGURE 8C.

$$d_j = s_j + n_j \quad (9.27)$$

$$x_j = d(j - \tau_R) = s(j - \tau_R) + n(j - \tau_R) \quad (9.28)$$

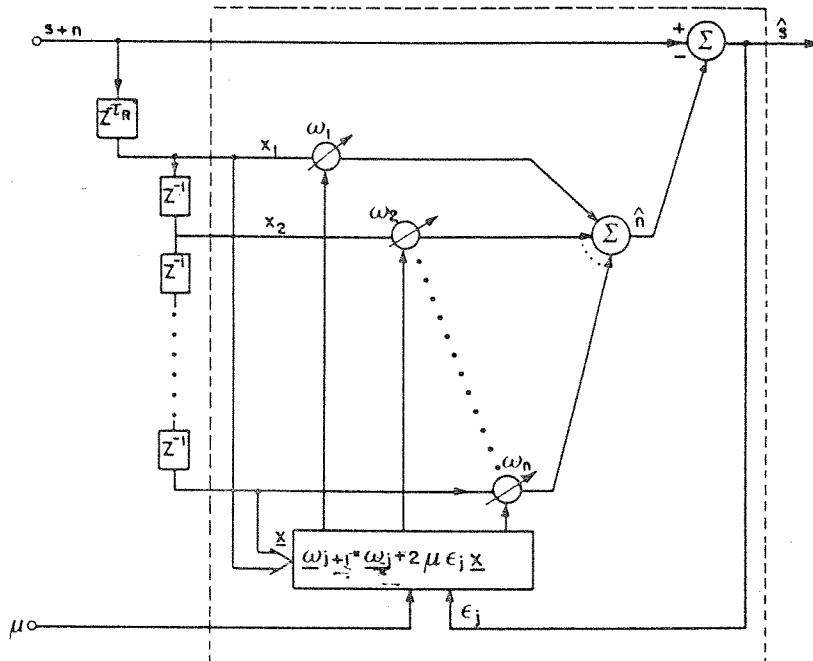


FIGURE 9. Noise canceller without reference.

Let the autocorrelation of the signal be $R_s(\tau)$. We choose τ_R such that $R_s(\tau_R) < \epsilon$, where ϵ is some small number. The delayed signal $s(j - \tau_R)$ in the reference will (almost) be uncorrelated with the primary signal. The reference $n(j - \tau_R)$ is correlated with the primary noise since it is periodic.

Adaptive noise cancellers can be implemented also in the frequency domain.^{16,33}

D. Adaptive Line Enhancer (ALE)

Consider now the problem of detecting a sine wave in white noise.⁷ We can apply the same configuration used in the last section for cancelling periodic noise without reference. Here, however, the signal is periodic and the noise is broad band. The adaptive combiner will supply the estimate of the periodic signal. Refer to Figure 10 and note that the only difference between the ALE and the noise canceller of Figure 9 is that here we take the output from the combiner.

The ALE has been extensively used⁹⁻¹³ to detect and track a moving spectral line buried in background broad band noise, with applications mainly in sonar detection and tracking and in communications.

V. IMPROVED ADAPTIVE FILTERING

A. Multichannel Adaptive Signal Enhancement

In many practical applications, we are given a number of channels. Each channel contains a signal contaminated by additive noise. The signals in the various channels are correlated with one another in some unknown fashion; they may, however, differ in their waveshapes and power spectra. The noise signals involved are assumed to be uncorrelated with one another and with the signals. One of the channels is denoted, the primary input, and it is desired to enhance its signal. A method for such enhancement was suggested by Ferrara and Widrow.²⁰ The basic idea is to use the correlative information in each channel by means of multichannel adaptive filtering; it is based on an earlier idea applied to antenna systems.³⁹

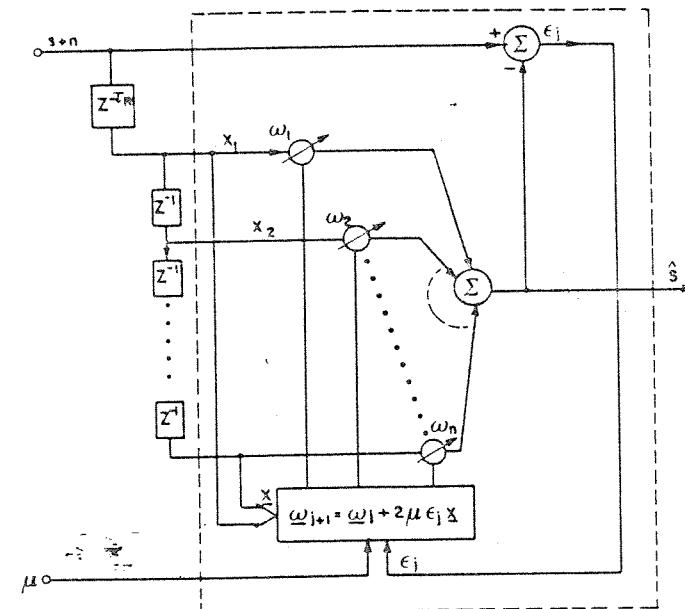


FIGURE 10. Adaptive line enhancer (ALE).

A multichannel recording of evoked potentials, where each channel represents the voltage monitored at a different location on the scalp, may serve as an example for such a problem.

Consider the i^{th} channel's output $x^{(i)}$

$$\begin{aligned} x^{(o)} &= d = s^{(o)} + n^{(o)} \\ x^{(i)} &= s^{(i)} + n^{(i)}; \quad i = 1, 2, \dots, M - 1 \end{aligned} \quad (9.29)$$

where M is the number of channels available.

Since we assume all signals $s^{(i)}$ to be somehow correlated with one another, we can arbitrarily choose the signal $s^{(o)}$, in the channel of interest, to be the "source" of all other signals. The signal in each one of the other channels can be expressed, therefore, as the output of some unknown linear filter inputted by $s^{(o)}$. Hence,

$$X(z^{(i)}) = S(z^{(i)}) + N(z^{(i)}) = H_i(z^{-1}) \cdot S(z^{(o)}) + N(z^{(i)}), \quad i = 1, 2, \dots, M - 1 \quad (9.30)$$

Here we describe these relationships in the z domain by means of the unknown linear filters, $H_i(z^{-1})$. Refer to Figure 11 for the signals model and the adaptive filter. Each signal $x^{(i)}$ is used as the desired input for an LMS adaptive filter (see Figure 5). The left side of Figure 11 is just an imaginary model. In practice, the M given channels are used directly as primary and reference inputs to the multichannel adaptive filter. In the process of adaptation, the weights of each LMS filter are simultaneously adjusted to minimize the power of ϵ_j . After convergence of all LMS filters, the output of the combiner is the best least squares estimate of the delayed primary signal $s^{(o)}$.

The delay in the primary channel is required to account for possible delays in the filters $H_i(z)$ and in the LMS filters. The distortions in the estimated signal as well as the noise power spectra of the output for the multichannel adaptive filter have been calculated.²⁰ The multichannel adaptive signal enhancer yields a substantial reduction in background noise but often at the expense of considerable signal distortion²¹ and computation load.

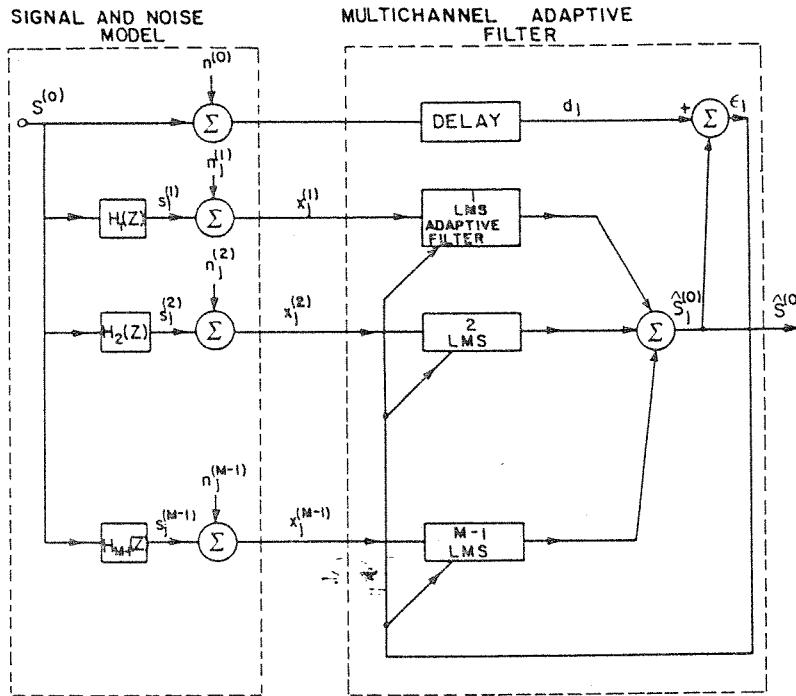


FIGURE 11. Multichannel signal enhancer.

B. Time-Sequenced Adaptive Filtering

An improvement for the conventional LMS filter has been suggested²¹ that enhances the signal with much less distortion. This new modification, termed the time-sequenced adaptive filtering, has been successfully applied to the problem of fetal ECG enhancement.²²

One of the problems with the conventional LMS is that of signal nonstationarities. The modified algorithm is especially suited for nonstationary signals with recurring statistical characteristics. These recurring events do not have to be periodic. The ECG is such a signal.

The conventional LMS adaptive filter uses the Widrow-Hoff LMS algorithm (Equation 9.14) to find the minimum of the error surface, a concave hyperparaboloidal surface (Equation 9.8). If the signal is highly nonstationary, composed of rapid recurring wavelets in noise, the LMS will not be able to track the nonstationarities. It would then converge to some best time invariant filter. The time-sequenced adaptive filter consists of several LMS filters, each slowly adapting to a specific "stationary" mode of the signal. The total filter, therefore, finds simultaneously several minima points on the error surface.

Consider the signal $d(k)$ given by:

$$d(k) = \begin{cases} s_1(k) + n(k), k \in (k_1^1, k_1^2), (k_1^3, k_1^4) \dots \\ s_2(k) + n(k), k \in (k_2^1, k_2^2), (k_2^3, k_2^4) \dots \\ \vdots \\ s_M(k) + n(k), k \in (k_M^1, k_M^2), (k_M^3, k_M^4) \dots \end{cases} \quad (9.31)$$

The signal $d(k)$ consists of M noisy processes $s_i(k)$, $i = 1, 2, \dots, M$, each one appearing in the signal at times k_i^j to k_i^{j+1} . The assumption is made that at each time only one process is present in the signal. As in the conventional LMS filter, a reference signal, $x(k)$, is given. Additional input is required here, called the sequence number, σ_k . This signal provides information concerning the type of signal, s , currently present in the signal, d . Namely, when $\sigma_k = i$, we assume that $x(k) = s_i(k) + n(k)$. Figure 12 shows the time-sequenced

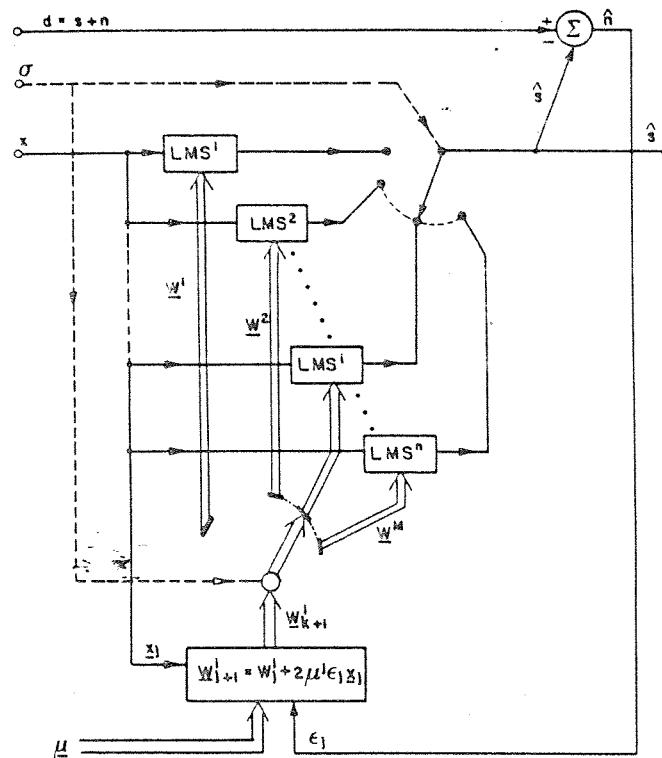


FIGURE 12. Time-sequenced adaptive filter.

adaptive filter. For each one of the M processes appearing in the signal, we set an adaptive filter. Each filter will adapt and find the minimum point appropriate for its own process. The filters are switched by two synchronized switches controlled by the sequence number, σ_k . Consider the case where $\sigma_k = i$, depicted in Figure 12. The output of the LMS algorithm is connected to the i th LMS filter, adjusting its weight vector \underline{w}^i . This is done using the constant μ^i , the i th element of the vector $\underline{\mu}$.

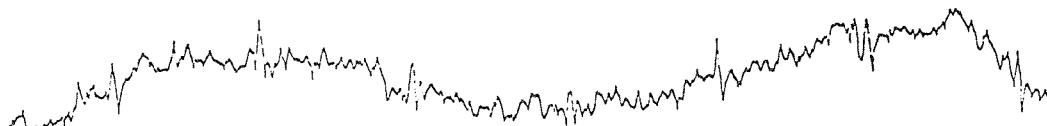
$$\underline{\mu}^T = [\mu^1, \mu^2, \dots, \mu^M] \quad (9.32)$$

The output of the i th LMS filter is connected to the summer to provide the error, ϵ_k . Note that when the sequence number changes, another filter will be selected. The same LMS algorithm will now adapt the weightings of the new filter using the appropriate constant, μ . Hence the adaptation process can be written as:

$$\underline{w}_{k+1}^i = \begin{cases} \underline{w}_k^i + 2\mu^i \epsilon_k X_k; & \sigma_k = i \\ \underline{w}_k^i; & \text{otherwise} \end{cases} \quad (9.33)$$

The computational load of the time-sequenced adaptive filter is almost the same as that required from the conventional LMS filter, since in both cases only one vector is being adjusted at a given time. The time-sequenced filter requires some more operations for the switching and selection of proper constants and weights. The convergence time of the filter is longer than that of the conventional LMS filter, since here each one of the filters is

(a) ABDOMINAL LEAD # 1 WITH MATERNAL ECG CANCELLED



(b) ABDOMINAL LEAD # 2 WITH MATERNAL ECG CANCELLED



(c) CONVENTIONAL LMS LEAST-SQUARES ENHANCING



(d) TIME-SEQUENCED ENHANCING



FIGURE 13. Enhancement of fetal ECG. (From Ferrara, E. R. and Widrow, B., *IEEE Trans. Biomed. Eng.*, 29, 458, 1982. © 1982 IEEE. With permission.)

adapted only in the periods when the appropriate signal is present. Also, the memory required for the time-sequenced filter is larger than that required by the conventional one.

When σ_k is given with no error, the time-sequenced filter converges to the correct optimal time varying filter. In practical applications, however, the sequence number, σ , is not perfect but subject to "jittering" which causes the filter to be less than optimal.

Example 9.5 — Fetal ECG Enhancement

The time-sequenced adaptive filter has been applied²² to the enhancement of fetal ECG. Fetal ECGs are the signals extracted, by means of surface electrodes, from the mother's abdomen. The fetal ECG contains strong maternal QRS complexes, maternal muscle noise (EMG), and power line interferences, in addition to the weak fetal QRS complexes. Both maternal and power line interferences can be removed effectively by means of conventional LMS adaptive noise canceller. The time-sequenced adaptive filter has been used²² to remove the EMG without severely distorting the fetal QRS. Two abdominal leads preprocessed for removal of maternal ECG and power line interferences were used. The sequence number, σ_k , was achieved by using matched filtering (see Chapter 1, Volume II) for R wave detection. Filters with 75 weight and a bias weight were used. Figure 13 shows the results of fetal ECG enhancement with conventional and with time-sequenced adaptive filtering. The superiority of the latter is clearly demonstrated.

REFERENCES

1. Davisson, L. D., A theory of adaptive filtering, *IEEE Trans. Inf. Theory*, 12, 97, 1966.
2. Mehra, R. K., Approaches to adaptive filtering, *IEEE Trans. Autom. Control*, 17, 693, 1972.
3. Johnson, C. R., The common parameter estimation basis of adaptive filtering, identification and control, *IEEE Trans. Acoust. Speech Signal Process.*, 30, 587, 1982.

4. Claasen, T. A. C. M. and Mecklenbrauker, W. F. G., Overview of adaptive techniques in signal processing, in *Signal Processing II: Theories and Applications*, Schüssler, H. W., Ed., Elsevier, Amsterdam, 1983.
5. Widrow, B., Adaptive filters, in *Aspects of Network and System Theory*, Kalman, R. E. and De-Claris, N., Eds., Holt Rinehart and Winston, New York, 1970.
6. Feintuch, P. L., An adaptive recursive LMS filter, *Proc. IEEE*, 64, 1622, 1976.
7. Widrow, B., Glover, J. R., Jr., McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., Zeidler, J. R., Dong, E., Jr., and Goodlin, R. C., Adaptive noise cancelling: principles and applications, *Proc. IEEE*, 63, 1962, 1975.
8. Widrow, B., McCool, J. M., Larimore, M. G., and Johnson, C. R., Jr., Stationary and nonstationary learning characteristics of the LMS adaptive filter, *Proc. IEEE*, 64, 1151, 1976.
9. Reddy, V. U., Egardt, B., and Kailath, T., Optimized lattice form adaptive line enhancer for a sinusoidal signal in broad band noise, *IEEE Trans. Acoust. Speech Signal Process.*, 29, 702, 1981.
10. Treichler, J. R., Transient and convergent behavior of the adaptive line enhancer, *IEEE Trans. Acoust. Speech Signal Process.*, 27, 53, 1979.
11. Richard, J. T. and Zeidler, J. R., Second order output statistics of the adaptive line enhancer, *IEEE Trans. Acoust. Speech Signal Process.*, 27, 31, 1979.
12. Bershad, N. J. and Feintuch, P. L., The recursive adaptive LMS filter — a line enhancer application and analytical model for mean weight behavior, *IEEE Trans. Acoust. Speech Signal Process.*, 28, 652, 1980.
13. Fisher, B. and Bershad, N. J., The complex LMS adaptive algorithm — transient weight mean and covariance with applications to the ALE, *IEEE Trans. Acoust. Speech Signal Process.*, 31, 34, 1983.
14. David, R. A. and Stearns, S. D., A comparison of adaptive line enhancers, in *Signal Processing II: Theories and Applications*, Schüssler, H. W., Ed., Elsevier, Amsterdam, 1983.
15. Yelderman, M., Widrow, B., Cioffi, J. M., Hesler, E., and Leddy, J. A., ECG enhancement by adaptive cancellation of electrosurgical interference, *IEEE Trans. Biomed. Eng.*, 30, 392, 1983.
16. Reed, F. A. and Feintuch, P. L., A comparison of LMS adaptive cancellers implemented in the frequency domain and the time domain, *IEEE Trans. Acoust. Speech Signal Process.*, 29, 770, 1981.
17. Furno, G. S. and Tompkins, W. J., A learning filter for removing noise interferences, *IEEE Trans. Biomed. Eng.*, 30, 234, 1983.
18. Shensa, M. J., Non Wiener solutions of the adaptive noise canceller with a noisy reference, *IEEE Trans. Acoust. Speech Signal Process.*, 28, 468, 1980.
19. Clover, J. R., Adaptive noise cancelling applied to sinusoidal interferences, *IEEE Trans. Acoust. Speech Signal Process.*, 25, 484, 1977.
20. Ferrara, E. R. and Widrow, B., Multichannel adaptive filtering for signal enhancement, *IEEE Trans. Acoust. Speech Signal Process.*, 29, 766, 1981.
21. Ferrara, E. R. and Widrow, B., The time sequenced adaptive filter, *IEEE Trans. Circuits Syst.*, 28, 519, 1981.
22. Ferrara, E. R. and Widrow, B., Fetal electrocardiogram enhancement by time sequenced adaptive filtering, *IEEE Trans. Biomed. Eng.*, 29, 458, 1982.
23. Parker, S. R. and Griffiths, L. J., Eds., Special issue on adaptive signal processing, *IEEE Trans. Acoust. Speech Signal Process.*, 29, 1981.
24. Weiss, A. and Mitra, D., Digital adaptive filters: conditions for convergence, rates of convergence, effects of noise and errors arising from implementation, *IEEE Trans. Inf. Theory*, 25, 637, 1979.
25. Friedlander, B., System identification techniques for adaptive signal processing, *Circuits Syst. Signal Process.*, 1, 3, 1982.
26. Friedlander, B., Lattice filters for adaptive processing, *Proc. IEEE*, 70, 829, 1982.
27. Gibson, C. J. and Haykin, S., Learning characteristics of adaptive lattice filtering algorithms, *IEEE Trans. Acoust. Speech Signal Process.*, 28, 681, 1980.
28. Soderstrand, M. A. and Yigil, M. C., An adaptive filter which adapts structure as well as weights, *Int. J. Electron.*, 55, 533, 1983.
29. Astrom, K. J. and Eykhoff, P., System identification — a survey, *Automatica*, 7, 123, 1971.
30. Cohen, A., Parameter estimation for medical diagnosis, *Proc. 3rd IFAC Symp. Ident. and Syst. Param. Est.*, North Holland, Amsterdam, 1973, 239.
31. Cohen, A., Transplantability tests for isolated perfused organs, *Med. Biol. Eng. Comput.*, 15, 349, 1977.
32. Dentino, M., McCool, J., and Widrow, B., Adaptive filtering in the frequency domain, *Proc. IEEE*, 66, 1958, 1978.
33. Mansour, D. and Gray, A. H., Unconstrained frequency domain adaptive filter, *IEEE Trans. Acoust. Speech Signal Process.*, 30, 726, 1982.
34. Agassi, O. and Messerschmitt, Nonlinear echo cancellation of data signals, *IEEE Trans. Commun.*, 30, 2421, 1982.

35. Verhoeckx, N. A. M., van den Elzen, H. C., Snijders, F. A. M., and van Gerwen, P. J., Digital echo cancellation for baseband data transmission, *IEEE Trans. Acoust. Speech Signal Process.*, 27, 768, 1979.
36. Lower, R. R., Stofer, R. C., and Shumway, N. E., Homovital transplantation of the heart, *J. Thoracic Cardiovas. Surg.*, 41, 196, 1961.
37. Chabries, D. M., Christiansen, R. W., Brey, R. H., and Robinette, M. S., Application of the LMS adaptive filter to improve speech communication in the presence of noise, *Proc. ICASSP*, IEEE, New York, 1982, 148.
38. Lim, J. S. and Oppenheim, A. V., Enhancement and bandwidth compression of noisy speech, *Proc. IEEE*, 67, 1586, 1979.
39. Widrow, B., Mantey, P. E., Griffiths, L. J., and Goode, B. B., Adaptive antenna systems, *Proc. IEEE*, 55, 2143, 1967.
40. Caraiscos, C. and Lin, B., A roundoff error analysis of the LMS adaptive algorithm, *IEEE Trans. Acoust. Speech Signal Process.*, 32, 34, 1984.
41. Normile, J. O. and Boland, F. M., Adaptive filtering with finite wordlength constraints, *IEE Proc.*, 130(E), 42, 1983.

INDEX

A

Acetylcholine (ACh), 12
 ACM, see Autocorrelation method
 Acquisition, 5—6
 Actin, 12
 Action-potential, 11—12
 Adaptation, 141
 Adaptive algorithm, 135, 142
 Adaptive delta modulation (ADM), 32
 Adaptive estimation, 135
 Adaptive filtering, 7, 141—160
 adaptation time, 141
 adaptive signal correction, 143—144
 adaptive signal estimation, 142—143
 adaptive system parameter identification, 142—
 143
 algorithm, 142
 frequency domain, 147
 general structure of filters, 142—143
 improved, 154—158
 lattice structure, 141—142
 least mean square, 141, 143—147
 line enhancement, 141
 multichannel, 141
 multichannel adaptive signal enhancement, 154—
 156
 noise cancellation, 141, 147—154
 performance index, 142
 time-sequenced, 141, 156—158
 Adaptive filters, 73
 Adaptive linear combiner, 144—145
 Adaptive line enhancer (ALE), 143
 noise cancellation, 154—155
 Adaptive segmentation, 101—106
 ADM, see Adaptive delta modulation
 Akaike information theoretic criterion (AIC), 97
 ALE, see Adaptive line enhancer
 Algorithms, 81
 Aliasing in frequency, 30
 All zero model, 83
 “Almost” periodic signals, 3—4, 15
 Almost stationary segments, 135
 Amount of information, 122
 Anaesthesia, 109
 Analog processing, 29
 Analog systems, 4
 Analog to digital (A/D) conversion systems, 29
A priori information, 5, 7—73, 130, 135, 141
A priori optimal filter design, 141
 AR, see Autoregressive model
 ARIMA, see Autoregressive integrated moving
 average
 ARMA, see Autoregressive moving average
 ARMAX, see Autoregressive moving average exogenous variables
 Autocorrelation, 26, 85, 89
 Autocorrelation coefficients, 93

Autocorrelation function, 5, 25
 Autocorrelation matrix, 93
 Autocorrelation measure (ACM) method, 102—103
 Autocorrelation method, 86, 106
 Autocovariance, 85
 Autoregressive integrated moving average
 (ARIMA), 84, 99—100
 Autoregressive (AR) model, 83—88
 comparison with other methods, 136
 least squares model, 85—88
 order estimation, 95—96
 spectral estimation, 109, 122—125
 Autoregressive moving average (ARM), 83, 85, 100
 Akaike information theoretic criterion, 97
 comparison with other methods, 136
 filters, 74
 lattice filters, 99
 order estimation, 95
 spectral estimation, 109, 126—132
 Autoregressive moving average exogenous variables
 (ARMAX), 82—83
 Availability of long records, 45
 AV node, 150
 Axon, 9—11

B

Background broad band noise, 154
 Band-limited continuous signal, 29, 31
 Band-limited spectrum, 135
 Bartlett window, 111
 Bayes' rule, 17—18
 Bias, 124
 Bilinear transformation method, 74
 Bioelectric potentials, 9
 Bioelectric signals, 6
 muscle, 12
 nerve cell, see also Nerve cell, 9—12
 neuron, 9
 origin of, 9—13
 volume conductors, 12—13
 Biomedical signals, see also specific topics, 1
 Blackman-Tukey algorithm, 110, 124
 Blackman-Tukey procedure, 109, 111—113, 117,
 124, 126, 135
 comparison with other methods, 136
 Block quantizer, 36
 Brain activities, 13

C

Capon's spectral estimation, 133—134, 136
 CAT-computed averaged transients, see Synchronous averaging
 Cell body, 9—11
 Central limit theorem, 26—27

Central nervous system (CNS), 9, 140

Cepstral analysis, 74, 76—77

Cepstrum, 76

Chebychev inequality, 60

Chemical signals, 6

Circadian rhythms, 101

Classification of signals, 3—4, 81

CNS, see Central nervous system

Coherence function, 72—73

Coherent averaging, 45

Colored additive noise, 130

Colored noise, 72

Comb notch lifter, 80

Communications, 154

Companding, 39

Complex cepstrum, 76—77, 80

Complex exponentiation, 76, 79

Complex frequency plane, 5

Complex logarithm, 76—77

Compression, 81, 90

Compressor, 38

Computation load, 155—157

Computer tomography (CT), 6

Conditional probability, 16—18

Conditional probability density function, 20

Constraint maximization, 123—124

Continuous random variable, 17, 19—20

Continuous signal processing, 5

Continuous signals, 4

Convolution integral, 66

Convolution theorem, 66, 115

Correlation analysis, 15, 23—26

Correlation coefficient, 23—24, 110

Correlation estimation, 56

Correlation function, 25—26

Correlation matrix, 25

Correlation vector, 129

Correlation window, 115—116

Cosine matrix, 129

Covariance, 23

Covariance lattice, 98

Covariance matrix, 25, 27

Criterion autoregressive transfer function (CAT), 97

Cross correlation function, 25

Cross-spectral density function, 72—73

Cross spectrum, 72

Cyclic convolution, 70

Cyclic convolution theorem, 70

D

Data reduction, 6

Data window, 115—118

Decimation in frequency, 70

Decimation in time, 70

Delta function, 30

Delta modulation (DM), 36

Dendrites, 9—11

Deterministic signals, 3, 5, 15

DFT, see Discrete Fourier transform

Difference equations, 43

Differential pulse code modulation (DPCM), 32, 36

Digital filters, 73—74

Digital signal processing, 29—44

discrete methods, 42—43

main tasks of, 29

quantization, see also Quantization, 29, 36—42

sampling, see also Sampling, 29—36

Digital technology, 4

Digital-to-analog conversion (D/A), 30

Dirichlet conditions, 65, 67

Discrete Fourier transform (DFT), 4, 68—70, 77,

79, 110—111

Discrete methods, 42—43

Discrete Parseval's theorem, 70

Discrete random events, 15

Discrete signal processing, 5

Discrete signals, 4—5

Discrete time processing, 29

DPCM, see Differential pulse code modulation

Durbin algorithm, 87—88, 98, 148

Durbin procedure, 87—88

Dye dilution curves, 80

E

ECG, see Electrocardiogram; Electrocardiographic signals

Echo cancellation in communication network, 149

EEG, see Electroencephalogram; Electroencephalographic signal

EEM, 106

Eigenvalue, 129—130

Eigenvector, 129—130

Electrical signal, 1

Electrocardiogram (ECG), see also Fetal ECG; Maternal ECG, 2, 150

interferences, elimination of, 150—151

nonuniform sampling process, 33—34

Electrocardiographic monitoring, 1

Electrocardiographic (ECG) signals, 3, 9, 15

Electroencephalogram (EEG), 102—103, 127, 135, 143, 150

gaussian distribution, 27

sample function, 26

spectral analysis, 109

Electroencephalographic (EEG) signal, 4, 9, 21—22

Electroencephalography, 1

Electromagnetic signals, 6

Electromyographic (EMG) signals, 9, 150

Blackman-Tukey method, 112

periodogram, 118

spectral analysis, 109

Electrosurgical interferences, elimination of, 151—152

Electrosurgical unit (ESU), 151

EMG, see Electromyographic signals

End plate, 12

Energy spectral density estimation, 132

Energy spectrum density, 67

Ensemble, 4, 21—22, 26
 Entropy, 122—123
 EP, see Evoked potential
 EPSP, see Excitatory postsynaptic potential
 Ergodicity, 26
 Ergodicity theorem, 26
 Ergodic signals, 3—4
 Events, 15
 Evoked potential (EP), 4, 45—46, 155
 Exact least square method, 142
 Excitable membrane, 10—12
 Excitatory postsynaptic potential (EPSP), 12
 Excitatory synapse, 11
 Expander, 39
 Expectations, 22—23
 Extracellular fluids, 10—11

F

Fast Fourier transform (FFT), 29, 68—70
 Fetal ECG, 6, 142—143
 time-sequenced adaptive filter applied to enhancement of, 156, 158
 Fetal heart monitoring, 6
 FFT, see Fast Fourier transform
 FHR, see Fetal heart monitoring
 Filter design theory, 5
 Filtering, 4—5, 141
 Final prediction error (FPE), 96—97
 Finite impulse response (FIR), 74
 Finite time averaging, 45—64
 availability of long records, 45
 estimation of variance and correlation, 53—56
 finite time estimation of the mean value, 45—53
 stationarity, 45
 synchronous averaging, see also Synchronous averaging 56—64
 Finite time estimation of the mean value, 45—53
 continuous case, 45—51
 discrete case, 51—53
 Finite wordlength (FWL) computations, 98
 FIR, see Finite impulse response
 First order adaptive sampling, 32—34
 Fixed parameters filter, 73
 Forgetting weights, 106
 Four sigma loading, 40
 Fourier series, 4
 Fourier transform (FT), 4—5, 29—30, 65—67, 104
 convolution theorem, 66
 Parseval's theorem, 66—67
 periodic signals, 67
 spectral estimation methods based on, 109—122
 Frequency convolution, 115—116
 Frequency domain analysis, 4—5, 65—80, 109,
 115—116
 adaptive filtering, 147
 cepstral analysis, 74, 76—77
 discrete Fourier transform, 68—70
 fast Fourier transform, 68—70

Fourier transform, see also Fourier transform, 65—67
 inverse discrete Fourier transform, 68
 inverse Fourier transform, 65
 linear filtering, 73—76
 noise cancellation, 154
 representation, 65
 spectral analysis, 71—73
 Frequency estimation bias, 124
 Frequency filtering, 5
 Frequency sampling method, 74
 FT, see Fourier transform
 Fundamental frequency, 78

G

Gastrointestinal signals, 127
 Gaussian distribution, 27
 Gaussian process, 15, 26—28
 Generalized likelihood ratio (GLR), 105—106
 General measurement and diagnostic system, 1—3

H

Hamming window, 97
 Hearing impaired, noise reduction for, 151—152
 Heart activities, 13
 Heart transplantation, 150—151
 Homomorphic filtering, 74, 77—80

I

IDFT, see Inverse discrete Fourier transform
 IFT, see Inverse Fourier transform
 Ignorance, 122
 IIR, see Infinite impulse response
 Ill conditioning, 98
 IMA, see Autoregressive integrated moving average
 Imaging, 6
 Impulse invariant method, 74
 Impulse response, 73
 Infinite impulse response (IIR), 74
 Inhibitory postsynaptic potential (IPSP), 12
 Inhibitory synapse, 11
 Input-output identification techniques, 90
 Interferences, 2
 Intracellular fluid, 9—11
 Inverse discrete Fourier transform (IDFT), 68, 79
 Inverse Fourier transform (IFT), 65
 Ion concentration changes, 9
 Ion selective electrodes, 9
 IPSP, see Inhibitory postsynaptic potential
 Iterative estimate, 89—90

J

Jittering, 158

Joint central moments, 23
 Jointly gaussian distributed, 27—28
 Joint moment, 23
 Joint probabilities, 16—18, 22
 Joint probability density function, 4, 22, 27
 Joint probability distribution function, 18
 Jump nonstationarities, 106

K

K-complexes, 7
 Kullback information, 97
 Kullback's divergence between conditional probability laws, 106

L

Lag window, 115—116
 Laplace transform, 4—5
 Laryngical disorders, 109
 Latencies estimation, 61—64
 Lattice coefficients, 98—99
 Lattice filter, 99
 Lattice inverse filter, 98
 Lattice representation, 98—99
 Lattice structure, 141—142
 Leakage, 110, 115—116, 133
 Least mean square (LMS) adaptive filter, 141, 143—148, 156

Least squares method, 85—88, 132
 Lifter, 76
 Linear filtering, 73—80
 Linearly independent variables, 24
 Linear prediction, 81—108
 Line enhancement, 141
 Line interference, 150, 158
 Line splitting, 124
 LMS, see Least mean square (LMS) adaptive filter
 Loading factor, 40
 Log likelihood ratio, 106
 Long pass lifter, 78
 Low-pass filter, 135
 LPC, 98—99

M

MA, see Moving average
 Matched filtering, 141, 158
 Maternal ECG, 142—143, 158
 Maximum entropy, 109
 Maximum entropy method (MEM), 109, 122—125
 Maximum entropy spectral estimation (MESE), 122
 Maximum entropy spectrum analysis (MESA), 122
 Maximum likelihood (ML), 90, 93—95, 97, 109
 comparison with other methods, 136
 spectral estimation, 109, 133—134
 Maximum likelihood method (MLM), see Maximum likelihood

Maximum likelihood spectral estimation (MLSE), 133
 Mean, 22—23
 Mechanical signals, 6
 MEM, see Maximum entropy method
 Menstrual secretions, 101
 MESA, see Maximum entropy spectrum analysis
 MESE, see Maximum entropy spectral estimation
 Midtread uniform quantizer, 37
 Midviser uniform quantizer, 38
 Mixed autoregressive moving average (ARMA) models, 90—95

 direct method, 90—93
 maximum likelihood method, 93—95
 ML, see Maximum likelihood
 Motor nerve, 12
 Moving average (MA) model, 74, 83—85, 89—90
 comparison with other methods, 136
 spectral estimation, 109, 125—126
 Multichannel adaptive filter, 141
 Multichannel adaptive signal enhancement, 154—156
 Multivariate gaussian process, 27—28
 Muscle, 12
 Muscle fatigue, 109
 Mutually exclusive events, 16
 Myelin sheath, 10
 Myosin, 12

N

n-bit quantizer, 38
 Negentropy, 97
 Nernst equation, 10
 Nerve cell, 9—12
 Nerve trunk, 10
 Neural information transfer, 9
 Neural network, 9
 Neurological disorders, 109
 Neuromuscular junctions, 10, 12
 Neuron, see also Nerve cell, 9
 Neurophysiological studies, 13
 Neurophysiology, 9
 Neurosignals, 7
 Noise cancellation, 141, 147—154, 158
 adaptive line enhancer, 154—155
 ECG interferences, 150—151
 electrosurgical interferences, 151—152
 frequency domain, 154
 hearing impaired, 151—152
 line interference, 150
 with reference input, 148—152
 without reference input, 152—154
 Noninvasive techniques, 6
 Nonperiodic signals, 3—4
 Nonrecursive algorithm, 142
 Nonstationary processes, 99—101
 Nonstationary signals, 3—4, 81, 84, 156
 Nonuniform sampling, 31—36
 Nonuniform sampling encoder (NSE), 34

Normal distribution, 37
 Normal equations, 86
 Normalized correlation function, 25
 Normalized covariance, 24
 N-pole complex sinusoidal signal, 124
~~nth~~ central moment, 23
~~nth~~ moment, 22
 Nuclear magnetic resonance (NMR), 6
 Nyquist rate, 30—31

O

Off-line processing, 29
 One-dimensional signals, 6
 One-sided Z transform, 43
 On-line processing, 29
 Optimal filters, 73
 Order estimation, see Process order estimation
 Outcome of experiment, 21

P

Parameter identification, 142—143
 Parametric families of distributions, 106
 Parametric model, 81
 PARCOR, see Partial correlation coefficients
 Parseval's theorem, 66—67
 Partial correlation coefficients (PARCOR), 88, 98—99
 Performance index, 142, 145
 Periodic signals, 3—4
 Periodograms, 109—110, 112—122
 advantage of, 114
 comparison with other methods, 136
 expected value, 114—116
 smoothing, 119—122
 variance, 116—117
 weighted overlapped segment averaging, 117—121

Phase unwarping, 77
 PHD, see Pisarenko's harmonic decomposition
 Piecewise-stationary, 101, 103
 Pisarenko harmonic decomposition (PHD), 109, 127—130
 comparison with other methods, 136

Pitch, 78
 Point processes, 7
 Pole model, 83
 Pole-zero model, 83
 Positive definite, 130
 Postsynaptic region, 12
 Power cepstrum, 76—77
 Power spectral density (PSD) function, 5, 71—72, 109
 Power vector, 129
 PQRST complex, 2
 PRBN, see Pseudo random binary noise
 Prediction, 75
 Premature ventricular contractions (PVC), 79

Preprocessing of signal, 135
 Presynaptic region, 11
 Prewhitenning filters, 135
 Primary input, 144, 154
 Probability density functions, 19—20, 26, 41—42
 Probability distribution functions, 18—19, 22, 27
 Probability of event, 15, 17—18
 Probability theory, 15—20
 Processing of signal, 1—2, 4—6
 Process order estimation, 95—98
 Prony's methods, 109, 130—132
 comparison with other methods, 136
 PSD, see Power spectral density function
 Pseudo random binary noise, 111

Q

QRS complex, 3, 7, 15, 151, 158
 Quantization, 29, 36—42
 analysis of noise, 39—40
 block quantizer, 36
 companding, 39
 compressor, 38
 expandor, 39
 four signs loading, 40
 loading factor, 40
 midtread uniform quantizers, 37
 midviser uniform quantizer, 38
 n-bit quantizer, 38
 probability density function, 41—42
 rough, 40—42
 sequential quantizers, 36
 signal-to-noise ratio, 40
 zero memory quantization, 36—39
 Quantization theorem, 41
 Quefrency, 76—77, 79—80

R

Random processes, 15—28
 correlation analysis, 23—26
 ensamble, 21—22
 gaussian process, 26—28
 probability theory, see also Probability theory, 15—20
 random signals, 21—23
 stationary in the wide sense, 26
 Random signals, 3—5, 21—23
 Random variables, 17—18, 21
 Real-time processing, 29
 Recognition, 6
 Reconstruction, 3
 Records alignment, 61—64
 Recursive algorithm, 142
 Redundancies, 2
 Reference inputs, 143, 145—146
 noise cancellation with, 148—152
 noise cancellation without, 152—154
 Reference window, 103—105

Reflection coefficients, 88
 Regression line, 23—24
 Relative frequency, 15
 Relative refractory period, 11
 Residual, 86
 Residuals correlation, 95
 Residuals flatness, 95—96
 Resting potential, 11
 Rough quantization, 40—42
 R-R interval, 3, 15
 Run length encoding, 34—36

S

Sample function, 21—22, 26
 Sampling, 29—36
 nonuniform, see also Nonuniform sampling, 31—36
 uniform, 30—31
 Sampling theorem, 30, 68
 SA node, 150
 Seasonal processes, 101
 Second differences method, 34
 Second order adaptive sampling, 32—34
 Segmentation, see also Adaptive segmentation, 1—2, 135

SEM, see Spectral error measure (SEM) method
 Sequential quantizers, 36
 Short pass lifter, 78
 Sidelobes, 109, 115—116, 133
 Signal correction, 143—144
 Signal distortion, 155
 Signal enhancement methods, 2
 Signal estimation, 142—143
 Signal nonstationarity, see Nonstationary signals
 Signal-to-noise ratio (SNR), 40, 45, 117, 141
 Singular values decomposition (SVD), 98, 127
 Skeletal muscle cells, 12
 Sleep states, 109
 Sliding windows, 102—105
 Smoothing, 75, 80
 periodogram, 119—122
 Sonar detection, 154
 Sonar tracking, 154
 Spectral analysis, 71—73, 109, 122—125
 Spectral error measure (SEM) method, 103—106
 Spectral estimation, 109—139
 autoregressive method, 109, 122—125
 autoregressive moving average methods, 109, 126—132
 Blackman-Tukey procedures, 109—113
 Fourier transform, methods based on, 109—122
 maximum likelihood, 109, 133—134
 moving average method, 109, 125—126
 periodogram, 109—122
 Pisarenko harmonic decomposition, 109, 127—130
 Prony's methods, 109, 130—132
 Spectrum, 71, 109
 Speech processing, 101, 135

Speech signals, 106
 Spindles, 7
 Standard deviation, 23
 Stationarity, see also Wide sense stationarity, 45
 Stationary modeling techniques, 100
 Stationary process, 3—4, 22
 Statistical averages, 22—23
 Statistically independent events, 17
 Statistically independent random processes, 22
 Steepest descent, 146
 Stochastic signal, 15
 Storage, 2—3, 6
 SVD, see Singular valued decomposition
 Synapse, 10—12
 Synaptic cleft, 11
 Synchronous averaging, 45, 56—64
 estimation of latencies, 61—64
 general case, 60—61
 records alignment, 61—64
 statistically independent responses, 58—59
 totally dependent response, 59
 Syntactic analysis, 7
 Syntactic methods, 106

T

Theoretic decision approach, 7
 Time average, 26
 Time cross correlation, 26
 Time domain, 4
 Time invariant filter, 73
 Time-sequenced adaptive filtering, 141, 156—158
 Time series analysis, 6, 81—109
 Time varying filters, 73
 Time window, 1
 Toeplitz matrix, 87
 Total refractory period, 11
 Transducer, 1—2, 6
 Transfer function (TF), 81—82
 Transient signal, 3—4
 Transmission, 2—3, 6
 Tremors, 109
 Trend-like nonstationarities, 102, 106
 Trend nonstationarity, 99—100
 Triangular smoothing window, 122
 Triangular weighting window, 111
 T-wave, 79
 Two-dimensional signals, 6
 Two points projection method, 32

U

Ultrasonic imaging, 6
 Uncorrelated independent variable, 24
 Uniform sampling, 30—31

V

Variance, 23

Variance estimation, 53—56
continuous case, 53—54
discrete cases, 54—56
VER, see Visual evoked response
Visual evoked response (VER), 79
Voltage triggered method, 32
Volume conductors, 12—13

W

Walsh spectral estimator, 134
Wavelet, 2, 7, 61, 80, 156
Weighted overlapped segment averaging (WOSA),
117—121, 124
Weighting sequence, 93
Weighting windows, 110
Whitening filter, 85, 90
White noise, 72
Wide sense stationarity, 26, 28
Widrow-Hoff LMS algorithm, 146, 156
Wiener filter, 73—76
Wiener-Hopf condition, 75
Wiener-Hopf equation, 146
Wiener-Khinchin relations, 71, 109—110, 112
Wiener-Khinchin theorem, 111

Window bias, 135
Windowing effects, 97
Woody filtering, 61
WOSA, see Weighted overlapped segment
averaging

X

X-ray analysis, 6

Y

Yule-Walker equations, 86, 93, 97—98, 127

Z

Z domain, 5
Zero memory quantization, 36—39
Zero order adaptive sampling, 32—34
Zero order hold (ZOH), 30
Zero padding, 110
Z transform, 4, 42—43, 77, 104

