AFWAL-TR-85-3080

AD-A168 174

COMPENDIUM OF METHODS FOR APPLYING MEASURED DATA TO VIBRATION AND ACOUSTIC PROBLEMS

Edited by
R. G. DeJong

Cambridge Collaborative, Inc
689 Concord Avenue
Cambridge, MA 02038

October 1985

Final Report for Period August 1980 - July 1985

Approved for public release; distribution is unlimited.

DTIC
ELECTE
MAY 2 7 1986
S D
D

COPY FILE COPY

FLIGHT DYNAMICS LABORATORY
AIR FORCE WRIGHT AERONAUTICAL LABORATORIES
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

86

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever; and the fact that the government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture use, or sell any patented invention that may in any way be related thereto.

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

This technical report has been reviewed and is approved for publication.

*Phyllis G. Bolds

Phyllis G. Bolds
Physicist
Data Analysis Group

Davey L. Smith

Davey L. Smith, Chief
Structural Vibration & Acoustics Branch
Structures and Dynamics Division

FOR THE COMMANDER

JAMES J. OLSEN
Assistant for Research & Technology
Structures & Dynamics Division

If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization please notify AFWAL/FIBG ***, W-PAFB, OH 45433 to help us maintain a current mailing list.***

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY N/A	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) 23016-85-1	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFWAL-TR-85-3080
6a. NAME OF PERFORMING ORGANIZATION Cambridge Collaborative, Inc. — 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION (AFWAL/FIBG) Flight Dynamics Laboratory AF Wright Aeronautical Laboratories
6c. ADDRESS (City, State and ZIP Code) 689 Concord Avenue Cambridge, MA 02138-1002	7b. ADDRESS (City, State and ZIP Code) WPAFB, OH 45433
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Flight Dynamics Laboratory — 8b. OFFICE SYMBOL (If applicable) AFWAL/FIBG	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F33615-80-C-3200
8c. ADDRESS (City, State and ZIP Code) WPAFB, OH 45433	10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
62201F	2401	240104	24010412

11. TITLE (Include Security Classification) See reverse	
12. PERSONAL AUTHOR(S) R.G. DeJong, Editor	
13a. TYPE OF REPORT FINAL	13b. TIME COVERED FROM Aug 1980 TO July 1985
14. DATE OF REPORT (Yr., Mo., Day) October 1985	15. PAGE COUNT 463
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES		
FIELD	GROUP	SUB. GR.
01	01	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Vibration, Sound, Noise; Acoustic Data, Data Analysis; Acoustic Measurements; Vibration Prediction; Dynamic Modeling;

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The scope of this report includes the measurement, analysis and use of vibration and acoustic data. The purpose of this report is then two-fold. First, it provides introductory material in an easily understood manner to engineers, technicians, and their managers in areas other than their specialties relating to the measurement, analysis and use of vibration and acoustic data. Second, it provides a quick reference source for engineers, technicians and their managers in the areas of their specialties relating to the measurement, analysis and use of vibration and acoustic data.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED

22a NAME OF RESPONSIBLE INDIVIDUAL	22b TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Phyllis G. Bolds	(513) 255-5200	AFWAL/FIBG

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

Block 11: Title

COMPENDIUM OF METHODS FOR APPLYING MEASURED DATA TO VIBRATION AND ACOUSTIC PROBLEMS

FOREWORD

This report was prepared by the staff of Cambridge Collaborative, Inc., R. G. DeJong, Editor, under contract No. F 33615-80-C-3200, entitled, "Compendium Methods for Applying Measured Data to Vibration and Acoustic Problems." Robert Merkle was the Contract Monitor.

Those who contributed to the writing of the report are as follows:

R.G. DeJong
J.E. Manning
R.E. Powell
D.H. Keefe
G.M. Glass

These authors wish to thank M. Manning, N. Vaccarello, and C. McLean for their assistance in preparing this report.

Accesion For
NTIS CRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution /
Availability Codes
Dist | Avail and / or Special
A-1
QUALITY INSPECTED 3

CHAPTER 1. INTRODUCTION

CHAPTER 2. MEASUREMENT TECHNOLOGY

CHAPTER 3. DIGITAL DATA ANALYSIS TECHNOLOGY

CHAPTER 4. TEST PROGRAM DESIGN

CHAPTER 5. ESTIMATION OF SYSTEM CHARACTERISTICS

CHAPTER 6. IDENTIFICATION OF TRANSMISSION PATHS

CHAPTER 7. IDENTIFICATION OF SOURCES

CHAPTER 8. VIBRATION RESPONSE PREDICTION

CHAPTER 1. INTRODUCTION

This report is a compendium of methods for applying measured data to vibration and acoustic problems. The scope of this report includes the measurement, analysis and use of vibration and acoustic data. In covering this broad scope it is the intent of this report to provide an assembly of information not found in any other single volume. The information assembled here is of an overview nature, providing the basics of the various technologies covered, while the reader is given a list of references that give more detailed information.

The purpose of this report is then two-fold. First, it provides introductory material in an easily understood manner to engineers, technicians and their managers in areas other than their specialties relating to the measurement, analysis and use of vibration and acoustic data. Second, it provides a quick reference source for engineers, technicians and their managers in the areas of their specialties relating to the measurement, analysis and use of vibration and acoustic data.

In order to serve this purpose this report has been organized more like a handbook than a text book. Each chapter is designed to be as much self standing as possible with references to specific sections of other chapters used when needed. In each section the basic ideas and methods of implementing the technology are given first. This is followed by a discussion of the underlying principles or implications where necessary. References are given to other material where more detailed information can be found. Illustrations are usually given of the application of the technology to a practical problem.

At the beginning of each chapter is a detailed outline of the contents of the chapter in addition to the more limited Table of Contents at the beginning of this report. Also included at the beginning of each chapter are lists of the figures, tables and symbols used in the chapter. At the end of each chapter is a list of references included there in.

The following paragraphs briefly describe the contents of each chapter of this report.

Chapter 2 presents an overview of the measurement technology of vibration and acoustic data. It is divided into four major sections dealing with sensing devices (transducers), signal conditioning devices (amplifiers, filters, etc.), recording devices, and system design. Each of these sections covers the general requirements of that category of devices as well as specific information about the selection and use of the different types of devices.

Chapter 3 presents a summary of various methods of analyzing measured data. It is assumed that the analysis is done digitally. Analog processing methods are covered under signal conditioning in Chapter 2. Chapter 3 is divided into four major sections dealing with digital sampling, single time series analysis, dual time series analysis, and multiple time series analysis. Common elements in the last three sections are statistical measures, correlation analysis, and spectral analysis.

Chapter 4 outlines the essential ingredients in a test program. They are the problem description, program objective, data requirements, test plan, and documentation of results. This chapter serves as a division between the presentation of the basic tools used in the measurement and analysis of vibration and acoustic data (Chapters 2 and 3) and the presentation of the methods of interpreting and applying the data to solve problems (Chapters 5 - 8). It is placed here to emphasize the interrelation between the gathering and manipulation of data on the one hand and the intended use of the data on the other hand.

Chapter 5 outlines some of the methods that use measured data to estimate the characterics of a system. The main focus here is on the use of the frequency response function. This chapter includes both the direct and indirect methods of estimating the frequency response function. It also outlines a number of methods used to estimate the damping of a system.

Chapter 6 outlines some of the methods that use measured data to identify the transmission paths of vibration and acoustic energy in a system. Methods for systems with both nondispersive and dispersive wave propagation are given. The methods include both time and frequency domain analyses.

Chapter 7 outlines some of the methods that use measured data to identify the sources of vibration and acoustic energy in a system. The methods include the use of the input/output model and the transmission model for systems with single or multiple sources.

Chapter 8 presents a number of methods for predicting the vibration response of a system. They include the propagating wave model, transmission line analysis, modal analysis, statistical energy analysis, finite element models, transfer function models, and extrapolation techniques. Although not all of these use measured data in themselves, they are often involved in the prediction or interpretation of measured data. Specifically these prediction methods are often used to obtain system models needed for the methods outlined in Chapters 5 - 7.

CHAPTER 2. MEASUREMENT TECHNOLOGY

By

J. E. MANNING
R. E. POWELL

CHAPTER 2. MEASUREMENT TECHNOLOGY

TABLE OF CONTENTS

	Page
LIST OF FIGURES	2-v
LIST OF SYMBOLS	2-vii
2.1 SENSING DEVICES	2-1
2.1.1 Categorization	2-1
2.1.1.1 Vibration Transducers	2-2
2.1.1.2 Pressure Transducers	2-4
2.1.2 Principles of Operation	2-4
2.1.2.1 Capacitance Type	2-5
2.1.2.2 Resistive Type	2-5
2.1.2.3 Piezoelectric Type	2-6
2.1.2.4 Inductive Type	2-6
2.1.2.5 Other	2-6
2.1.3 General Requirements	2-7
2.1.3.1 Sensitivity	2-8
2.1.3.2 Dynamic Range	2-9
2.1.3.3 Frequency Response	2-9
2.1.3.4 Physical Properties	2-10
2.1.3.5 Environmental Effects	2-10
2.1.3.6 Signal Conditioning Requirements	2-10
2.1.3.7 Calibration	2-11
2.1.4 Accelerometers	2-11
2.1.4.1 Types and General Use	2-11
2.1.4.2 Piezoelectric Accelerometers	2-13
2.1.4.3 Piezoresistive Accelerometers	2-21
2.1.4.4 Selection Criteria	2-22

2.1.4.5 Calibration..... 2-24
2.1.4.6 Mounting Procedures..... 2-26
2.1.4.7 Specialized Units..... 2-27

2.1.5 Microphone and Pressure Transducers 2-28

2.1.5.1 Condenser Microphone..... 2-30
2.1.5.2 Electret Microphone..... 2-35
2.1.5.3 Piezoelectric Pressure Transducer.... 2-35
2.1.5.4 Piezoresistive Pressure Transducer... 2-37

2.1.6 Strain Gages 2-38

2.1.6.1 General Types and Uses..... 2-39
2.1.6.2 Selection Procedure..... 2-43
2.1.6.3 Calibration..... 2-46
2.1.6.4 Bonding and Surface Preparation..... 2-46
2.1.6.5 Special Uses..... 2-47

2.1.7 Other Special Purpose Transducers 2-47

2.2 SIGNAL CONDITIONING EQUIPMENT 2-49

2.2.1 General Requirements 2-49

2.2.2 Identification of Basic Elements 2-50

2.2.3 Conditioning for Piezoelectric Transducers ... 2-50

2.2.4 Conditioning for Piezoresistive Transducers... 2-55

2.2.5 Conditioning for Condenser and Electret Microphone and Other Capacitive Devices 2-57

2.2.6 Conditioning for Strain Gages 2-60

2.2.7 Amplifiers 2-64

2.2.7.1 AC/DC Amplifier..... 2-67
2.2.7.2 Amplifier Inputs..... 2-67
2.2.7.3 Frequency Response and Input/Output Impedance..... 2-68
2.2.7.4 Special Features..... 2-70
2.2.7.5 Overload Performance..... 2-70
2.2.7.6 Line and Battery Operation..... 2-70
2.2.7.7 Calibration Procedures..... 2-71
2.2.7.8 Data Monitoring..... 2-72
2.2.7.9 Cabling and Connectors..... 2-72

2.2.8 Filters 2-73

2.2.8.1 Types Available..... 2-73
2.2.8.2 Passive..... 2-75
2.2.8.3 Active..... 2-79

2.2.9 Encoders 2-82

2.2.9.1 Introduction..... 2-82
2.2.9.2 Purpose of Encoding..... 2-82
2.2.9.3 Multiplexers..... 2-84
2.2.9.4 Continuous Wave Modulation..... 2-85
2.2.9.5 Pulse Modulation..... 2-87
2.2.9.6 Analog-to-Digital Converters..... 2-90

2.2.10 Cabling Requirements 2-93

2.2.11 Transmitters..... 2-97

2.3 RECORDING DEVICES 2-99

2.3.1 General Requirements 2-99

2.3.2 Permanent Recording Devices.....2-100

2.3.2.1 Analog Magnetic Tape Recorders 2-100
2.3.2.2 Graphical Hard Copy 2-106
2.3.2.3 Digital Recording Devices 2-109

2.3.3 Transient Recorders.....2-110

2.4 SYSTEM DESIGN 2-113

2.4.1 General Requirements 2-113

2.4.2 Design Procedures 2-114

2.4.2.1 Accuracy Requirements.....2-116
2.4.2.2 Number of Data Channels.....2-117
2.4.2.3 Frequency Bandwidth.....2-119
2.4.2.4 Dynamic Range.....2-122
2.4.2.5 Noise.....2-124
2.4.2.6 Linearity.....2-126
2.4.2.7 Interference.....2-127
2.4.2.8 System Design Summary.....2-127

REFERENCES.....2-129

LIST OF FIGURES

Figure		Page
2-1.	Seismic Vibration Transducer.....	2-3
2-2.	Amplitude and Phase Response for Simple Seismic Transducer.....	2-14
2-3.	Compression-Type Piezoelectric Transducer.....	2-15
2-4.	Simplified Charge Source Equivalent Circuit.....	2-17
2-5.	Simplified Voltage Source Equivalent Circuit.....	2-18
2-6.	Air Condenser Microphone.....	2-31
2-7.	Representation of Condenser Microphone.....	2-33
2-8.	Electret Condenser Microphone.....	2-36
2-9.	Foil Strain Gage.....	2-40
2-10.	Temperature Compensation for Strain Gage.....	2-42
2-11.	Strain Gage Test Set-up.....	2-48
2-12a.	Signal Conditioning with a Voltage Amplifier.....	2-51
2-12b.	Signal Conditioning with a Charge Amplifier.....	2-51
2-13.	Signal Conditioning for a Piezoresistive Transducer.....	2-56

2-14. Signal Conditioning Circuit for Condenser Microphone..... 2-59

2-15. Potentiometer Circuit for Strain Gages..... 2-61

2-16. Bridge Circuit for Strain Gages..... 2-63

2-17a. Passive Low Pass Filter..... 2-76

2-17b. Passive High Pass Filter..... 2-76

2-17c. Passive Low Pass Filter Used With Transducer..... 2-78

2-17d. Passive High Pass Filter Used with Transducer..... 2-78

2-18. General Circuit for Signal Conditioning..... 2-94

2-19. Bandwidth and Noise for Various Tape Speeds..... 2-101

2-20. Frequency Response and Noise for Direct Recording..... 2-105

SYMBOLS

A to D	Analog to digital
ADC	Analog to digital converter
AM	Amplitude modulation
$A(\omega)$	Amplitude of frequency response
a	Acceleration
C	Capacitance
CMRR	Common mode rejection ratio
^{o}C	Degrees Celsius
dB	Decibel
d/dt	Differentiation with respect to time
E_o	Output voltage
E_p	Preamplifier output voltage
^{o}F	Degrees Fahrenheit
f	Frequency-Hertz
f_c	Cut-off frequency
FET	Field effect transistor
FM	Frequency modulation
ft	Feet
g	Acceleration of gravity
$H(\omega)$	Amplitude response
Hz	Hertz
ips	Inches per second
j	$\sqrt{-1}$
k	Gage factor
kHz	Kilo Hertz
L	Length
LSB	Least significant bit
μ	Micro 10^{-6}
μV	Micro Volts
MHz	Mega Hertz

m	Meters
ma	Milliamp
mm	Millimeters
mV	Milli Volts
ns	Nanosecond
Pa	Pascal
PAM	Pulse amplitude modulation
Pc	Pico-Coulomb
PCM	Pulse code modulation
Pf	Pico-Farad
PM	Phase modulation
psi	Pounds per square inch
PR	Piezoresistive
psig	Psi gage
Q	Output charge
R	Resistance
rms	Root mean square
s	Seconds
sec	Second
V	Volts
VCA	Voltage controlled amplifier
Z	Impedance
ω	Radian frequency
%	Percentage
Ω	Ohms
~	Approximately
π	Pi
"	Inches
λ	Wavelength
$\theta(\omega)$	Phase of frequency response

CHAPTER 2. MEASUREMENT TECHNOLOGY

This chapter of the report presents information on vibration and acoustic measurement systems. The chapter is divided into four major sections dealing with sensing devices, signal conditioning, recording devices, and system design.

2.1 SENSING DEVICES

The first item to be considered in designing a vibration and acoustic measurement system is the transducer. In general terms this is a device that converts the physical variable that we wish to measure into a signal that can be observed or recorded for later data processing. This section of Chapter 2 is limited to information on vibration and acoustic transducers. Later sections will deal with conditioning the electrical signals provided by the transducers and other aspects of measurement system design.

2.1.1 Categorization

The most commonly used transducers for vibration and acoustic measurements are

1) accelerometers
2) microphones (and other pressure transducers), and
3) strain gages

These will be discussed in detail in this chapter. We will focus on devices that are practically useful for vibration and acoustic problems. This means that they can measure extremely small vibratory motions and pressures over a very wide range of frequencies.

A large variety of transducers are available. These vary with regard to the physical quantity that is measured, the principal of operation, the type of output signal, sensitivity and dynamic range, frequency response, physical characteristics, and durability. However, within the limitations of vibration and acoustic measurements, we can categorize transducers into two major groups: transducers that sense motion (vibration) and transducers that sense force (or pressure).

2.1.1.1 Vibration Transducers

Vibration transducers can be divided into three groups: "fixed reference" types, "differential" types, and "inertial" or "seismic" types. In the fixed reference transducer the motion at one point is measured relative to the motion at a fixed reference point. Examples include certain capacitance and potentiometer-type displacement transducers and several optical devices where the optical transducer is mounted at a fixed point near, but not in contact with, the vibrating structure. The use of a fixed reference transducer is, of course, limited to applications in which the required fixed reference point is available. In practice the use of this type of transducer is usually limited to a laboratory setting.

The differential vibration transducer provides a measurement of the motion at one point on a vibrating surface relative to another. The most common example is a strain gage. Transducers measuring an angular motion might also be termed differential transducers. The differential transducer is sometimes limited by a low signal-to-noise ratio, since it must measure the difference between two motion variables that are approximately equal and quite small.

The third type of vibration transducer is the inertial or seismic transducer. This device can be represented by a mass-spring system as shown in Figure 2.1. The transducer is fixed to

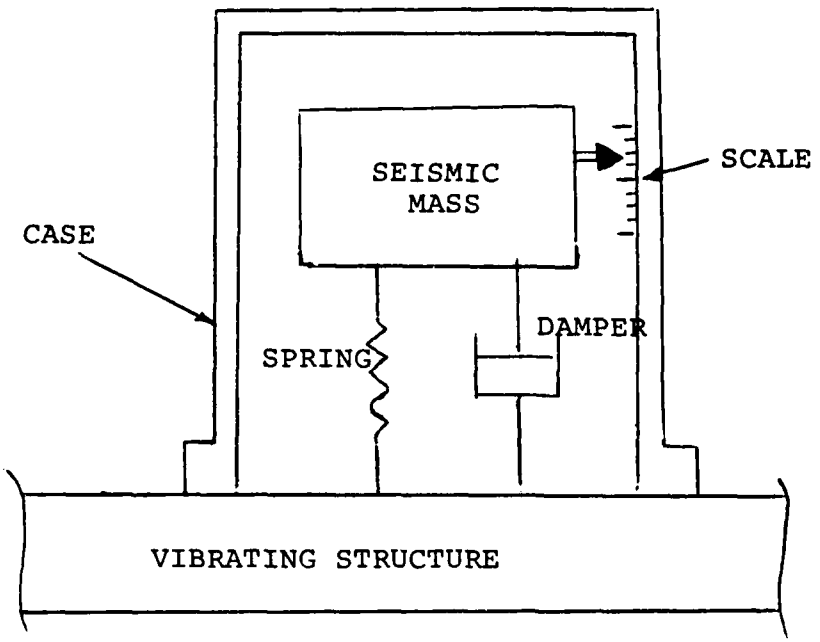


Fig. 2.1 Seismic Vibration Transducer

the vibrating structure, and the motion at the attachment point is inferred from the motion of the mass relative to the case. Depending on the mass-spring resonance frequency, the damping, and the frequency range of interest this type of transducer can be used to measure acceleration, velocity, or displacement.

2.1.1.2 Pressure Transducers

Pressure transducers, force transducers, and microphones are fundamentally the same. The differences are in application. A pressure transducer is typically used to measure the pressure in a fluid or gas while the force transducer is used to measure the force applied to a structure. A microphone is a pressure transducer used to measure acoustic pressures in air. Another type of transducer not mentioned above is the hydrophone. This transducer is used to measure acoustic pressures in water or a fluid.

2.1.2 Principals of Operation

Although several mechanical devices exist to measure vibration, [2.1] they are important only as a means to record peak levels during transportation. These so-called shock recorders and vibrating reed devices will not be discussed further.

Transducers of practical importance for vibration and acoustic measurement systems are electro mechanical devices. A variety of materials and physical principals are used to convert a motion or force into an electrical signal. A summary of these devices follows.

2.1.2.1 Capacitance Type

The capacitance between two plates is measured. In a vibration transducer of this type one plate is attached to the vibrating structure and the other is held fixed. The displacement of the vibrating plate causes a change in distance between the plates and a change in capacitance. The requirement for a fixed reference point and the nonlinearity between capacitance and plate separation distance preclude use of this type of device for most applications.

A second capacitance device that is commonly used is the condensor microphone. In this transducer a thin diaphram serves as one plate of the capacitor. At frequencies below its resonance frequency, the displacement of the diaphram is proportional to the applied pressure. Thus, a measurement of the acoustic pressure is obtained by measuring the change in capacitance. Advantages include high sensitivity and wide dynamic range.

2.1.2.2 Resistive Type

In this type of transducer the physical variable to be measured causes a change in resistance across the transducer leads. The most common example is a strain gage in which the strain applied to fine wires in the gage causes a change in resistance. The strain gage can be used directly to measure strain on a structure or as part of a transducer such as a strain gage accelerometer. Transducers using piezoresistive materials also fall within this category. In such a material an applied stress produces a change in resistance. Thus, a piezoresistive pressure transducer or accelerometer can be designed. These piezoresistive materials provide a much higher sensitivity than conventional wire strain gages, have much better signal to noise ratios, and therefore, are commonly used.

The major advantage of the resistance-type transducer is that it can have a good frequency response down to DC while still having a good frequency response at high frequencies.

2.1.2.3 Piezoelectric Type

Certain materials produce an electric charge that is proportional to an applied stress. Use of these piezoelectric materials in vibration and pressure transducers is common, and a large variety of different designs and manufacturers are available. Most commonly, a piezoelectric crystal as used as the spring element in an inertial transducer. Since the material stiffness is high, the resonance frequency of the spring-mass system is high. This allows the transducer to provide an accurate measure of the sensed dynamical quantity over a wide range of frequencies. Such a transducer is called a piezoelectric or crystal transducer. As a general purpose transducer for vibration measurements piezoelectric accelerometers are unequalled. Their only serious limitation is an inability to measure DC accelerations. However, since their frequency response can extend down to a fraction of one Hertz, they can be used for most applications.

2.1.2.4 Inductive Type

Inductive transducers rely on converting the variable to be measured into a change in the magnetic flux of an electric circuit. In the self-generating type, an electromotive force is generated as a result of a conductor moving through a magnetic field.

2.1.2.5 Other

In some measurement problems it is necessary to use sensing devices having unique characteristics. Often these specialized

transducers use one of the operating principals discussed above. However, two additional principles of operation should be mentioned. These are optical systems and reflected wave systems.

In the optical system a beam of light is used as the sensing element. By either detecting the intensity of the beam or by using the interference between a reference beam and a beam reflected from the test surface, it is possible to determine vibratory displacement or velocity.

In the reflected wave system a microwave or ultrasonic beam is used as the sensing element. By detecting the interference pattern between the incident and reflected beams or by measuring the time delay for beam pulses that are reflected from the test surface it is possible to measure vibratory displacement.

Both the optical and reflected wave systems allow a measurement to be made without contacting the surface. However, they require a fixed reference frame and are usually limited to low frequencies.

2.1.3 General Requirements

The general requirement for a transducer is that it provide an output signal that accurately represents the physical variable that we wish to measure. This general requirement can be expressed in terms of the following factors:

1) sensitivity
2) dynamic range
3) frequency response
4) physical properties
5) environmental effects
6) signal conditioning requirements
7) calibration

These factors will be discussed in general in the paragraphs below. Later sections of this chapter will deal with detailed information for specific types of transducers.

2.1.3.1 Sensitivity

The transducer sensitivity is expressed as the ratio of its electrical output to the input variable it is to measure. For example, the sensitivity of an accelerometer could be expressed as volts per meter per second squared ($V/m/s^2$) or millivolts per g of acceleration (mV/g). Passive transducers, in which the output is a change in electrical characteristics such as capacitance or resistance, require an external supply voltage or current. Their sensitivity is expressed either with a specified supply voltage (or current) or as a ratio of the sensitivity to the supply voltage. For example, most condenser microphones are designed for a specific supply voltage, typically 200V. The sensitivity of these microphones is expressed in millivolts per pascal (mV/Pa) with the supply voltage required to be 200V. On the other hand, the sensitivity of a piezoresistive accelerometer is typically expressed as millivolts per g per volt (mV/g/V). It is implied that the sensitivity of such a transducer can be changed by changing the supply voltage.

Although it is not always specified, the sensitivity of a transducer is often determined under dynamic conditions. In this case it is important to note that the output and input may not be in similar terms. For example, although it is not common, it is possible to specify the sensitivity as an rms voltage output per peak to peak acceleration input.

Transducers are classed in families with a range of sensitivities. Trade-offs exist between sensitivity, dynamic range, and frequency bandwidth.

2.1.3.2 Dynamic Range

The dynamic range of a transducer defines the range of amplitudes over which the transducer provides an accurate output signal. On the high end, the maximum input to the transducer is limited by nonlinearities and the resulting signal distortion. In some designs the maximum limit is defined by limitations in the structural strength of the transducer. In such a case, the maximum limit may be lower for continuous signals than for short transient signals or vibration shocks. In addition, the limit for positive and negative signals can be significantly different.

At the lower end the transducer output is typically limited by noise either from the transducer itself or from the associated signal conditioning equipment. In either case the noise is added to the output signal making it difficult or impossible to identify the output signal. The noise is usually specified as a root mean square value.

2.1.3.3 Frequency Response

The frequency response defines the range of frequencies over which the transducer can be used to obtain accurate data. It is typically specified as an amplitude response as a function of frequency and a phase response. The amplitude response is simply a plot or tabulation of the transducer sensitivity as a function of the frequency of the input signal. The phase response is the number of degrees of phase lag or lead between the input and the output. For most transducers the frequency range is set by the frequencies between which the amplitude response is within a certain limit, i.e. ±10% or ±0.5 dB, etc. The phase response is often not specified. However it can be of importance in measurements of the time waveforms for transient signals. Although the phase response of the ideal transducer has no lag or lead (0^0 phase), a linear phase over the frequency range of

interest is acceptable since it corresponds to a simple time delay of the signal without distortion.

2.1.3.4 Physical Properties

The size and weight of a transducer are usually important in a particular application since it is necessary that the transducer not affect the variable being measured. Often a trade-off exists between the transducer size and its sensitivity and dynamic range. Thus, although a high sensitivity and wide dynamic range are always desired characteristics, the resulting transducer size and/or weight may not be acceptable.

In determining the size and weight of a particular transducer it is important to include the cable connector.

2.1.3.5 Environmental Effects

Transducers may be exposed to a large variety of environments. The most important environmental factor for vibration and acoustic transducers is temperature. In many cases the transducer sensitivity is a function of temperature. Other environmental factors to be considered are vibration (in the case of a pressure transducer), fluctuating pressures (in the case of a vibration transducer), humidity, magnetic fields, and nuclear radiation.

2.1.3.6 Signal Conditioning Requirements

The requirement that a transducer be small and lightweight precludes the incorporation of all but the simplest signal conditioning equipment within the transducer itself. Therefore, in selecting a transducer some consideration must be given to its signal conditioning requirements. Since the required electronics usually cost more than the transducer itself a selection is often based on available electronics. Considerations include the required input impedance and gain, the requirement for a voltage

or current power supply, the required frequency response range, and the required dynamic range (see Section 2.2 for detailed information).

2.1.3.7 Calibration

Most transducers are supplied with a calibrated sensitivity. Although this sensitivity should be time-invariant, it is beneficial to calibrate the transducers before they are used in a measurement system. It is also convenient to have a system level calibration. This type of calibration is usually achieved by a known insert voltage source at the input to the signal conditioning equipment. Although this calibration does not apply to the transducer it serves to calibrate the other components of the system.

2.1.4 Accelerometers

2.1.4.1 Types and General Uses

A transducer that measures the acceleration of a vibrating surface is generally referred to as an accelerometer.

The primary advantage in using the accelerometer for vibration measurements is its capability to provide accurate data over a wide range of amplitudes and frequencies. Although a number of transducers have been designed to provide a measure of vibratory velocity or displacement, the accuracy of the accelerometer is such that in many measurement programs the output of the accelerometer is electronically integrated to obtain velocity and displacement.

In its most basic form it is a seismic device that can be represented by a mass-spring oscillator as shown in Figure 2.1. At frequencies well below the resonance frequency of the oscillator, the displacement of the mass relative to the case is proportional to the case acceleration.

As a general rule the seismic mass displacement (and, therefore, its output) increases as resonance frequency is decreased, but its useful frequency range is reduced. A tradeoff exists between accelerometer sensitivity and frequency bandwidth. Units capable of measuring very high frequency vibrations have a low sensitivity, while units with a high sensitivity generally have a more limited frequency range.

It can be equivalently stated that the force exerted by the spring element is proportional to the case acceleration. Thus, by measuring either the displacement or force, we can obtain a measure of acceleration.

In some older designs a stylus is attached directly to the seismic mass and used to write on a rotating drum or a moving strip of chart paper. The operation of this type of device is easy to understand, but it has no role in a modern vibration and acoustic measurement system.

In the modern accelerometer the spring element is a solid-state material that is electrically responsive to the applied force. Two types of material are commonly used. In the most common type of accelerometer a piezoelectric crystal is used as the spring element. This material generates an electrical charge that is proportional to the stress on the material. In the second commonly used type of accelerometer, a piezoresistive material is used that has an electrical resistance proportional to the applied stress.

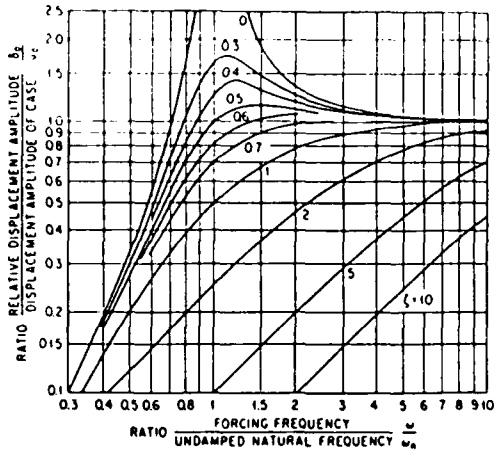
Many other types of accelerometers have been designed using other types of electrically responsive materials and other mechanical configurations. These other types will be briefly described in this chapter. However, greater attention will be placed on the piezoelectric and piezoresistive accelerometers, since these types of transducers play a major role in almost every vibration measurement system.

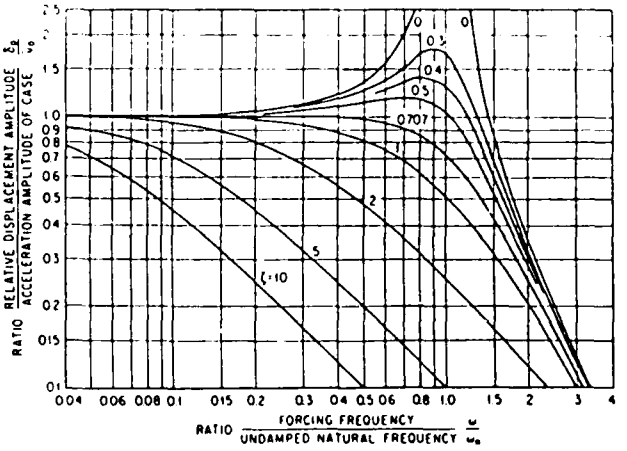
A theoretical calculation of the response of the mass spring oscillator shown in Figure 2.1 results in the amplitude ratio and

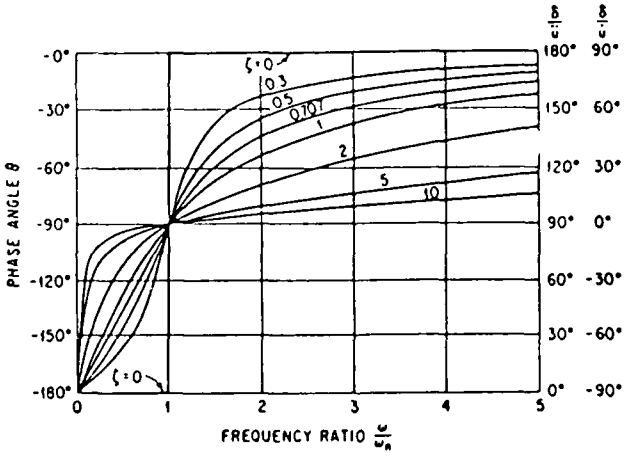
phase curves shown in Figure 2.2. These curves show the theoretical accelerometer response as a function of frequency and the amount of damping. The accuracy of the accelerometer extends to the highest frequency, approximately 60% of the resonance frequency, when the damping is ~ 0.65. The phase response for this value of damping is also fairly linear, which is a desirable characteristic. Although damping is required to provide the highest frequency response for a given design, most commercially available accelerometers have very little damping. In the limiting case of zero damping reasonable accuracy (5%) exists up to 20% of the resonance frequency. The phase accuracy is without error in this limiting case at frequencies up to the resonance frequency. Thus, as a practical rule in selecting an accelerometer the resonance frequency should be five times the upper frequency limit of interest. The amplitude accuracy will be within ~ 5% and the phase accuracy will be excellent.

2.1.4.2 Piezoelectric Accelerometers

The piezoelectric accelerometer is a seismic transducer in which the seismic mass is attached to a piezoelectric crystal. Two basic designs exist. In the compression-type design the vibration of the seismic mass causes the piezoelectric crystal to undergo a compressive stress proportional to the case acceleration. A typical design is shown in Figure 2.3. In this design a preload is applied to the mass so that the crystal is always under a compressive load. This prevents loss of contact between the mass and the crystal and improves the linearity of the accelerometer for high amplitudes. In the shear-type design the seismic mass and crystal are arranged so that an acceleration of the accelerometer case causes a shear stress on the crystal. A primary advantage of the shear-type design is that it can be made to have a very low sensitivity to deformations of the accelerometer case. This type of design can also be successfully used in very small transducers.

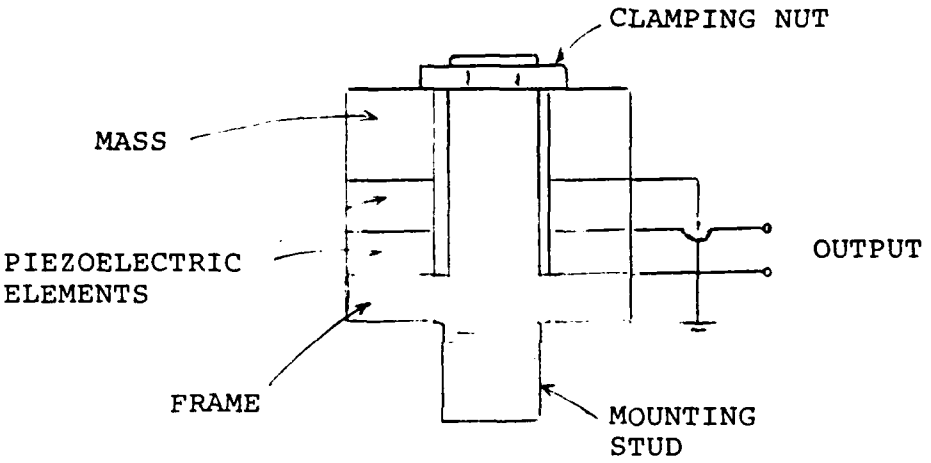






Data from Ref. 2-2

Fig. 2.2 Amplitude and Phase Response for Simple Seismic Transducer



NRL Type C-4
accelerometer

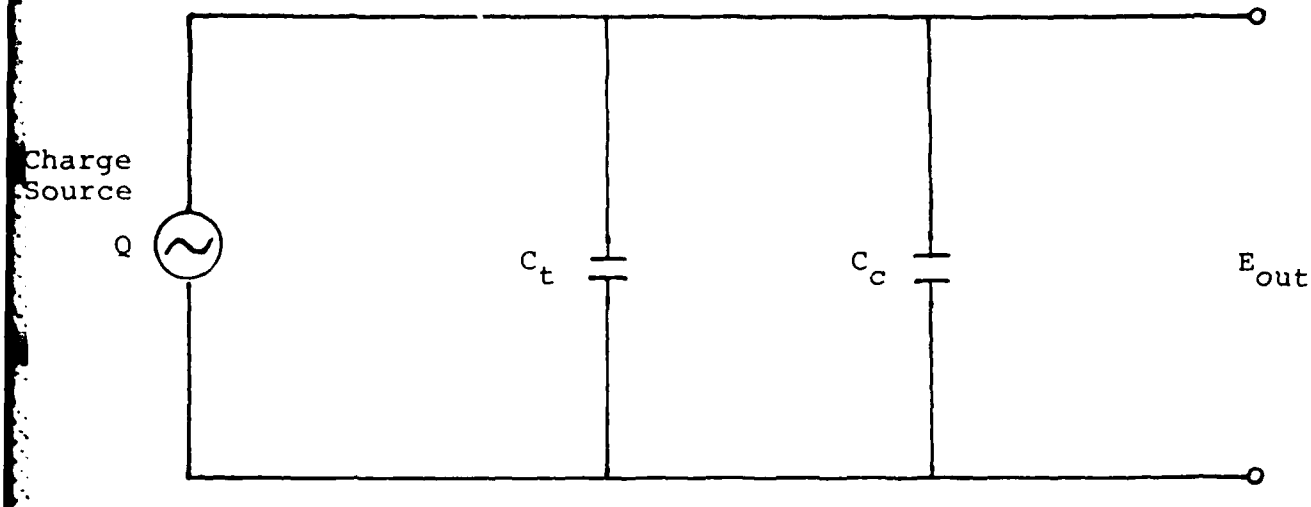
Fig. 2-3 Compression Type Accelerometer

The compression-type design can result in a very rugged transducer with a high sensitivity and high resonant frequency. Its disadvantages are primarily due to its inherent sensitivity to case deformation. Typically, this type of design is more susceptible to base strain, acoustic pressures, and temperature transients than the shear-type.

Several manufacturers provide excellent piezoelectric accelerometers with both types of design so that the user should focus more on the specifications for the transducer being considered than the type of design that is used.

The piezoelectric accelerometer produces an electrical charge that is proportional to the acceleration. Two simplified equivalent circuits are shown in Figure 2.4 and 2.5. In both circuits the capacitance element includes the capacitances of the piezoelectric crystal and the cable from the accelerometer to the signal conditioning amplifier. Note that the equivalent voltage source, Figure 2.5, depends on the cable capacitance. Thus, a change in cable produces a change in transducer sensitivity. This feature is sometimes inconvenient, and it is common to use charge amplifiers rather than voltage amplifiers for signal conditioning. Further discussion of the signal conditioning requirements is presented in Section 2.2.3.

Typically the data sheet for a piezoelectric accelerometer will present both the charge sensitivity, in either picocoulombs per g (pC/g) where g is the acceleration of gravity - 960.6 cm/sec^2 - or picocoulombs per meter per second squared ($pc/m/s^2$), and the voltage sensitivity in millivolts per g (mV/g) or $mV/m/s^2$ at one particular frequency in the range 50 to 200 Hz. The range of sensitivities is from 0.4 pC/g (1.2 mV/g) for a very light-weight (1/2 gm) subminiature accelerometer to 100 pC/g (100 mV/g) for a larger (60 gm) unit. Voltage sensitivities typically range from 1.2 to 100 mV/g. Specially designed accelerometers for measurement of high-amplitude shocks may have a lower sensitivity to prevent the output from being too large. On the other end of

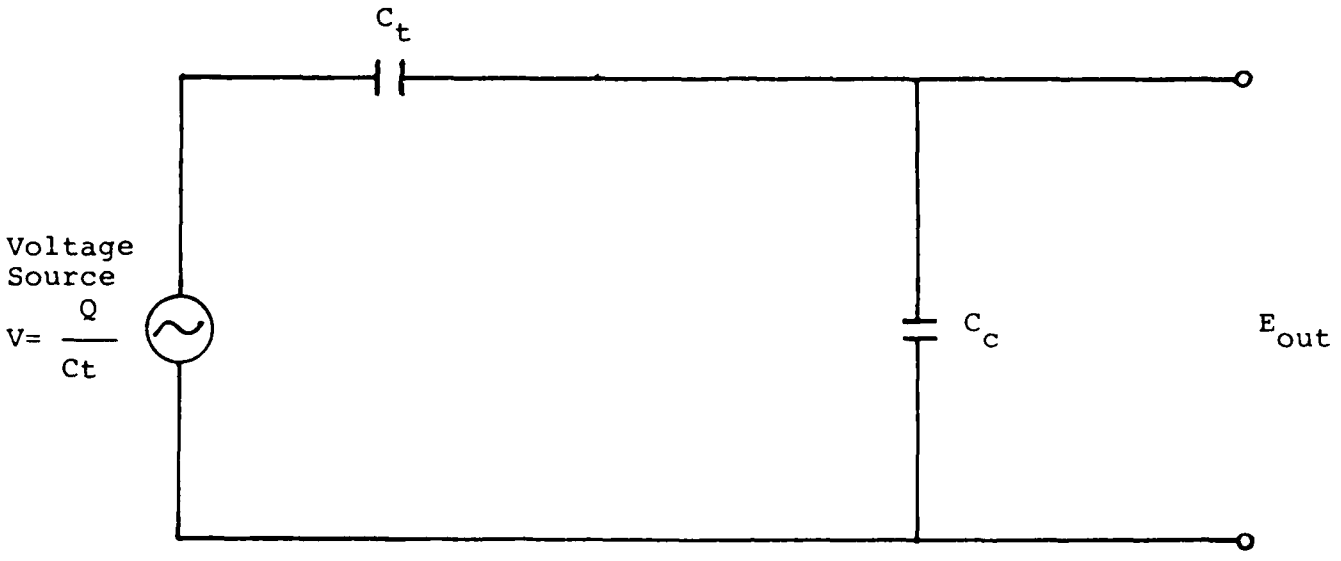


$$\frac{E_{out}}{Q} = \frac{1}{C_t + C_c}$$

C_t Transducer Capacitance

C_c Cable Capacitance

Fig. 2.4 Simplified Charge Source Equivalent Circuit



$$\frac{E_{out}}{V} = \frac{C_t}{C_t + C_c}$$

Fig. 2.5 Simplfied Voltage Source Equivalent Circuit

the scale some piezoelectric accelerometers have built-in amplifiers which can increase the resulting sensitivity to very high values near 10,000 pC/g.

Note from Figure 2.4 that the ratio of the charge sensitivity to the voltage sensitivity is equal to the total capacitance of the accelerometer and its cable. Since small accelerometers must use a small crystal they have a lower value of capacitance. For example, the capacitance of the subminiature accelerometer is only 300 pF (including cable) while that for the 60 gm high sensitivity accelerometer is 1000 pF.

The dynamic range of the piezoelectric accelerometer defines the range of acceleration levels over which the sensitivity is constant within a specified tolerance. The lower limit of the dynamic range is set in practice by the electrical noise of the signal conditioning amplifier or tape recorder. Since the noise can be expressed as an equivalent rms level at the input to the amplifier, either pC in the case of a charge amplifier or mV in the case of a voltage amplifier, the lower limit for the piezoelectric accelerometer is increased by selecting a unit with the highest sensitivity. A typical high-quality charge amplifier noise floor is 10^{-2} rms pC over a bandwidth of 10 to 100,000 Hz. Use of such an amplifier with an accelerometer having a sensitivity of 100 pC/g results in a noise floor of 10^{-4} g rms. Filtering of the amplifier output makes it possible to measure accelerations as low as 10^{-5} g within a limited frequency range.

The upper limit of the dynamic range is set by nonlinearities of the piezoelectric material or by potential damage to the transducer. Typical values are in the range 2,000 to 5,000 g with specially designed high shock transducers having an upper limit as high as 100,000 g.

In many cases the upper limit of the dynamic range is set not by the accelerometer, but by the maximum input to the signal conditioning amplifier. Thus, in selecting a transducer careful

consideration should be given to the amplifier that will be used. Since the amplifier usually costs more than the accelerometer, it is better to buy a new accelerometer to meet the objective of a measurement program than to buy a new amplifier.

A major limitation of the piezoelectric accelerometer is its inability to measure vibration at very low frequencies. Although the resistance of the piezoelectric material is very high it is not infinite so that leakage of the charge occurs at low frequencies. The finite input impedance of the signal conditioning amplifier also limits the low frequency response. For the voltage amplifier the low frequency cutoff, f_c, is determined by the system time constant, RC, where R is the input resistance of the amplifier and C is the total capacitance of the accelerometer and its cable. The cut-off frequency is given by

$$f_c = \frac{1}{2\pi} \frac{1}{RC} \qquad (2\text{-}1)$$

The charge amplifier has an input impedance that is extremely high. Because of its design the low frequency cutoff does not depend on the accelerometer but on the amplifier design.

The upper limit of the frequency response is set either by the resonance frequency of the transducer or by the amplifier. Typically, the resonance of a piezoelectric accelerometer is lightly damped. Thus, the sensitivity of the accelerometer shows a large peak at the resonance frequency that is 15 to 30 dB above the sensitivity at lower frequencies. For most commercially available accelerometers the frequency response is specified with a ±5% deviation of the sensitivity. Typically, its upper limit is 20% of the units mounted resonance frequency, which is usually specified in the data sheet. The mounting technique will often reduce the resonance frequency, and, therefore, the upper limit of the frequency response.

The lower limit of the accelerometer is typically specified to be 1 or 2 Hz. However, the input impedance of the signal conditioning amplifier must be sufficiently high to obtain this lower limit. In addition, the leakage resistance of a piezoelectric accelerometer decreases with increasing temperature. This may affect the low frequency sensitivity, and special consideration must be given when low frequency measurements are required in a high temperature environment.

2.1.4.3 Piezoresistive Accelerometers

The piezoresistive accelerometer is also a seismic transducer. However, high-sensitivity piezoresistive materials are used as the sensing elements. These materials exhibit a change in resistance due to an applied strain and are used in semiconductor strain gages (see Section 2.1.6). In fact, the sensing element in the piezoresistive accelerometer can be considered to be a strain gage.

The primary advantage of the piezoresistive accelerometer is that it provides both static and dynamic response. Although the upper frequency limit for the piezoresistive unit is somewhat lower than for the piezoelectric, the frequency range over which valid data can be collected is very large. (up to 30kHz in commercial units).

In some piezoresistive accelerometer designs the seismic mass is damped in order to extend the frequency response to a higher limit. The phase response is affected, but by proper choice of the damping coefficient (see Section 2.1.4.1), the phase can be made to be approximately a linear function of frequency. A linear phase results in a simple time delay of the signal without distortion. The damping also reduces the transducer response at its resonance frequency. Although the transducer does not provide a valid measurement at its resonance frequency, a high signal output at this frequency can overload

the signal conditioning amplifiers and filters. This overload problem is particularly severe for shock tests where ringing of an undamped transducer can result in a large signal output.

The static or DC response capability of the piezoresistive accelerometer is also desirable in shock tests where the objective is to measure a transient waveform without distortion.

A disadvantage of the piezoresistive design is the requirements for a regulated constant current power source in some commercial units and a regulated constant voltage power source in other units. This increases the complexity of the signal conditioning equipment. In addition, the choice of piezoresistive types is more limited than for the piezoelectric accelerometers. Subminiature units are generally unavailable, and the maximum temperature at which piezoresistive accelerometers can operate is typically 120°C.

2.1.4.4 Selection Criteria

The selection between a piezoelectric and a piezoresistive accelerometer should be based on the requirement for low frequency response. If data below 20 Hz are not required, a piezoelectric unit with a voltage amplifier is the best choice. For a higher cost, the voltage amplifier can be replaced with a charge amplifier in order to reduce the lower frequency limit. With some signal conditioning charge amps it is possible to extend the frequency response range to less than 1 Hz.

An alternative to the piezoelectric accelerometer and charge amp is the piezoresistive accelerometer with a constant current power supply. Although the transducer itself provides DC measurement capability, the DC voltage from the power supply must be blocked with an AC amplifier.

To obtain true DC measurement capability the piezoresistive accelerometer must be used with a Wheatstone bridge circuit. The bridge offers excellent accuracy for both DC and dynamic

measurements but requires the signal conditioning equipment to have bridge balancing circuits.

In applications where either a piezoelectric or piezoresistive unit can be used, the piezoelectric accelerometer should be selected because of its greater versatility.

The next step in selecting an accelerometer is to consider the sensitivity, upper frequency limit, and weight. These three factors are typically interrelated and must be considered together. The sensitivity must be sufficiently high that the noise floor of the signal conditioning electronics is below the lowest vibration amplitude of interest. For example, if a measurement is required of vibration amplitudes down to 10^{-3} g's and the signal conditioning amplifier noise at full gain is 10^{-2} mV rms referred to the input, an accelerometer sensitivity of at least $10^{-2}/10^{-3}$ = 10 mV/g is required. A similar example can be given for a charge amp in which case the amplifier noise is expressed in pC and the accelerometer sensitivity in pC/g.

In many measurement programs consideration must also be given to the maximum value of the acceleration to be measured. For example, if the maximum acceleration is 100 g and the maximum voltage to the amplifier is 10 V, then the accelerometer sensitivity must be less than 10,000/100 = 100 mV/g.

After determining the range of sensitivities needed for the measurement, consideration must be given to the frequency response range and the accelerometer weight. The upper frequency limit typically extends to 20% of the transducers mounted resonance frequency. When manufacturers data are given for either the mounted resonance frequency or an upper frequency limit, it is assumed that the transducer is stud-mounted to a flat steel surface.

Other mounting conditions may lower this resonance frequency and thereby reduce the upper frequency limit. The mounting problem can be somewhat alleviated by using a very lightweight

accelerometer. However, the sensitivity and durability of such units may not be sufficient.

The weight of the accelerometer can also adversely affect the measurement on very lightweight structures by the mass-loading effect. In general, the point impedance of the structure must exceed the mechanical impedance of the accelerometer for the measurement to be valid.

When possible the lightest weight accelerometer having a sensitivity that is 3 to 10 times greater than the minimum requirement should be used. This criteria provides a minimum signal-to-noise ratio of 10 dB and reduces the problems associated with mounting a heavier transducer.

Additional factors to be considered are a) transverse sensitivity, b) temperature range, c) acoustic sensitivity, d) sensitivity to base strain, e) magnetic sensitivity, f) durability, and g) type of connector. In most applications these additional considerations will not influence a decision on which accelerometer to use.

2.1.4.5 Calibration

Two methods of accelerometer calibration exist - absolute and relative. In an absolute calibration the accelerometer is excited to a vibration level that can be measured absolutely within a certain error bound. The absolute measurement is typically carried out using either optical interferometer or reciprocity techniques. Other absolute measurement techniques exist, but are useful only for static or very low frequency measurements. For example, an absolute calibration of an accelerometer with DC frequency response is achieved by simply inverting the accelerometer in the earth's gravitational field.

A relative calibration is carried out by comparing the output of an accelerometer with the output of a standard unit that has been absolutely calibrated. Several manufacturers sell

standard accelerometers for such a purpose. These units have been calibrated absolutely by the manufacturer and are designed so that the accelerometer to be calibrated can be directly mounted on the standard accelerometer which in turn is mounted on a small mechanical shaker. No special requirements need be placed on the shaker since the calibration is based on the relative output of the two accelerometers.

Although absolute calibration apparatus can be set up, the comparison method is generally preferred for routine calibrations because of its simplicity. A standard accelerometer should be kept in a well-controlled laboratory setting and used periodically to check the calibration of all accelerometers. The standard accelerometer should periodically be sent back to the manufacturer for an absolute calibration.

Because of the potential damage to accelerometers during their use, each unit should be calibrated at the end of a measurement program. The standard accelerometer should be calibrated at least once a year.

Before calibration the accelerometers should be carefully inspected for mechanical damage and discarded if any has occurred.

The actual calibration can be carried out using either a swept sine or random excitation. The choice of technique depends in large part on the available data processing equipment. Availability of a two-channel analyzer makes the random technique attractive. In either case the calibration should be carried out over the full frequency response range of the transducer.

It is difficult to obtain an accurate calibration in the field. However, an end-to-end calibration of the entire instrumentation system is very useful in checking that all gains are properly set and that equipment is operating properly. Small battery-operated shaker tables are available for this purpose. Prior to mounting the accelerometer on the test structure, it can be mounted on the shaker table and subjected to the calibration

excitation. The output signal from the instrumentation chain can be recorded or monitored to determine that its level is consistent with the accelerometer sensitivity and all amplifier gains.

In cases where the accelerometers must be mounted on the structure before the measurement equipment is assembled, it is not feasible or worthwhile to remove the accelerometers for a field calibration. A careful, systematic laboratory calibration procedure using a standard reference accelerometer is sufficient. In cases where accuracy is of critical importance this calibration should be carried out both before and after the measurements.

2.1.4.6 Mounting Procedures

The method by which an accelerometer is mounted to a test structure is critically important to its accuracy at high frequencies. Typically, the upper limit to the frequency response is set at 20% of the mounted resonance frequency. The highest resonance frequency is obtained by using a steel stud that screws into threaded holes in the base of the accelerometer and in the test structure. To prevent a lowering of the mounted resonance frequency the surface of the test object should be smooth and flat. The hole in the test object must be at right angles to the surface.

In most cases the calibration data provided by the manufacturer is obtained using this mounting technique.

A second method for mounting an accelerometer is to bond a small mounting pad, which can be obtained from the accelerometer manufacturer, to the test object using a rigid cement. Most manufacturers recommend one or more types of cement. Proper use of these cements does not degrade the accelerometer performance. However, the use of a soft bond or a bond that is too thick will result in a reduced mounted resonance frequency.

Other methods of mounting accelerometers include wax, two-sided adhesive tape, magnets, and long probes. Use of these techniques is discouraged except for surveys where quantitative measurement accuracy is not needed.

2.1.4.7 Specialized Units

Many piezoelectric accelerometers are designed for use under special conditions. Shock accelerometers are designed to have a low sensitivity and a very high resonance frequency. In addition, these specialized transducers are designed to minimize zero shift. This phenomenon occurs when piezoelectric materials are subjected to high stress. It results in a small voltage output after the shock is over and can cause a large distortion of the frequency spectrum with low frequency levels being too high.

Other specialized accelerometers are designed with signal conditioning FET amplifiers built in to the transducer case. These units provide a low impedance output, and, therefore, do not require use of charge amplifiers to obtain a good low frequency response. In addition, the built-in amplifier can have some gain so that the effective transducer sensitivity can be made to be very high.

A disadvantage of this type of unit is its requirement for a power supply and typically a fairly low upper limit to the maximum acceleration it can measure.

A third type of specialized accelerometer is designed for use at high temperatures. In these units a quartz sensing element is used to minimize the influence of temperature on the performance. These units can be used at temperatures as high as 260^{o}F and as low as -200^{o}F. However, their sensitivity is generally quite small.

2.1.5 Microphones and Pressure Transducers

Pressure measurements are required in a variety of applications. In many cases the transducers to be used must provide an accurate measure of a static or slowly varying pressure. Although the amplitude of the pressure to be measured can range over many orders of magnitude, the consideration of frequency response is often not of concern. This is not the cases in a vibration and acoustic measurement system. In this case the frequency range over which the pressure transducer is to operate is one of its most important attributes.

In this section we will limit our attention to pressure transducers that are designed to measure fluctuating pressures over a range of frequencies. This type of transducer is commonly referred to as a dynamic pressure transducer or, when it is used to measure acoustic pressures in air, as a microphone.

In many transducer catalogs microphones and dynamic pressure transducers are listed together. Although these two types of transducers are similar, there are some important distinctions. In general terms a microphone is a dynamic pressure transducer used to measure the fluctuating or "acoustic" pressure in air. Unlike a general purpose dynamic pressure transducer, which may be able to measure pressures up to several hundred or several thousand psi, the microphone is limited to measuring acoustic pressures, which are well below the standard atmospheric pressure of 14.7 psi. The general purpose dynamic pressure transducer may also have a frequency response extending down to very low frequencies or to DC. The microphone, on the other hand, is designed to measure only fluctuating pressures, and its frequency response typically extends down to 5 Hz. Thus, there are many applications for which a microphone is unsuitable because of its limited ability to measure pressure signals with a large amplitude and low frequency content. Examples include measurements of shock waves, ballistic pressures, hydraulic

pressures, and combustion pressures. On the other hand, microphones typically have a very high sensitivity and can be used to measure the very low fluctuating pressures that are typical of sound fields. It is important to remember that an acoustic level of 65 dB, which is an average level for speech, corresponds to an rms pressure approximately 5×10^{-6} psi. A high-sensitvity pressure transducer with a voltage sensitivity of 100 mV/psi would provide an output of less than 1μV.

A variety of physical phenomena have been used to design dynamic pressure transducers and microphones. These include piezoelectric effects, variations in capacitance, electromagnetic induction, electrodynamic induction, magnetostriction, and variations in resistance. Although the moving-coil electrodynamic or "dynamic" microphone has been in popular use for many years, its low sensitivity and susceptibility to vibration-induced response eliminate this type of transducer from modern measurement systems.

As a practical matter, we can limit the discussion to three types of microphones or dynamic pressure transducers. These are:

1) condenser,
including both air-condenser and electret types;

2) piezoelectric
including ceramic and crystal types; and

3) piezoresistive,
including strain-gage types.

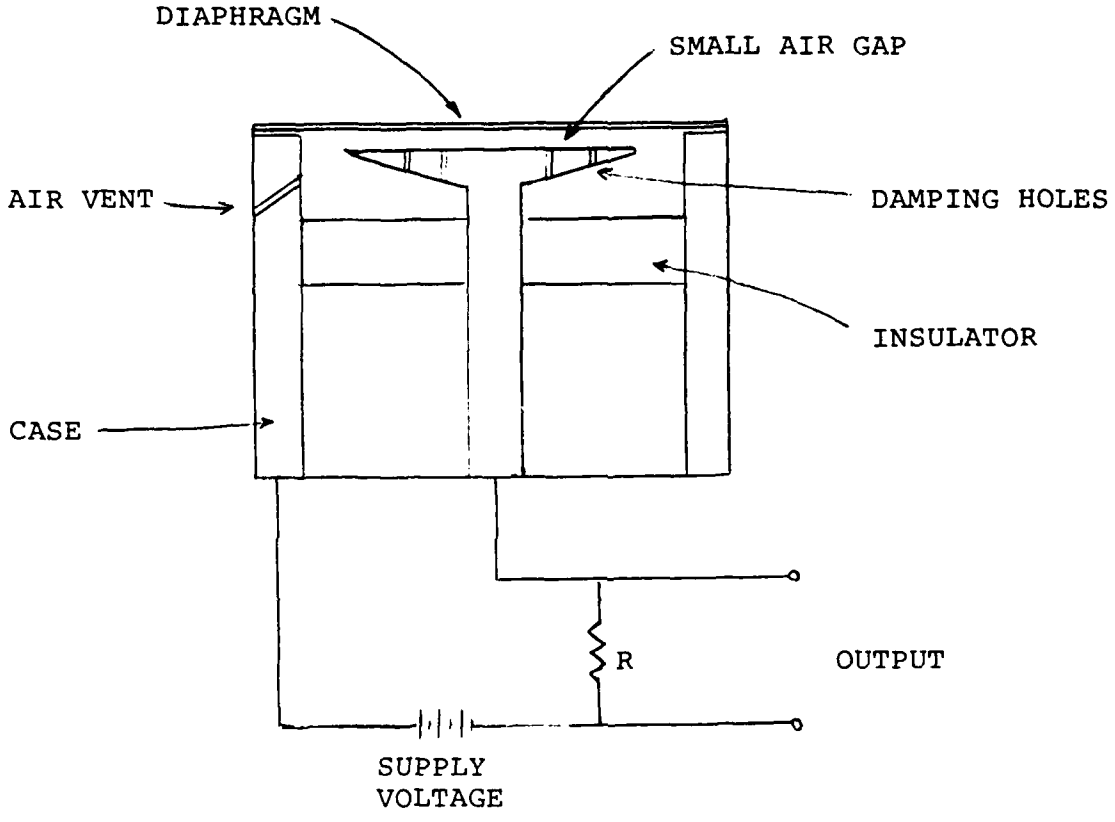
Of these three, the air-condenser microphone offers the highest accuracy, but at the expense of increased fragility and more stringent signal conditioning requirements. The electret microphone offers high accuracy, no need for a power supply, less

weight than an air-condenser microphone, but has a reduced dynamic range. The piezoelectric microphone or dynamic pressure transducer is significantly less affected by environmental factors than the condenser microphone and can be made to withstand very severe conditions. Its accuracy is sufficient for many measurement applications. The piezoresistive pressure transducer offers many of the advantages of the piezoelectric transducer with the added capability of extending the frequency response to DC. Its upper frequency limit is higher than condenser microphones although not as high as the piezoelectric units, but in cases where a frequency response to DC is required this transducer is a good choice. Its sensitivity is much lower than condenser microphones.

2.1.5.1 Condenser Microphone

The condenser microphone is shown in Figure 2.6. A thin diaphragm is located a small distance from a backing plate to form a variable capacitance. Pressure on the diaphragm causes it to move relative to the backing plate which results in a change in capacitance. In most cases a small vent hole is used to equalize the static pressure on each side of the diaphragm. This vent prevents changes in barometric pressure, which are often larger than the acoustic pressures which are to be measured, from causing the diaphragm to deflect beyond its linear range. This vent hole also eliminates the ability of the microphone to measure pressures at very low frequencies and sets a lower limit on the frequency response, typically 5 Hz.

At high frequencies a mechanical resonance of the diaphragm occurs which sets an upper limit to the frequency response. In many designs the response at this resonance is damped by having small holes in the backing plate. The flow through these holes combined with the acoustic radiation pressures damp the diaphragm



Example from Ref. 2-3

Fig. 2-6 Air Condenser Microphone

resonant response and extend the useful frequency range to higher frequencies.

In air-condenser or condenser microphones, a large DC voltage (~ 200 V) is applied to the two plates of the variable capacitor. A simplified equivalent circuit for this type of microphone is shown in Figure 2.7. An exact analysis of the circuit shown in Figure 2.7 is beyond the scope of this section due to the nonlinearities involved. An approximate analysis is presented in Section 2.2.5. There is one point of immediate concern in evaluating the condenser microphone. Due to the very small capacitance of this type of transducer, ~ 20 pF for a 1/2" microphone, the signal conditioning amplifier must be located in the immediate proximity of the microphone. The capacitance of even a short length of cable can cause a reduction in signal amplitude and higher distortion. In most commercially available units the cable capacitance is avoided altogether by screwing the microphone cartridge directly onto the signal conditioning amplifier. This greatly increases both the size and weight of the transducer unit, and often eliminates the condenser microphone from consideration.

The sensitivity, dynamic range, and frequency response of the condenser microphone are strongly dependent on the microphone size. A large microphone allows the diaphragm to be more flexible and to have a larger capacitance. This results in a higher sensitivity and allows measurement of extremely low sound pressure levels. On the other hand the greater flexibility of the diaphragm results in large deflections which increases the distortion at high amplitudes. The greater diaphragm flexibility also reduces the diaphragm resonance and thereby limits the upper limit of the frequency response.

The polarization voltage also influences the microphone performance. A low voltage results in distortion at high amplitudes. Thus, most condenser microphones require a

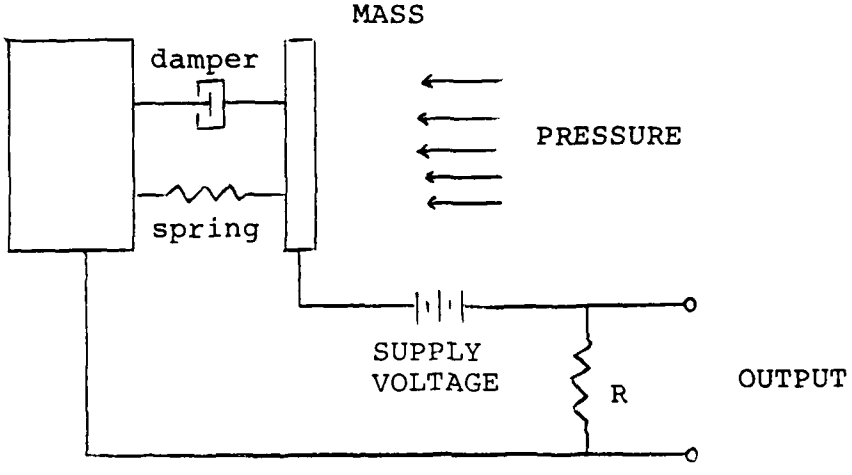


Fig. 2-7 Representation on Condenser Microphone

polarization voltage of 200 V. Although units are available that require only 28 V polarization, the upper limit of the dynamic range for these units is typically 10 to 20 dB (a factor of 3 to 10 in amplitude) lower than for units with higher polarization voltage.

Condenser microphones are available with either a free-field response, or a pressure response. Precision measurements of a localized sound source are made using a free-field microphone pointed directly towards the source (0^{o} incidence). Pressure measurements in a diffuse field are made using a smaller pressure microphone (either 1/8", 1/4" or 1/2" diameter). Use of a 1" microphone in diffuse field measurements distorts the sound field in a manner discussed below. Finally, pressure measurements in a small coupler or enclosed volume are made using a pressure microphone with an optional probe tube attachment.

At high frequencies microphone size also can cause a distortion of the measured sound field. Thus, measurements obtained with the microphone diaphragm at right angles to a high frequency sound source may be significantly different than measurements obtained with the microphone turned 90^{o}. Typically, a grid is used over the diaphragm. This grid both protects the microphone from damage and alters the directivity of the microphone so that its response is more uniform for different angles of sound incidence.

The use of specially designed nose cones mounted on free-field microphones can reduce the directional variation of the frequency response. For example, the directional variation of frequency response of a 1" condenser microphone with nose cone is less than 3 dB at 10 kHz for both randomly incident sound and sound at an arbitrary angle of incidence.

2.1.5.2 Electret Microphone

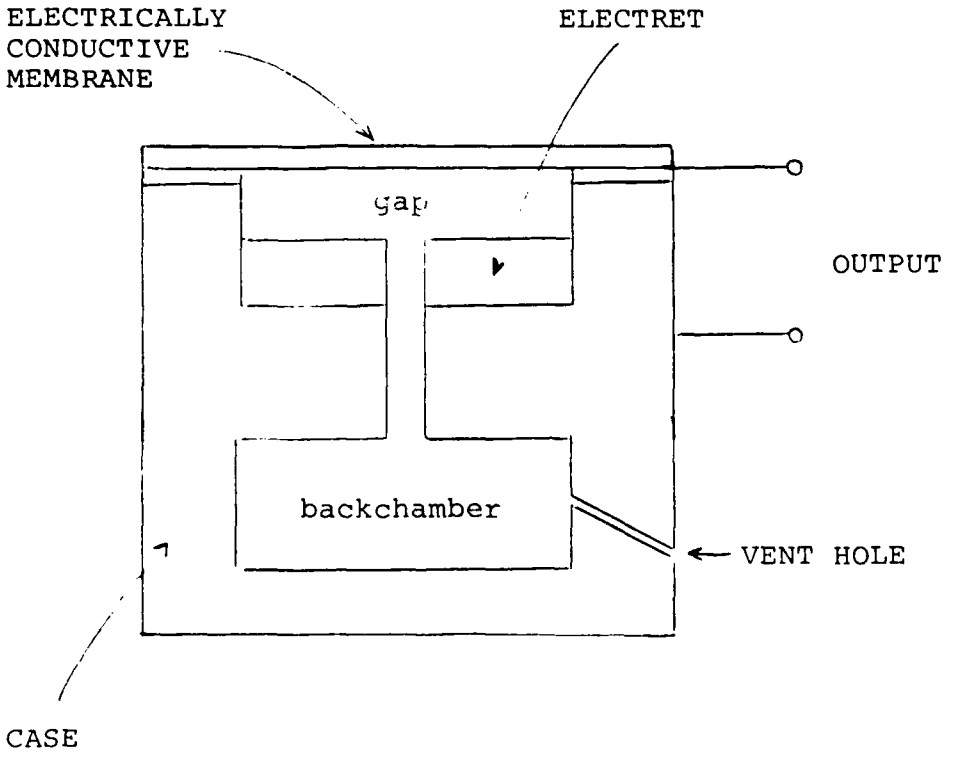
The electret microphone is a type of condenser microphone in which the required charge is permanently imbedded in a layer of electret material between the diaphragm and the backplate, as shown in Figure 2.8. The need for a polarizing voltage is eliminated, which in turn simplifies the requirements for the signal conditioning amplifier. Other advantages of the electret microphone over the air-condenser microphone include; a lower cost, lower susceptibility to environmental factors, and greater mechanical durability. Disadvantages are a somewhat lower sensitivity, smaller dynamic range, and lower stability over long periods of time.

2.1.5.3 Piezoelectric Pressure Transducer

The piezoelectric pressure transducer uses a piezoelectric crystal, typically of quartz or polycrystalline ceramic materials. The principle of operation is that the pressure on the end of cross sectional area of the transducer applies a compressive force on the piezoelectric crystal, and the crystal element produces an excess of electric charge in phase with the force.

Piezoelectric pressure transducers and compressive-design piezoelectric accelerometers are very similar except for the addition of the seismic mass to the accelerometer (Section 2.1.4.2).

Advantages of piezoelectric pressure transducers are durability, response to very high frequencies, capability to measure high pressure levels, and similarity to piezoelectric accelerometers so that the same signal conditioning equipment can be used. Units are available that are insensitive to humidity and corrosive elements in the environment and that can be used over a wide temperature range (-270 to 260^{o}C).



Example from Ref. 2-3

Fig. 2-8 Electret Condenser Microphone

The main disadvantages of piezoelectric transducers are a high sensitivity to vibration relative to other transducer types, and the transducers cannot be used for static pressure measurements.

The sensitivity of a piezoelectric pressure transducer is specified in terms of a charge sensitivity (pC/psi), and sensitivities range from 0.35 pC/psi for general purpose transducers up to 8.0 pC/psi for high sensitivity transducers. The sensitivity may also be specified as a voltage sensitivity (mV/psi), the voltage developed across the piezoelectric crystal per unit pressure.

The dynamic range is the range of pressures over which the sensitivity is constant within a specified tolerance (1% or less). The dynamic range tends to be on the order of four decades of pressure, and various transducers are available for pressures in the range of 0.01 to 80,000 psi.

Frequency response is limited at high frequencies by the resonant frequency of the crystal, and this resonant frequency is in the range of 100-500 kHz. The low frequency response is controlled by the signal conditioning instrumentation (Section 2.2.3) and is typically 2-5 Hz.

2.1.5.4 Piezoresistive Pressure Transducer

Piezoresistive (PR) pressure transducers are similar in operation and characteristics to PR accelerometers (Section 2.1.4.3). The difference between the transducers is that the accelerometer has a seismic mass to convert a force to an acceleration signal, whereas the force per unit area of the transducer diaphragm is the variable sensed by a PR pressure transducer.

The principal of operation is based upon the piezoresistive effect, wherein the electrical resistance of silicon and other semiconductor strain gage elements is proportional to the applied

mechanical stress. The variable resistance of the PR strain gage is used in a Wheatstone Bridge circuit within the transducer. An external power supply is needed to power the bridge, and commercial PR pressure transducers are available with either a regulated DC voltage power supply or a regulated DC current power supply.

The clearest advantage of PR pressure transducers over other types is the ability to measure static pressure, and additionally measure pressures with flat frequency response at frequencies up to 200 kHz. Additional advantages are a very low output impedance, and negligible reponse to thermal transients. Disadvantage of PR pressure transducers are their somewhat limited temperature (-50^{o} to $120^{o}C$), and the need for a regulated power supply.

The voltage sensitivity of a PR pressure transducer is specified in units of mV/psi, and transducers are commercially available with sensitivities in the range from 0.15 to 150 mV/psi. The gage pressure full scale range in which they operate can be from 2000 to 2 psig. Units are available for both gage and differential static and oscillatory pressure measurements.

Piezoresistive transducers can be very small and light. For example, commercial units are available with a 2.4 mm face diameter, 19 mm body length, and a mass of three grams.

2.1.6 Strain Gages

Strain gages are used in vibration and acoustic measurement systems for two purposes: first, they are used to obtain a direct measurement of the dynamic strain at a point on a test object; second, they are used as an element within a transducer. This section is oriented toward the use of commercially-available transducers rather than their design. Thus we will limit the discussion to the use of strain gages for a direct measurement of

strain. It should be pointed out, however, that the piezoresistive accelerometer and pressure transducers are an implementation of a strain gage transducer using solid state technology (Section 2.1.4.3. and 2.1.5.).

2.1.6.1 General Types and Uses

The strain gage under consideration is a thin foil or wire that exhibits a change in resistance due to the mechanical strain exerted on the gage. The sensitivity of the gage is characterized by a gage factor $\underline{k}$, where

$$\underline{k} = \frac{\Delta R/R}{\Delta L/L} \qquad (2\text{-}2)$$

$\underline{k}$ is the gage factor, R is the initial gage resistance, L is the initial gage length, ΔR is the change in resistance, and ΔL is the change in length. The strain exerted on the gage is $\Delta L/L$ so that the fractional change in gage resistance, $\Delta R/R$, is proportional to strain. The gage factor is determined by the material of the gage wire or foil with typical values for metallic foil gages in the range 2 to 4.

To provide a reasonable sensitivity the gage resistance must be high. Thus, the wire or foil used for the gage is very thin and difficult to handle. In most commercially-available gages the wire or foil is bonded to some type of backing material. The wire or foil is also folded several times into a grid, as shown in Figure 2.9, so that the resistance is increased. This type of gage is typically referred to as a bonded strain gage.

In a semi-conductor gage, a piezoresistive material is used in place of the foil or wire. This can result in a large increase in sensitivity (effective gage factors that are ~ 50 times the gage factor of a conventional foil gage).

By proper choice of signal conditioning equipment (see Section 2.2.6) a strain gage can be used to measure both static

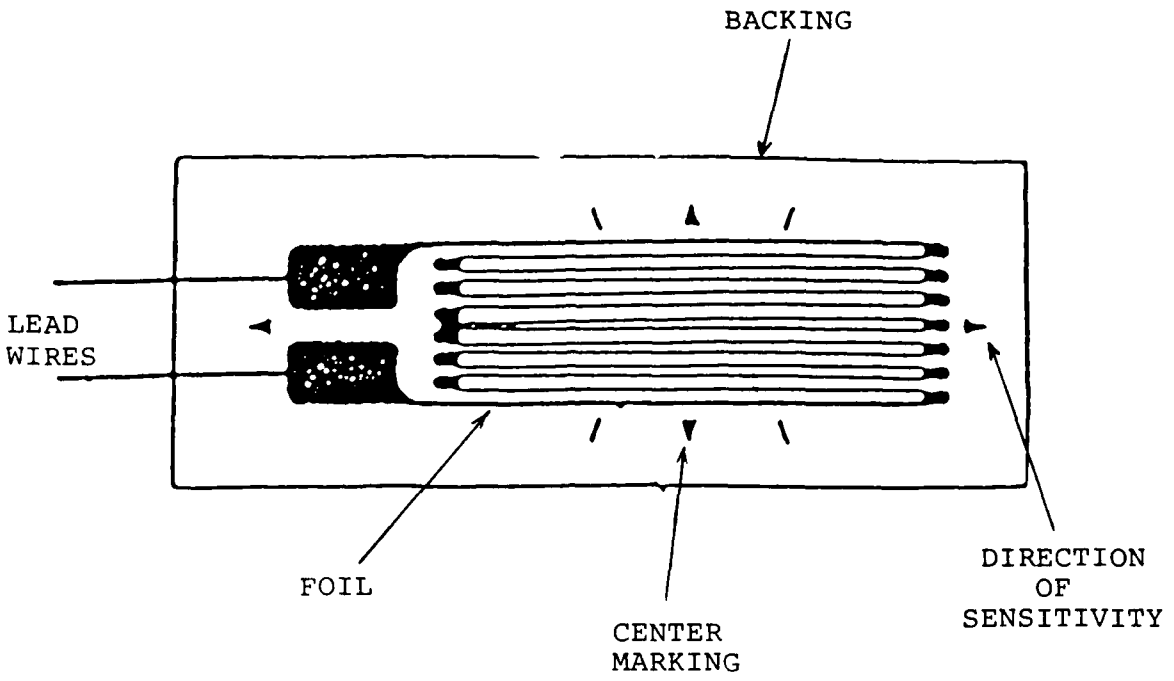


Fig. 2.9 Foil Strain Gage

and dynamic strains. Because of their small size and light weight they can be used to collect valid dynamic data up to very high frequencies (~ 100,000 Hz). However, because the strain amplitude is typically very small at these high frequencies, it is necessary to select gages with a high sensitivity in order to obtain valid measurements at these frequencies. Because of this requirement, bonded strain gages are generally classified as being for either static or dynamic measurements. The static gage has a relatively low gage factor, but is less affected by temperature changes than a gage intended for dynamic measurements.

The effect of temperature is to change both the gage factor and the gage resistance. However, if the time period of the temperature change is large compared to the duration of the dynamic measurement, the change in resistance will not significantly affect the accuracy of the dynamic measurement. At low frequencies and for static measurements the effects of temperature fluctuations are often compensated for by using pairs of gages with one active and one dummy gage. The active gage is bonded to the test specimen while the dummy gage, which is identical to the active gage, is bonded to a piece of material that is unloaded and at the same temperature as the test specimen. The active and dummy gages are placed in a Wheatstone bridge as shown in Figure 2.10. Changes in temperature cause the resistance of both gages to change equally so that the bridge stays in balance and generates no output. In addition, if the material on which the dummy gage is mounted has the same thermal expansion as the test specimen, then the strain due to thermal expansion is also compensated for.

The change in gage factor due to temperature affects both static and dynamic measurements. Fortunately, this change is small - on the order of 1 to 2% over a temperature range of $100^{o}F$ - for metallic gages. Most strain gages are designed to be sensitive only to strains in one direction - along their

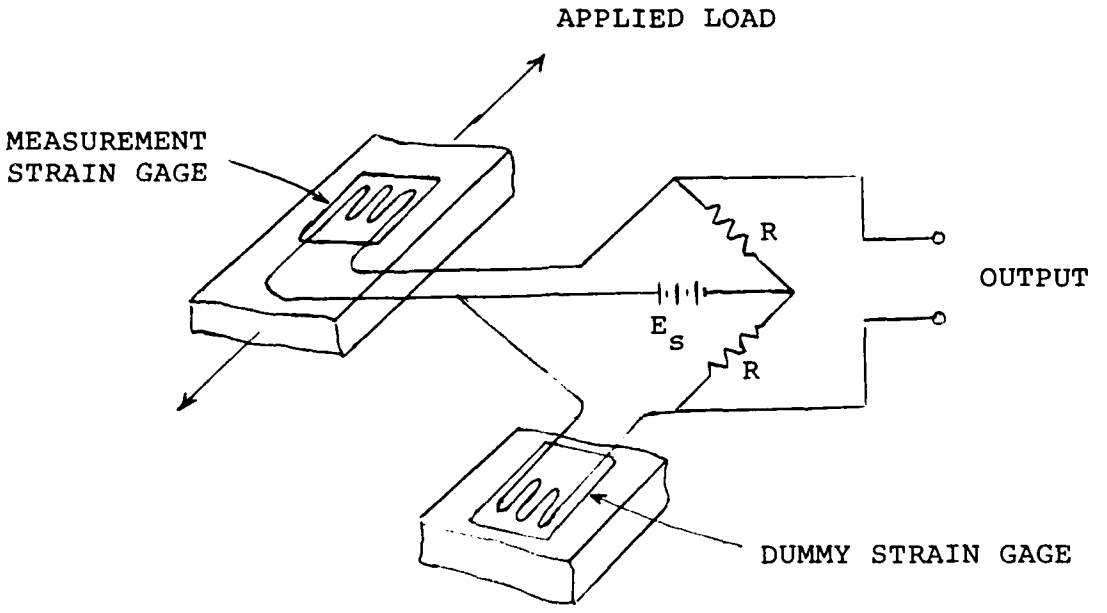


Fig. 2-10 Temperature Compensation for Strain Gages

longitudinal axis. Gages such as the one shown in Figure 2.9 have a small transverse sensitivity due to the corners at each end of the grid. By making the corners thicker, as shown in the Figure, the transverse sensitivity can be reduced to less than 1% of the main axis sensitivity.

2.1.6.2 Selection Procedure

The first step in selecting a strain gage is to determine if its use is necessary. The inherent problems in using bonded strain gages include: 1) difficulty in bonding the gage to the test specimen, 2) difficulty in attaching the lead wires, 3) general fragility of the gage, 4) pickup of electromagnetic noise, and 5) a generally low sensitivity. Because of these problems the strain measurement should be replaced by an acceleration measurement when possible. However, in cases where a strain measurement is needed the metallic foil or semi-conductor gages should be used.

The choice between a metallic foil and a semiconductor gage is based in large part on the differing sensitivities of these two types of gages. Although the semiconductor gage is more expensive, more difficult to apply, and generally more sensitive to temperature changes, it provides a gage factor and sensitivity that is 50 to 100 times greater than that of the metallic foil gage.

To obtain the gage sensitivity in $\mu V/\mu strain$ we assume a current-source power supply that provides the maximum allowable current to the gage. This maximum current is limited by heating of the gage and is specified by the manufacturer. A typical value is 25 ma. Using this current-supply the voltage across the gage is given by

$$E_{out}(t) = \underline{k} \quad I \quad R \quad \varepsilon(t) \qquad (2\text{-}3)$$

where E_{out} is the output voltage, $\underline{k}$ is the gage factor, I is the maximum current for the gage, R is the gage resistance, and $\varepsilon(t)$ is the strain to be measured.

For a typical bonded metallic foil gage

$$\underline{k} = 2.0,$$

$$I = 25 \text{ ma, and} \tag{2-4}$$

$$R = 120 \text{ ohm}$$

The sensitivity is

$$\frac{E_{out}}{\varepsilon} = 2.0 \quad 25 \quad 10^{-3} \quad 120 \quad \text{volts/unit strain} \tag{2-5}$$

$$= 6.0 \ \mu V/\mu\text{strain}$$

If a high quality, low-noise strain gage amplifier is used, the noise floor over a 25 kHz bandwidth is approximately 3 μV. The equivalent level is 0.5 μstrain. Thus, the metallic foil gage can generally be used to measure strain levels above 1.5 μstrain*. Even lower levels of strain can be measured by filtering the output into a narrower frequency band.

The noise level of 0.5 μstrain is sufficiently low for measurements of strain that are to be used to determine if dynamic stresses are within ultimate strength or fatigue limits. However, it must be understood that the vibration level corresponding to a strain level of 0.5 μstrain is quite high. It is difficult, if not impossible, to determine precisely a vibration level from a strain level. However, as an approximate rule, the average acceleration level of a structure can be related to the average strain level by the equation

*the signal to noise ratio at this level of strain would be 10 dB

$$a = \omega c_{\ell}\ \varepsilon \qquad (2\text{-}7)$$

where a is the acceleration, ω is the radian frequency, c_{ℓ} is the speed of longitudinal waves in the material of the structure, and ε is the strain level. A dynamic strain amplitude of 0.5 μstrain at 1000 Hz on a steel or aluminum test object (c_{ℓ} = 17,500 ft/sec) is found to correspond to an acceleration level of

$$a = 1.7 \text{ g's} \qquad (2\text{-}8)$$

This acceleration amplitude is well above the noise floor of a typical accelerometer. Thus, the example illustrates that strain gages (other than piezoresistive transducer design) should not be used if a measurement of vibration is desired.

A second factor in selecting a strain gage is its size. A bonded gage measures the average strain over its active length. The advantage of a longer length is a higher sensitivity. However, the ability to measure strain in an area of a steep strain gradient due to a stress concentration is impaired.

The third factor to be considered is environmental conditions. Temperature, humidity, pressure, electromagnetic fields, corrosive liquids, and other physical factors can damage the gage and affect its performance. Fortunately, in a "normal" environment these factors do not have a major effect on strain gage performance. Thus, a bonded metallic foil strain gage can provide accurate dynamic strain data in most measurement programs. The sensitivity of the metallic foil gage is generally sufficient for strain measurements on structures, and this type of gage should be used if possible. However, where a higher sensitivity is needed a semiconductor gage can be used.

2.1.6.3 Calibration

A direct calibration of bonded strain gage sensitivity is impossible, since once the gage is bonded to a test structure it cannot be removed. Thus, the gage can only be used once and cannot be calibrated first and then used for a measurement. Most strain gage manufacturers use statistical methods of quality control. Based on calibration data from a large number of gages off the production line they are able to specify an error-bound on the gage factor.
The specification of this error-bound by a reputable manufacturer is the only practical means to determine gage measurement accuracy.

Although a direct calibration of a gage sensitivity is not possible, it is good practice to measure the gage resistance before and after measurement tests. The resistance should not change, unless temperature changes have occurred, and should be within manufacturers' specifications.

2.1.6.4 Bonding and Surface Preparation

The bonding of a metallic or semiconductor gage to the test structure is critical to achieving satisfactory performance. On the one hand the bond must have sufficient rigidity to transmit the strain from the test object to the gage. On the other hand the bond must not be so rigid that it increases the local stiffness of the test object and thereby reduces the strain at that point. The best approach is to use a very thin layer of a cement recommended by the gage manufacturer and to follow carefully the instructions for cleaning and preparing the surface. The bonding of gages to rough or curved surfaces is particularly difficult and should be avoided if possible.

Failure to obtain a good bond over even a small area of the gage results in poor data accuracy. First time users of strain gages should experiment with their bonding procedure. One

approach to this experimentation is to bond gages to the top and bottom surfaces of a small cantilevered beam as shown in Figure 2.11. Under bending deformation the output of the two gages should be equal and of opposite signs. Thus, if the two gages are connected in series, the composite change in resistance should be zero. Any measured change in resistance is due to either a poor bond or an error in the gage factor. A sufficient number of samples described above should be made and tested until confidence in the bonding technique is developed.

2.1.6.5 Special Uses

Bonded strain gages have many special uses. They are often used as the sensing element in transducers designed to measure variables other than strain. For example, semiconductor gages can be mounted on a thin diaphram and used as the sensing element in a pressure transducer (see Section 2.1.5). Other examples include designs to measure acceleration (see Section 2.1.4), displacement, force, angular motion, and torque.

2.1.7 Other Special Purpose Transducers

The vibration and acoustic transducers described above provide measurements at a single point on the test object. For many measurement problems it is necessary to measure the variation of vibration over a spatial region. The most common example is when a spatial derivation of the vibration is needed. Measurements of acoustic intensity or angular vibration both require that a derivative be measured.

The usual technique for measuring a spatial derivative is to use two or more sensors a small distance apart. The derivative is expressed as the difference in the signal output from these transducers. Since the signals to be measured are small and a difference between signals is needed, the dynamic range of these multiple transducer devices is limited.

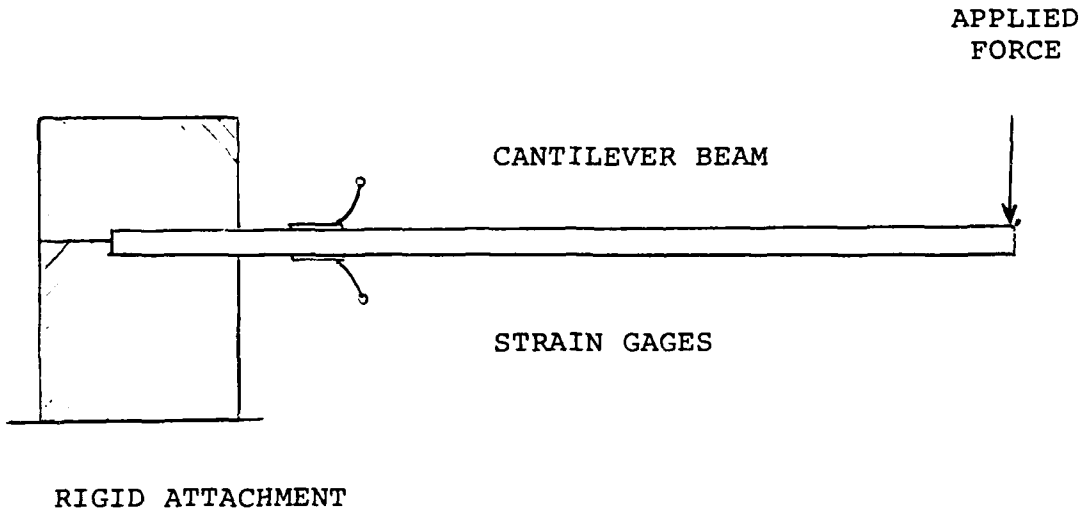


Fig. 2-11 Strain Gage Test Set-up

Other special purpose transducers provide direct measurements of angular vibration or acoustic pressure gradient. Use of these transducers can require special mounting procedures. However, their use is often required to meet the objectives of the measurement program.

2.2 SIGNAL CONDITIONING EQUIPMENT

2.2.1 General Requirements

The sensing devices described in the previous section provide an electrical signal representing a physical variable that we wish to measure, such as acceleration, strain, or pressure. The signal from the transducer must be conditioned before it can be analyzed or recorded. This conditioning includes several factors. For all passive transducers in which the variable to be measured is represented as a changing resistance or capacitance, the conditioning equipment must provide a voltage or current power supply to obtain an electrical signal from the transducer. The signal must then be transferred from the transducer to a remote, centralized location where that data is to be recorded or analyzed. Its voltage range and source impedance must be conditioned to be within the requirements of the recording device or analysis system. The conditioning equipment must carry out these functions while preserving the desired accuracy of the measurements and within available space, weight, and electrical power. In addition, the equipment should facilitate calibration of the vibration and acoustic measurement system and should allow flexibility in the choice of a recording or analysis device. Finally, the signal conditioning equipment must operate properly under the environmental conditions in which it is to be used.

2.2.2 Identification of Basic Elements

The basic elements for signal conditioning are:

1) cables
2) power supplies
3) impedance transformers and preamplifiers
4) amplifiers
5) filters
6) encoders and transmitters, including A to D converters

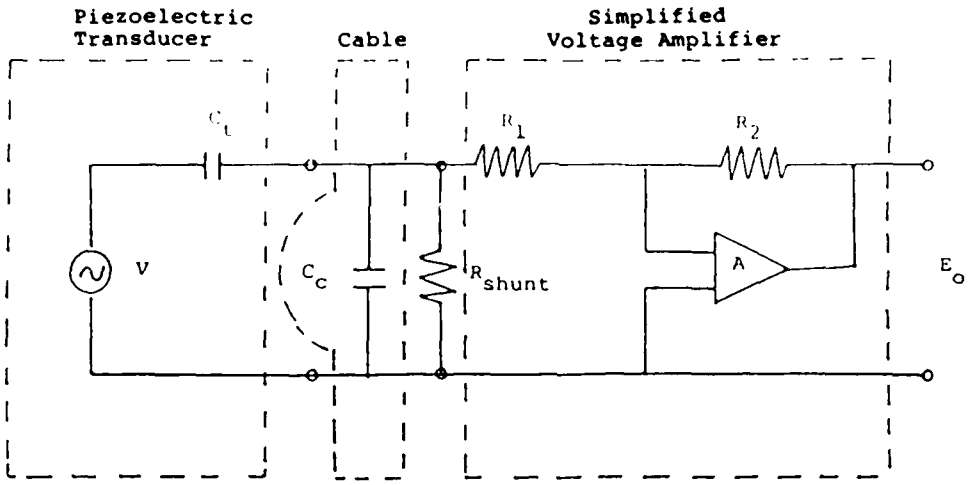
Although the selection of the signal conditioning elements depends in large part on the transducers being used, the cost of these items makes it desirable to select equipment that is of sufficient versatility that it can be used in a variety of applications.

The life of most commercially available signal conditioning equipment, given reasonable maintenance, is 3 to 5 years or longer. Thus, in selecting equipment consideration should be given not only to the present application but to future needs.

2.2.3 Conditioning for Piezoelectric Transducers

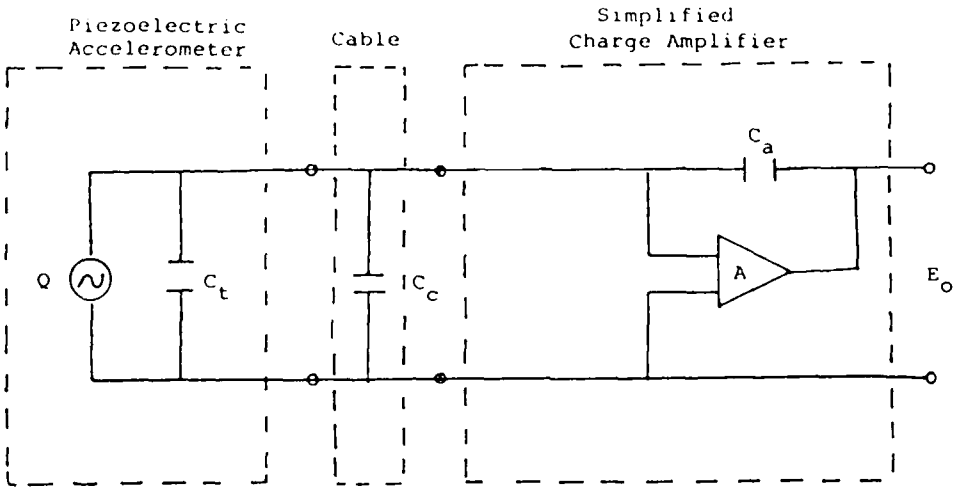
The piezoelectric transducer signal is a fluctuating electrical voltage. The output can be equivalently represented by the charge equivalent circuit in Figure 2.12.a or the charge equivalent circuit in Figure 2.12.b The transducer sensitivity is expressed either as a charge sensitivity (i.e. coulombs per unit acceleration) or a voltage sensitivity (i.e. volts per unit acceleration). The voltage sensitivity is usually the open-circuit voltage that would exist if the transducer were attached to a signal conditioning amplifier with infinite input impedance.

The capacitance of the piezoelectric transducer is small which results in a high source impedance. Because of this high



$$\frac{E_o}{V} = \frac{jw\ R\ C_t}{1 + jw\ R\ (C_t + C_c)} \qquad \text{with } R = \frac{R_1\ R_s}{R_1 + R_s}$$

Fig. 2.12.a Signal Conditioning with a Voltage Amplifier



$$\frac{E_o}{Q} = \frac{1}{C_a}$$

Fig. 2.12.b Signal Conditioning with a Charge Amplifier

impedance, special care must be given in selecting the transducer cable and preamplifier. Although it is possible to condition the output of a piezoelectric transducer with a general purpose amplifier, (these are discussed in Section 2.2.7), most measurement applications require the use of a specially designed preamplifier. These preamps, which may be included in a general purpose amplifier, are in the form of either a high impedance Field Effect Transistor (FET) or a charge amplifier.

The effect of a cable from the transducer to the amplifier is to add a shunt capacitance as shown in Figure 2.12. This capacitance has no effect on the charge sensitivity of the transducer, but reduces the voltage sensitivity by the ratio $C_t/(C_t + C_c)$, where C_t is the transducer capacitance, and C_c is the cable capacitance. Because of the effect of the cable capacitance on the voltage sensitivity, many transducer manufacturers provide a sensitivity including a specific cable which they provide. Other manufacturers provide transducers with FET signal conditioning circuits built into the transducer case. For these transducers the circuit representations shown in Figure 2.12 are not applicable.

To condition the signal from a piezoelectric transducer, it is necessary to use a high input impedance voltage amplifier or a charge amplifier. A simplified schematic of the voltage amplifier is shown in Figure 2.12a. The amplifier gain is governed by the ratio R_2/R_1 while the input impedance is equal to the value of R_1. Although the input impedance can be made to be very large, it is not infinite so that the voltage amplifier interacts with the piezoelectric transducer. The interaction reduces the transducer output at low frequencies. The frequency response of the transducer and voltage amplifier is governed at low frequencies by the RC time constant, where R is the input impedance of the amplifier, and C is the combined impedance of

the transducer and cable. The low frequency amplitude response is given by

$$|H(\omega)| = \sqrt{\frac{1}{1 + \frac{1}{(\omega RC)^2}}} \qquad (2\text{-}9)$$

where $|H(\omega)|$ is the amplitude response and ω is the radian frequency. This function is -1 dB at $\omega RC = 2$ and -3 dB at $\omega RC = 1$. At frequencies well below 1/RC the frequency response has a slope of 6 dB/octave. For example, if we consider a transducer having a capacitance of 950 pf with a cable capacitance of 50 pf and a signal conditioning amplifier with an input impedance of 10 MΩ , the cut-off frequency (at which the frequency response is -3 dB) is equal to

$$f_c = \frac{1}{2\pi}\frac{1}{RC} = \frac{1}{2\pi}\frac{1}{10^7}\frac{1}{10^3 10^{-12}} = 16 \text{ Hz} \qquad (2\text{-}10)$$

The transducer output at frequencies below 16 Hz will be greatly attenuated and the measured data at these low frequencies will be invalid.

In some cases the attenuation of low frequencies aids in the signal conditioning. For example, if the frequency range of interest lies above 100 Hz, there is no need to preserve the low frequency response of the transducer. In fact, the signal at low frequencies may only serve to overload the recording equipment. In this case a shunt resistance can be added to the input of the voltage amplifier to adjust the cut-off frequency to a desired value

$$f_{\text{cut-off}} = \frac{1}{2\pi}\frac{1}{C_t + C_c}\left[\frac{1}{R_{\text{shunt}}} + \frac{1}{R_{\text{amp}}}\right] \qquad (2\text{-}11)$$

where $f_{cut\text{-}off}$ is the frequency at which the response will be 3 dB down, C_t is the transducer capacitance, C_c is the cable capacitance, R_{shunt} is the shunt resistance, and R_{amp} is the input impedance of the voltage amplifier.

The two major problems associated with a voltage amplifier - the effect of cable capacitance on transducer sensitivity and the roll-off of low frequency response - may not be acceptable for a particular data collection system. In this case the signal conditioning should be done with charge amplifiers. A simplified schematic for this type of amplifier is shown in Figure 2.12.b. This type of amplifier detects the electrical charge across the piezoelectric element of the transducer. Since the voltage at the input to the charge amplifier is driven to a very small value (zero in the case of an ideal amplifier where the gain of the operational amplifier is infinite) the cable capacitance has no effect on the output voltage of the charge amplifier. This output voltage, E_0, is given by

$$E_o = V \frac{C_t}{C_a} \qquad (2\text{-}12)$$

where V is the open circuit voltage output for the transducer, C_t is the transducer capacitance, and C_a is the feedback capacitance for the charge amplifier. The charge amplifier output is more commonly written in terms of the charge output of the transducer, Q;

$$E_o = Q \frac{1}{C_a} \qquad (2\text{-}13)$$

where the factor $1/C_a$ becomes the gain of the charge amplifier.

The output of the charge amplifier does not exhibit the low frequency roll-off seen previously for the voltage amplifier.

This allows piezoelectric transducers to be used at very low frequencies. In cases where this low frequency response is not needed, electronic filters can be used following the signal conditioning charge amplifier. However, care must be taken to avoid overloading the charge amp. Depending on the overload characteristics of the charge amp it may or may not be possible to determine if it has been overloaded by looking at the output of the filter.

As a general rule charge amplifiers are preferred over voltage amplifiers for signal conditioning on piezoelectric transducers. However, because of the higher cost and susceptability to electromagnetic pickup when long cables must be run from the transducer to the charge amplifier, voltage amplifiers are also used.

The dynamic range for the signal conditioning amplifier is set on the high end by the maximum voltage output of the amplifier, which is typically 1 to 10 v, and on the low end by the amplifier noise. The high source impedance of the piezoelectric transducer requires a high impedance signal conditioning preamplifier. Field Effect Transistor (FET) input stages are typically used. The noise produced by these amplifiers depends in part on the source impedance.

2.2.4 Conditioning for Piezoresistive Transducers

The main signal conditioning requirement for a piezoresistive (PR) accelerometer or pressure transducer is an external power supply. The output signal from a PR accelerometer or pressure transducer is a static or fluctuating voltage requiring little further conditioning, since PR transducers have a low output impedance and very low noise level.

The equivalent circuit for a PR transducer is illustrated in Figure 2.13. The variable resistance in this circuit is directly proportional to the sensed quantity, acceleration or pressure.

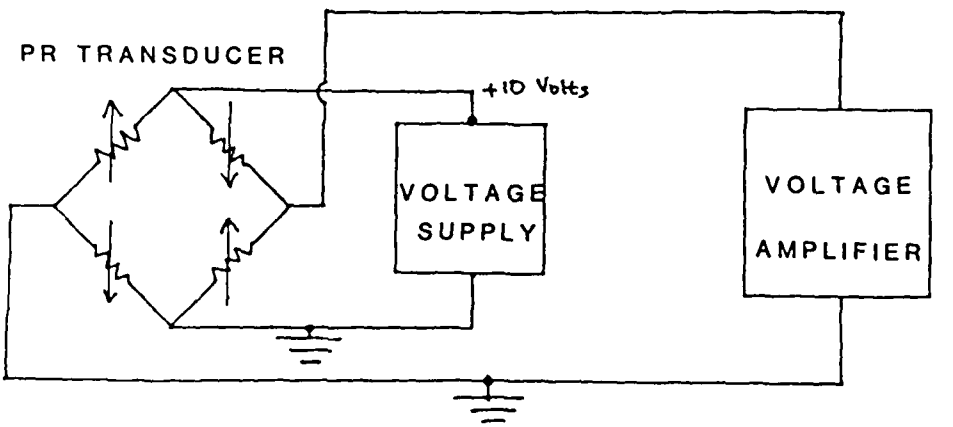


Fig. 2.13 Signal Conditioning for a Piezoresistive Transducer

The low output impedance from the bridge makes the system insusceptibile to electromagnetic and electrostatic noise.

The most important requirements of a PR signal conditioner is to provide a D.C. supply voltage to the bridge of the transducer. There are two commercial types of conditions for PR pressure transducers, both using a Wheatstone Bridge strain gage, variable resistance configuration. The signal conditioner in the first type provides a highly-regulated, constant current to the transducer bridge. The specific current can be adjusted on the signal conditioner to obtain the nominal transducer sensitivity. The constant current type of power supply is useable over a wide temperature range.

The second type of signal condition provides a highly regulated, constant voltage to the transducer bridge. The sensitivity of the transducer is directly proportional to the DC supply voltage, and setting the supply voltage to factory specification results in the nominal transducer sensitivity (in mV/psi).

In order to make static pressure measurements, the signal conditioner should have a zero-balance adjustment to null the preamplifier output in the absence of an input signal. Since variations in the supply voltage result in stray drift in the calibration, it is important that the power supply be regulated.

Commercial signal conditioners use an attenuator amplifier which provides a normalized output (±2.5 V) for inputs of 20,40, 60,80 and 100% of the full scale range of PR transducers. Such conditioners can be used for either accelerometers or pressure transducers.

2.2.5 Conditioning for Condenser and Electret Microphones and Other Capacitive Devices

Condenser and electret microphones generate a signal proportional to fluctuations in capacitance. The polarizing

voltage across the capacitance is supplied by an external power supply in the case of a condenser microphone, and supplied by the embedded electret layer in the case of an electret microphone. The output signal from the condenser or electret microphone must be further conditioned by a pre-amplifier due to the high output impedance of these capacitive devices.

The equivalent circuit for the condenser microphone, cable and the input stage of the preamplifier is illustrated in Figure 2.14. The microphone is modeled as a current source $j\omega C_T E_o$, where C_T is the transducer compliance and E_o the open circuit voltage. The cable compliance is C_c and the input stage of the preamplifier has an input capacitance C_p in parallel with the input resistance R_p. The preamplifier voltage signal E_p is related to E_o by:

$$E_p/E_o = \frac{C_T}{C_T + C_c + C_p + (\omega R_p)^{-1}} \qquad (2\text{-}14)$$

The input resistance is large $(\sim 10^{11}\ \Omega)$ so that Equation (2-14) is:

$$E_p/E_o = \frac{C_T}{C_T + C_c + C_p} \quad , \quad \omega R_p (C_T + C_c + C_p) \gg 1 \qquad (2\text{-}15)$$

throughout most of the frequency range. Thus the preamplifier voltage E_p is simply proportional to the open circuit voltage and hence the acoustic pressure impinging on the condenser microphone diaphragm.

The signal to noise ratio is enhanced by increasing E_p/E_o. Hence, cable capacitance is eliminated by screwing the microphone directly onto the preamplifier. The transducer capacitance is on

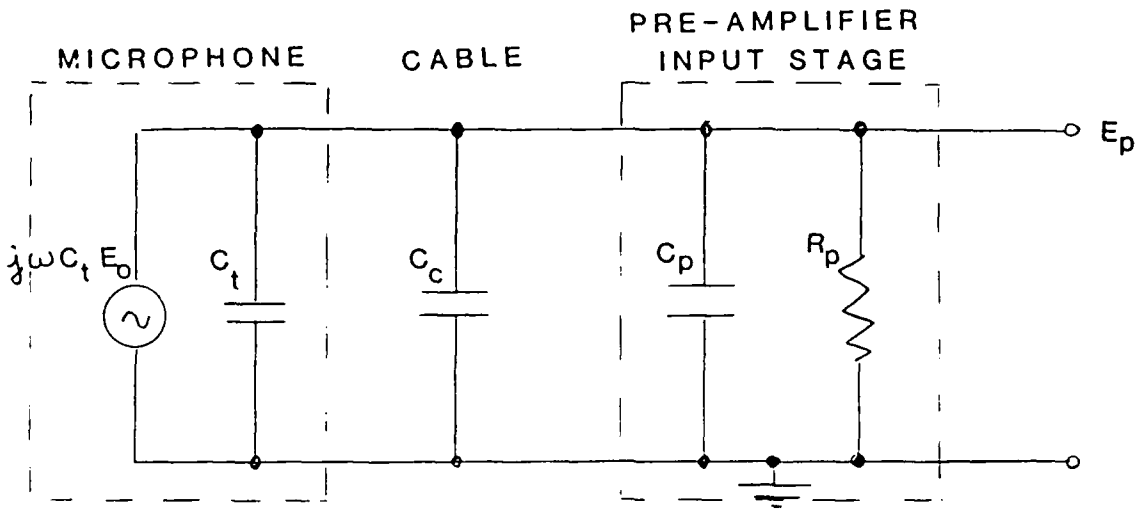


Fig. 2.14 Signal Conditioning Circuit for Condenser Microphone

the order of 60 pF, and the preamplifier input capacitance is on the order of 0.4 pF. Thus the voltage loss is less than 1 dB.

At very low frequencies such that $\omega R_p (C_T + C_c + C_p) << 1$, the transmission is:

$$E_p/E_o = \omega R_p C_T \qquad (2\text{-}16)$$

The transmission rolls off at low frequencies at the rate of 6 dB/octave. The response of commercially available condenser microphones rolls off in the range of 2-5 Hz.

The preamplifier output stage has a low output impedance (50 Ω), and thus can drive large cables. Peak output voltages and currents are typically in the range of 4 V and 15 mA, respectively.

The self-generated preamplifier noise spectrum is of the order of 70 μV for a 1/4" condenser microphone and 120 μV for a 1/8" condenser microphone. The preamplifier operates in the temperature range from -20°C to 100°C.

Commercial preamplifiers have adjustable gain settings, typically -20 dB, 0 dB and 20 dB.

The condenser microphone requires an external power supply, and the requirements for this power supply are discussed in Section 2.1.5.1. In contrast, the electret microphone does not require an external power supply.

2.2.6 Conditioning for Strain Gages Transducers

Strain gages and strain gage transducers exhibit a change in resistance that is proportional to the signal being measured. Signal conditioning depends on whether or not a requirement exists for measurements of static or DC data.

The simplest method for conditioning the output of the strain gage is to use the "potentiometer" circuit shown in Figure 2.15. In this circuit the ballast resistor is selected so that

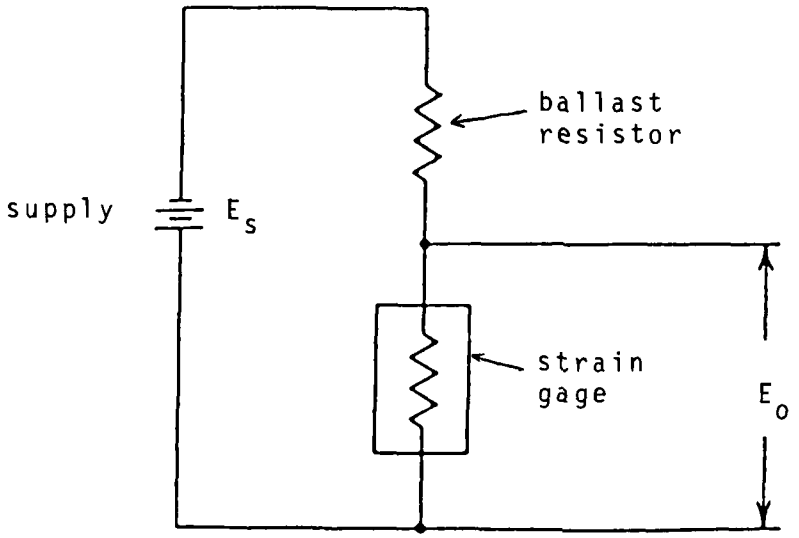


Fig. 2.15 Potentiometer Circuit for Strain Gages

its resistance is large compared to that of the strain gage. Its purpose is to provide a relatively constant current through the gage that is independent of changes in gage resistance due to applied strain. In applications where a high degree of accuracy is required, the DC voltage supply and ballast resistor should be replaced by a constant current source.

The output signal for the potentiometer circuit is the voltage across the gage. This signal contains a DC component which is present without strain being applied to the gage. Although this DC voltage changes with the static strain on the gage, the accuracy of the potentiometer circuit is poor for static measurements, since the voltage change to be measured is very small compared to the DC voltage present without strain. This potentiometer circuit should only be used for dynamic measurements where the DC output of the circuit can be blocked using an AC amplifier.

In cases where a static component of stress is to be measured, the "Wheatstone Bridge" circuit is used. This circuit is shown in Figure 2.16. It allows the bridge to be balanced under a condition of zero signal so that the output voltage is zero. The primary advantage of the Wheatstone Bridge circuit is that it can be used for both static and dynamic measurements. In addition, it allows greater freedom in connecting the strain gages to form temperature compensating networks. Its disadvantage is the need for bridge balancing before each measurement and the potential requirement for a differential rather than single-ended amplifier.

Although the bridge circuit can be used for dynamic measurements, it offers no particular advantage over the simpler potentiometer circuit for this type of measurement.

Two types of signal conditioning amplifier are available commercially. One type is specially designed for dynamic measurements. The gage is excited by a constant-current source

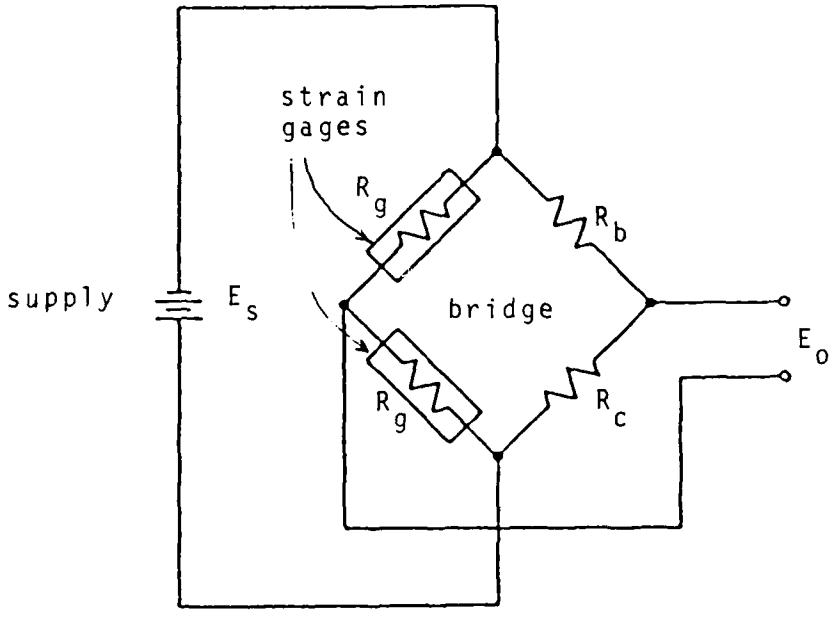


Fig. 2.16 Bridge Circuit for Strain Gages

which can be adjusted over a range of currents. An AC low-noise amplifier is then used.

2.2.7 Amplifiers

The role of the amplifier in a data collection system is to convert the output of the transducer or signal conditioning preamplifier (if present) to a signal input for a recording device or analysis equipment. This requires a voltage amplifier with a gain that can be adjusted over a wide range of values.

Four options exist for gain control. The simplest is a manual gain which is adjusted by a multi-position switch on the front panel of the amplifier. To achieve high accuracy the gain switch has fixed steps, usually in a 1, 2, 5, 10.... sequence or 1 dB steps. However, continuously adjustable gain verniers are also commonly used in conjunction with a fixed gain switch having 10 or 20 dB steps.

The manual gain adjustment is the simplest form of gain control and allows the design of relatively low cost amplifiers. It is also easy to use. Its major shortcoming is in multi-channel data collection systems where the manual adjustment of each channel is too time consuming. In addition the need to keep a log book in which the gains of each channel are recorded introduces a major source of human error into the measurement system.

The second option for gain control is a manual adjustment with a gain status readout. An electrical signal is provided from each amplifier with the gain encoded usually in digital format. This signal can be multiplexed with other gain status signals and recorded on one channel of the data recording system. Although the use of the gain status readout is significantly more complex than the simple manual system, it eliminates the need to record the gains in a log book, and, thereby, eliminates that source of error.

A third method for gain control is to use a remote gain-setting option. Two techniques are possible. In the first the amplifiers are set up in a master-slave arrangement. The master device is used to control the gains of the other amplifiers. This greatly simplifies the gain setting procedure. However, it eliminates the flexibility of changing the gain of a single amplifier without changing all amplifier gains. The second technique allows the gain of each amplifier to be set independently. Typically, a digital signal is used to set the gain, although it is possible to accomplish the same objective using voltage controlled amplifiers (VCA's). However, the accuracy in setting the gain using VCA's is not as high as that achieved using a digital gain signal. Many manufacturers follow the IEEE-488 interface standard for the digital gain control signal. This greatly simplifies the process of including gain controlled amplifiers in a computer-aided data collection system.

The fourth method for gain control is to use automatic gain-setting or "auto-gain" amplifiers. These units automatically set the gain of the amplifier so that the output signal is within a limited dynamic range. The auto-gain amplifier detects the average output signal level over a short period of time set by a time-constant switch. If the average level is above a set threshold, the gain is reduced by a fixed amount. Similarly, if the average level is below a lower limit threshold, the gain is increased by a fixed amount. In most commercially available auto-gain amplifiers the averaging time constant, the gain-decrease and gain-increase thresholds, and the step size of the gain change can be adjusted within certain limits. The amplifier's characteristics can, therefore, be adjusted to match the characteristics of the recording device.

The performance of the auto-gain amplifier is seriously degraded for certain types of data. In particular, these amplifiers are not useful for measurement of single transient

events or shocks. Preceeding the start of the event in a period of "no signal" the gain of the amplifier will be set to its maximum value. Due to the high gain the amplifier will overload when the event occurs. After a duration of one time constant the gain will be reduced by one step (typically 6 dB). At this point the event may be over or an overload may still occur, since the gain can be reduced only one step each time constant. In either case valuable data has been lost.

Many auto-gain amplifiers allow the auto-gain feature to be disabled. This should be done when using these amplifiers for measurement of transient data such as blasts, impacts, and other shocks.

The use of auto-gain amplifiers increases the cost and complexity of the measurement system. Therefore, they should only be used when needed to reduce the dynamic range of the output signal. The most common application is in large multichannel data collection systems where the data are to be recorded on a magnetic tape. The dynamic range of the tape is quite limited and often is significantly less than the range of preceeding transducers and amplifiers. This limitation combined with some uncertainty in predicting the level of the data to be measured makes the auto-gain amplifier a worthwhile improvement to the measurement systems.

When many data channels are needed, the gain status of the individual auto-gain amplifiers can be multiplexed and recorded on a single channel of the recording device.

The primary role of the instrumentation amplifier in a data collection system is to amplify the signal from the signal conditioning preamplifier so that its voltage range is compatible with recording equipment or data analysis equipment. Although the accuracy of the amplifier is of critical importance in selecting a particular model, most commercially available amplifiers have accuracies that are significantly better than other elements of

the instrumentation chain. Therefore, in practice a selection is usually made on the basis of flexibility and convenience of use.

2.2.7.1 AC/DC Amplifier

The first step in selecting an amplifier is to decide on whether an AC or DC amplifier is needed. The AC amplifier is simpler to use and typically less expensive. Its frequency response extends down to appproximately 1 Hz, which is adequate for most vibration and acoustic measurements. An even lower frequency response is available on many AC amplifiers. However, if true DC response is required, a DC amplifier must be used.

The upper limit of the amplifier frequency response can extend to the megaHertz region. However, this extended frequency response is not needed for vibration and acoustic measurements (excluding, ultrasonic and acoustic emission measurements), so that the frequency response for most amplifiers designed for these types of measurements extends up to 100 or 200 kHz.

2.2.7.2 Amplifier Inputs

The second decision to be made in selecting an amplifier is to determine if a single-ended or differential input is needed. A single-ended input is more commonly used and provides good performance in systems where a common ground can be provided for all elements of the instrumentation chain. In measurement systems where the amplifiers must be located a hundred feet or more from the amplifiers and where the inductive pickup of noise voltages from nearby power lines and motors is possible, spurious voltage signals appear on both the active and ground input to the amplifier. Use of well shielded cable is a necessity for these situations. But in many cases it is necessary in addition to use a differential input.

The performance of a differential input amplifier is specified by its common-mode rejection ratio (CMRR). This ratio

expresses the ability to reject signals that are common to both the active and ground inputs to the amplifier. Values for the CMRR are typically in the range 60 to 90 dB.

The use of differential mode inputs requires special connectors and cables. The cables should be twisted-pair conductors with a separate isolated ground for the shield. Due to the inconvenience of changing all cables and connectors, a decision on the need for differential inputs should be made early in the measurement system design.

2.2.7.3 Frequency Response and Input/Output Impedance

As general requirements in selecting an amplifier, the frequency response should be flat over the bandwidth of interest, and the input and output impedances must be sufficiently high at the input and low at the output so as not to load the other electronic devices connected to the amplifier.

We assume that a signal conditioning amplifier has been used to condition the transducer signal and to provide a low impedance input to the amplifier. Thus, the requirement that the input impedance of the amplifier be significantly greater (a factor of 100 for good accuracy) than the source impedance is not severe. In fact, many amplifiers provide a sufficiently high input impedance that they can be used directly to provide both transducer signal conditioning and gain.

The requirement on the amplifier output impedance is also not severe since the input impedance of most recording devices and analysis equipment is typically many thousands of ohms.

The frequency response of the AC amplifier is typically specified by the frequencies at which the amplitude response is down 3 dB. To obtain greater accuracy over the frequency range of interest the upper and lower frequency limits of the amplifier should be set approximately two octaves (a factor of four) above and below the limits of interest. Although this procedure

increases the amplifier noise by allowing signals with frequency content outside the range of interest, it is required to obtain high accuracy at frequencies near the upper and lower frequency limits of interest.

The phase accuracy of the amplifier is limited to a narrower frequency range than the amplitude response. To obtain good phase accuracy (~ ± 10 degrees) the upper and lower frequency limits of the amplifier should be one decade (a factor of 10) beyond the frequency range of interest.

Several commercially available amplifiers provide low-pass and/or high-pass filters that can be set to limit the frequency response range. These filters are useful in limiting noise due to out-of-band signals. However, they have a significant effect on the frequency response near their cut-off frequencies and should be carefully used.

Although it is a fairly simple task to construct a fixed-gain instrumentation amplifier using operational amplifiers, most measurement systems use a set of rack-mounted, commercially obtained amplifiers. The primary reason for this is the availability of special features that simplify the data collection process. The most common feature is an adjustable gain. Options include manual adjustment in fixed steps over a full 90 to 100 dB range and a remote controlled feature whereby the gain can be set by an electronic computer. Gain accuracy is typically very good (± 1%) over the entire range.

Many amplifiers can be obtained with input connectors and a power supply for a remote signal conditioning preamp. This feature is particularly useful for high impedance transducers where the signal conditioning equipment must be located remote from the amplifier/recording station and near the transducers.

2.2.7.4 Special Features

Special calibration features are also available on many units. These allow a simple calibration signal to be switched into the input circuit of the amplifier. This signal serves as a system calibration for the amplifier and recording devices.

Other features to be considered include battery operation, special power supply voltages and live frequencies, rack or cabinet mounting, and monitoring meters.

2.2.7.5 Overload Performance

The response of the amplifier to an overload signal and its recovery time is important to the performance of the overall instrumentation system. Ideally, the amplifier should simply clip the signal at an upper voltage unit, say ± 5 volts, and recover instantly when the signal falls below the ± 5 volt limit. Amplifier designs approach this ideal with recovery times under 1 msec. However, some older designs and particularly some DC amplifiers have poor recovery times of several seconds.

A fast recovery time is particularly important in auto-gain amplifiers, since data will be lost in the recovery period after each gain change. In addition it is important that the switching transients be as small as possible.

2.2.7.6 Line and Battery Operation

Many amplifiers allow either line or battery operation. Often it is possible to operate several units from a single power supply. This reduces the overall system cost and simplifies system maintenance, since failure of the power supply is not uncommon.

The amplifier performance is strongly tied to the power supply. Any ripple or noise in the power supply DC voltage can be transmitted to the amplifier output. Thus, it is best to

obtain both power supply and amplifier from the same manufacturer.

Battery operated amplifiers offer two advantages. First, they allow a unit to be easily transportable since the required batteries are not very large and are often included within the amplifier housing. A second advantage is the elimination of a ground through the power supply. Although the input ground is usually isolated from the power supply, the isolation is not always total so that the possibility of a ground loop exists. An obvious disadvantage in the battery supply is its finite life. Battery life is particularly limited at lower temperatures and often the battery performance below 40^{o}F is unacceptable.

Most amplifier power supplies offer a range of line frequencies up to 400 Hz and voltages up to 250 V. Particular combinations are available either as standard features or as options.

2.2.7.7 Calibration Procedures

The gain accuracy of an amplifier can easily be calibrated using an oscillator and digital AC volt meter. One convenient approach is to connect two amplifiers in series. The gains can then be adjusted over the entire range of the amplifiers without a change in output level so that the accuracy of the volt meter does not come into play.

The distortion and output noise can also be checked using an oscillator and a spectrum analyzer. In performing this check it is important to realize that the power supply can influence both distortion and noise and that the output noise is dependent on gain. The highest output noise occurs at maximum gain. However, if the noise is referred to the input (i.e. the output rms noise is reduced by the amplifier gain factor), then the minimum noise typically occurs within the middle range of gain settings.

Many amplifiers offer a calibration option. This allows a single frequency source to be switched into the output. This offers a method for system calibration of the amplifier and subsequent recording appartus. It is not a replacement for a careful laboratory calibration.

2.2.7.8 Data Monitoring

In most data collection efforts it is desireable to monitor each data channel to insure that overloads are not occurring and that the measured data are above the noise floor. The appropriate place to monitor the data is at the output of the amplifiers. The use of monitoring meters or equipment at the input to the amplifier or at the input to the signal conditioning amplifier may result in a loading of these devices that affects their accuracy.

Monitoring devices include overload lights, meters, and oscilloscope displays. Overload lights are generally sufficient if combined with the ability to selectively monitor a single channel using a meter or oscilloscope. The oscilloscope is preferred since it allows the actual signal to be monitored rather than its level.

2.2.7.9 Cabling and Connectors

Most instrumentation amplifiers are single ended and use BNC connectors. This type of connector combined with high quality shielded cable offers good performance over a variety of conditions.

Even with high quality cables and connectors cable failure is a problem. It is tempting and possible to make your own BNC cables. However, commercially available cables with a molded rubber boot at the cable connector to provide stress relief offer much more reliable performance and are worth the extra cost.

Most manufacturers offer other types of connectors as options. The most significant case in which a conventional BNC connector cannot be used is for a differential input or for cases in which the cable must provide power to a preamplifier. There is no set standard for these cases so that care must be taken in ordering the equipment to see that the proper connectors are obtained.

2.2.8 Filters

In many measurement applications the frequency bandwidth over which the transducer provides an output exceeds the frequency range of interest. In this case electronic filters should be used to condition the signal by limiting its bandwidth. In most data recording systems it is necessary to limit the input signal bandwidth in order to realize the maximum signal-to-noise ratio. Also, in digital signal processing systems it is necessary to limit the bandwidth of the input 90 gnol = > signal to the A to D converter in order to prevent aliasing. These are two cases in which the use of filters is an important feature of the measurement system.

Filters are also used to process vibration and acoustic data in the frequency domain. Although the trend in data processing is toward digital filtering, which is discussed in detail in Chapter 3, electronic analog filters still play an important role in many data processing systems.

2.2.8.1 Types Available

Filters are generally categorized by their frequency response. Four basic types exist: low-pass, high-pass, band-pass, and band-rejection or notch filters. The ideal low-pass filter passes all frequencies below a certain cut-off frequency with unity gain and rejects all frequencies above this frequency. A filter of this type may be used preceeding an A to

D converter to prevent aliasing in a digital signal processing system or it may be used to reject the output of a piezeolectric accelerometer at the transducer's resonance frequency.

The ideal high-pass filter rejects frequencies below a certain frequency limit and passes all higher frequencies. This type of filter is often used in vibration and acoustic measurements on moving vehicles, where the low frequency motions of the vehicle are not of immediate interest.

The high-pass filter can also be combined with the low-pass filter to form a band-pass filter. This type of filter is commonly used to analyze vibration and acoustic data. Acoustic data are often analyzed using octave or one-third-octave band-pass filters. In the octave filter the ratio of the upper frequency limit of the pass-band to the lower frequency limit is a factor of two. Thus, a filter with a pass-band from 200 to 400 Hz would be categorized as an octave filter. The octave band center frequency is equal to $\sqrt{2}$ times the lower frequency limit of the pass-band. For the example above the band center frequency is 283 Hz. Where a choice of frequency band is possible, the ANSI/ISO standards for band center frequencies should be adhered to.

The band-rejection or notch filter is used to reject a certain band of frequencies. It is most commonly used to reject a narrow band of frequencies centered on 60 Hz in order to eliminate electromagnetic pickup due to ground loops and other problems in the measurement system. Although notch filters can be used to eliminate the signal at 60 Hz and its harmonics, a better approach is to reduce the electromagnetic pickup by eliminating all ground loops and by shielding the components of the measurement system.

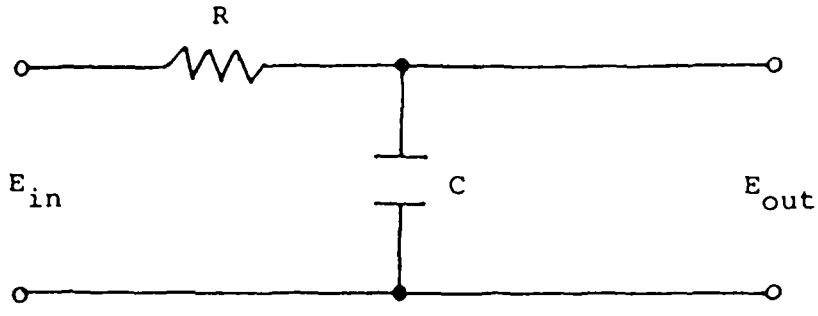
Although a filter can be implemented in a variety of physical forms, the electronic filter is most common and should be used in all vibration and acoustic measurement systems.

Two basic types of electronic filters exist: active and passive. These two types operate in the analog domain as opposed to digital filters which are described in Chapter 3.

The active filter is designed using a combination of resistors, capacitors, and amplifiers. A characteristic of the active filter is that it allows the design of filters with a broad range of desirable frequency response characteristics. Passive filters use a circuit of resistors, capacitors, and inductors. Although elimination of the amplifier offers certain advantages, the many disadvantages of a passive filter preclude it from use in all but a few very simple applications. There are cases where the need for the desirable features of a passive filter - low noise, wide dynamic range, low cost, and lack of a requirement for a power supply - outweigh the disadvantages.

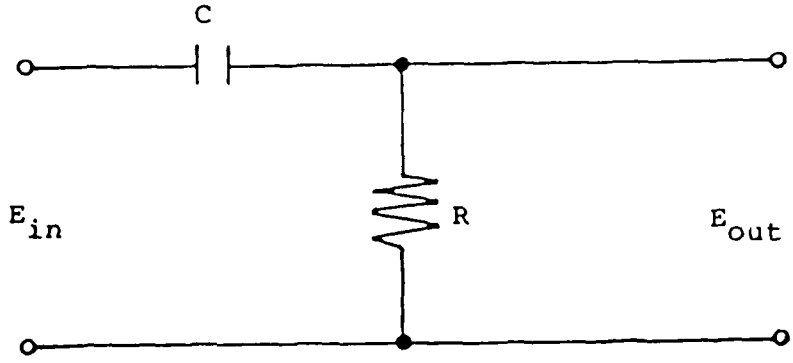
2.2.8.2 Passive

Simple, passive low-pass and high-pass filters are shown in Figures 2.17a and b. These simple designs illustrate the advantages and disadvantages of the passive filter. The use of a single resistor and capacitor results in very low noise (limited by thermal noise in the resistor), wide dynamic range (typically limited by the maximum voltage across the capacitor), and low cost (part costs are negligible unless a very large capacitor is needed). On the other hand the performance of the design is not great. The attenuation at high frequencies is moderate with the amplitude decreasing 6 dB/octave. In addition, the filter response near the cut-off frequency shows a gradual change in amplitude rather than a sharp cut-off. By adding more stages to the filter design it is possible to increase the performance. However, in such a design it is necessary to take into account the interaction of the stages by an impedance analysis. The performance can also be increased by use of inductors in the circuit. However, inductors are generally susceptible to



$$\frac{E_{out}}{E_{in}} = \frac{1}{1+j\omega RC}$$

Fig. 2.17.a Passive Low Pass Filter



$$\frac{E_{out}}{E_{in}} = \frac{jwRC}{1+j\omega RC}$$

Fig. 2.17.b Passive High Pass Filter

magnetic pickup and tend to saturate, which lends to nonlinearity and signal distortion. The interaction effects in multiple stage circuits employing several inductors are particularly difficult to deal with.

As a practical rule filters using inductors should not be used for vibration and acoustic measurement systems. However, the simple passive low-pass and high-pass filter designs shown in Figures 2.17a and b are of practical usefulness. The application of a simple passive low-pass filter to reduce the signal from a piezoelectric accelerometer at its resonance frequency is shown in Figure 2.17c. The filter design must take into account the impedance characteristics of both the accelerometer and the signal conditioning amplifier. The need to consider the interaction between the filter, the accelerometer, and the amplifier makes it desirable in many cases to replace the passive filter with an active filter. However, great care must be taken to assure that the signal at the accelerometer resonance does not overload the input stages of the active filter. Often it is very difficult to determine from the output of the active filter that the input has been overloaded. The very wide dynamic range and low noise of the passive low-pass filter assures that it will not be overloaded. The most common application of this design is for shock tests where accelerometer "ringing" at its resonance frequency can be a large problem.

A second practical application of the passive filter is as an inexpensive high-pass filter. A filter design is shown in Figure 2.17d. The characteristics of many transducers and their associated signal conditioning amplifiers are such that the frequency response extends to very low frequencies. In many cases, in particular for acoustic and vibration measurements on moving vehicles, the low frequency response is not of interest and serves only to overload the recording equipment. An active filter can be used and is often built into the signal

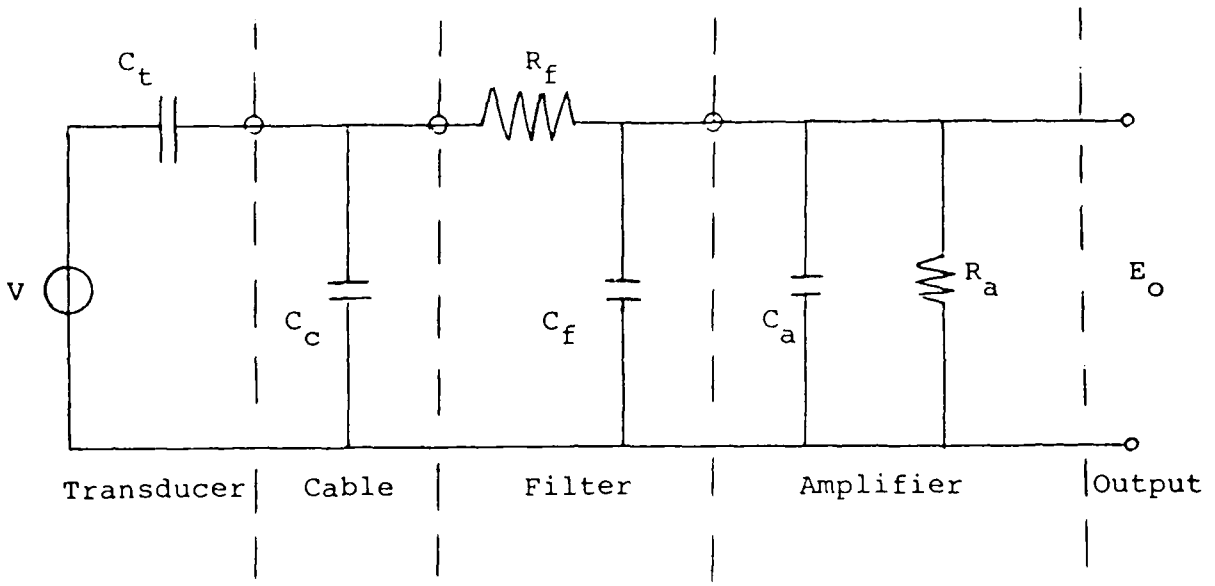


Fig. 2.17.c Passive Low Pass Filter Used with Transducer

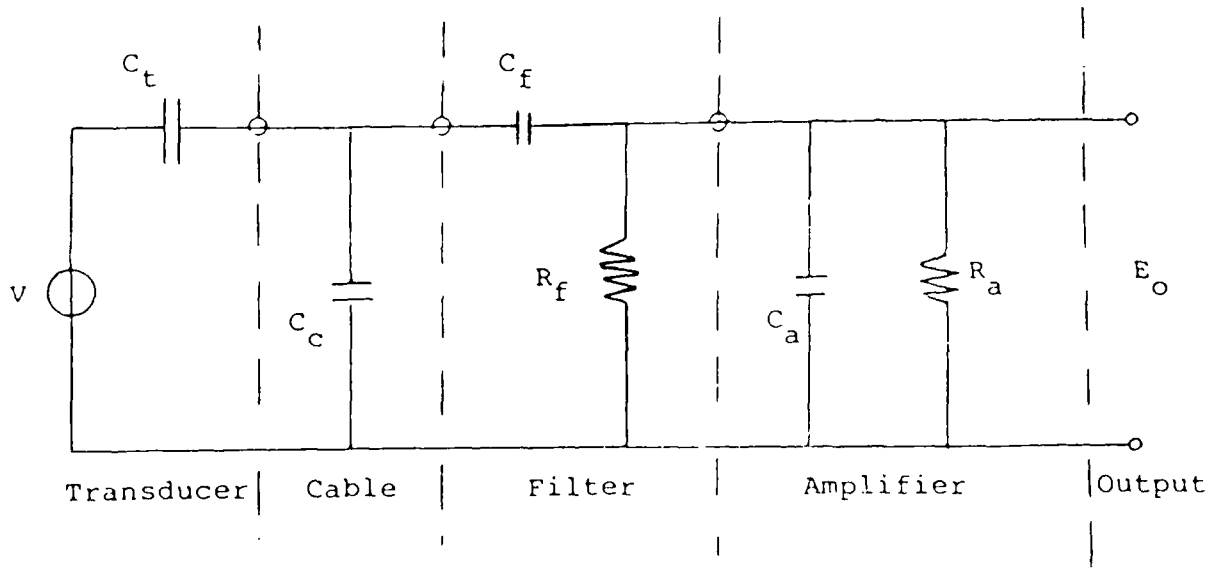


Fig. 2.17.d Passive High Pass Filter Used with Transducer

conditioning amplifier. However, if it is not, the passive high-pass design is attractive because of its low cost.

2.2.8.3 Active

Active filters are preferred for most vibration and acoustic measurement systems. Although their cost is significantly higher than passive filters, they provide much higher performance and are more conveniently used. Since their performance can be made to be independent of source or load impedance, they can be used for a variety of transducer and signal conditioning amplifiers.

The performance of a filter can be described by its frequency response characteristic or transfer function, $H(j\omega)$, where

$$H(j\omega) = A(\omega)\ e^{j\theta(\omega)} \qquad (2\text{-}17)$$

$j = \sqrt{-1}$, ω is the radian frequency, $A(\omega)$ is the ratio of the amplitudes of the output and input signals, and $\theta(\omega)$ is the phase shift between the input and output. The amplitude response is typically plotted in dB [$20 \log_{10} A(\omega)$], while the phase response is plotted in degrees.

For the ideal filter the amplitude response is one in the pass-band and zero at the other frequencies. The phase response of the ideal filter is linear, which results in a simple delay of the signal without distortion.

An actual filter can only approximate the characteristics of the ideal filter. In most cases the requirement for a specific amplitude response function conflicts with the requirement for a linear phase response. Many filters are optimized for either the best amplitude response or the best phase response.

Active electronic filters are often designed using a pole-zero model. With this model, the frequency response is given by

$$H(j\omega) = \frac{\sum_{n=0}^{N} (j\omega)^n a_n}{\sum_{m=0}^{M} (j\omega)^m b_m} \qquad (2\text{-}18)$$

where N determines the number of zeros in the jω-plane , M determines the number of poles, and the coefficients a_n and b_m determine the filter type.

The two most popular filter types are the Butterworth and Bessel filters. Both are all-pole filters with no zeroes, N=0. The coefficients for the Butterworth filter are selected to give an optimally flat amplitude response function, while those of the Bessel filter are selected to give an optimally linear phase response within the pass-band. The performance of both filter types in the stop band is dictated largely by the number of poles, M. At high frequencies the amplitude response of an M-pole Butterworth or Bessel filter decreases 6 x M dB/octave. Thus, a 4-pole filter has a roll-off of 24 dB/octave, while an 8-pole filter roll-off is 48 dB/octave.

For vibration and acoustic measurements of data, which can be considered to be periodic or random stationary signals, the 4-pole Butterworth filter is a good choice for signal conditioning. For cases in which the signal to be measured is a single or repetitive transient the 4-pole Bessel filter should be used. Many commercially available electronic filters allow a simple switch selection between these two filter types.

Higher performance filters can be obtained by cascading a 4-pole filter or by using a 6 or 8-pole filter. However, it should be pointed out that the performance obtained by cascading two 4-pole filters does not equal that of a single 8-pole filter. A second method for achieving higher performance is to use a filter with both poles and zeros. The elliptic filter

introduces zeros to increase the roll-off near the cut-off frequency. The penalty is pass-band ripple and possible large variations in the attenuation in the stop-band. However, the elliptic filter can be an excellent choice for an anti-aliasing filter where a sharp cut-off is often needed. A 6-pole, 6-zero elliptic filter can be designed to give over 80 dB of attenuation within an octave above the cut-off frequency and a pass-band ripple of ±0.1 dB.

Modern electronic filters offer, in addition to excellent filtering performance, a variety of other features. These include adjustable gain, remote control, and many are programmable through an RS-232 or IEEE-488 Standard communications line.

2.2.9 Encoders

2.2.9.1 Introduction

Encoding of signals from measurement instrumentation is another step of signal conditioning used to prepare the signals for transmission or recording. Encoders change the format of the signals while preserving their content of useful information. Encoding is necessary when the characteristics of the source signals are not compatible with the characteristics of the transmission or recording medium. The format of the encoded signals is chosen to preserve the signal information during transmission or recording. The original signals are recovered for analysis by decoders which perform the inverse operation of the encoders.

A common example of the encoding function is the FM (frequency modulation) process, which preserves the DC and low frequency information in a signal by encoding the signal's instantaneous amplitude variations into frequency variations of a high frequency carrier wave. The encoded signal may then be transmitted by radio wave or recorded on magnetic tape, neither of which media possess DC capability.

In this section, the purpose and types of encoders are presented, and a more detailed discussion is given on a particularly important sequence of encoding - analog-to-digital conversion and PCM encoding.

2.2.9.2 Purpose of Encoding

Encoding is used to match the signal characteristics to the characteristics of the transmission or recording media. The purpose is to avoid loss of information and to optimize the amount of information which can be recovered at the other end of a medium with finite information capacity.

The characteristics which may be altered by encoding are:

- o Number of independent channels
- o Frequency content and bandwidth
- o Dynamic range
- o Information-bearing quantity

The number of independent sensor channels needed for a test may greatly exceed the number of channels available for transmission or recording. The sensor channels may each have associated information, such as gain settings, which must also be preserved for later analysis.

The frequency content of the desired information must somehow be encoded into the available frequency range of the transmission channel. Some signal channels may require more frequency range than others, so that different encoding schemes might be useful for different channels.

The dynamic range of amplitudes from sensor signals is often greater than that available on the transmission medium. If the purposes of the test require more dynamic range than the transmission or recording medium provides, then the range of signal amplitudes must be encoded to "fit" within the range of the medium.

The process of encoding test signals for the reasons stated above often requires that the signal quantity used to transmit the desired information be changed. A simple example is the conversion of voltage signals to current loop for transmission through a long twisted pair cable. In this case, electomagnetic pickup which would reduce the signal's dynamic range can be greatly reduced. In other encoding schemes the signal information may be used to modulate a carrier wave or to modulate a digital pulse train.

2.2.9.3 Multiplexers

Multiplexers are devices which combine two or more channels of information into a single channel. The recovery of the original channels is performed by a demultiplexer. The process of multiplexing applies to both analog and digital operations, but the device called a multiplexer usually operates in the analog domain. There are two basic forms of operation - time multiplexing and frequency multiplexing.

Time multiplexers combine the input channels serially into a single output channel. The term commutator is sometimes used since the process is similar to (and was once performed by) the action of the commutator on an electric motor. Each input channel is sampled for a fixed time (dwell time) before the next is selected. The first channel is not sampled again until after the last channel's dwell time. The output of the time multiplexer as it samples each channel once is called a frame. The number of frames per second is the effective sampling rate, therefore the input signals must obey the sampling theorem (as discussed in Section 3.1.1.1) in order to avoid frequency aliasing when decoded. The dwell time per channel must be long enough for the following stages of equipment to settle on (track) the new channel after switching from the previous one. The transition between channels must be fast enough to reduce crosstalk interference between channels to an acceptable level. The product of number of channels and the dwell time gives the minimum time per frame and therefore sets the maximum frame rate. An additional consideration which may apply is the use of frame synchronizing or identifying pulses, which must be distinguishable from the input channels.

Frequency multiplexing is the process of adding (mixing) the outputs of FM modulators with different carrier frequencies. The input channels can then be separated (decoded) by band pass filtering and FM demodulation. The important characteristics are

the bandwidths around each carrier and the carrier separations in frequency. The dynamic range increases with increasing bandwidth but decreases due to channel cross-talk if the carriers are not adequately separated. Since the total bandwidth available is generally fixed, there is a tradeoff between dynamic range and the number of channels that may be accomodated.

2.2.9.4 Continuous Wave Modulation

Continuous wave modulation is an encoding method where the signal input is used to modify the instantaneous amplitude, frequency, or phase of a high frequency carrier wave. The technique is used whenever the frequency range of the signal must be translated to the limited frequency range of analog transmitters or recorders. The signal recovery is performed by demodulators. The applications of continous modulation are not limited to analog fields - the limitations of telephone and radio transmission as well as magnetic recording media must be addressed in even the most recent telecommunication and digital computing fields.

There are three basic types of continuous modulation: amplitude modulation (AM), frequency modulation (FM), and phase modulation (PM). The encoded carrier signal can be described by the relation:

$$y(t) = A(t) \cos \psi(t) \qquad (2\text{-}19)$$

Amplitude modulation causes the amplitude of the encoded carrier wave to vary with the input signal amplitude. For AM encoding, the carrier frequency is a constant ω and $\psi(t) = \omega t$. The signal $x(t)$ is encoded into the carrier amplitude by

$$A(t) = A_0 + a\, x(t) \qquad (2\text{-}20)$$

where the amplitude of x(t) should not exceed A_0/a, a condition known as overmodulation. The lower bound on the dynamic range provided by AM encoding is set by interference from additive noise or by gain variations. With dynamic range varying between 40 dB (fair) and 20 dB (poor), the use of AM encoding for acoustic and vibration signals is generally avoided.

In FM encoding, the signal dynamic range can be increased because the signal is encoded into the instantaneous frequency of the carrier, and the carrier amplitude is constant

$$y(t) = A \cos \psi(t) \tag{2-21a}$$

$$\psi(t) = \int \omega(t)\, dt \tag{2-21b}$$

$$\omega(t) = \omega_o + \Delta\omega\, x(t) \tag{2-21c}$$

where ω_o is called the carrier center frequency and $\Delta\omega$ is the maximum frequency excursion. The maximum frequency excursion is often expressed as a percentage of the carrier frequency, called the deviation. FM tape recorders typically use deviations of 40%, while much smaller deviations must be used for radio frequency telemetry. The ratio of maximum frequency excursion to the maximum frequency of the input signal is called modulation index. The modulation index is also called the phase excursion since it represents the deviation of the the total carrier phase (in radians). FM systems are capable of encoding low frequency and DC signals.

FM encoding is insensitive to amplitude errors and requires only that the carrier amplitude exceed the noise amplitude. There is a tradeoff between frequency bandwidth and dynamic range in FM encoding which is further discussed in Section 2.4, System Design. The dynamic range available in FM encoding is generally 30 dB (fair) to 50 dB (good), which makes it an acceptable medium

for many acoustic and vibration measurements. Most laboratory tape recorders can operate in an FM mode with the encoders and decoders built into the machine's electronics. FM transmission is used by 300 bit-per-second modems (modulator-demodulators) for digital communication over ordinary telephone lines.

Phase modulation (PM) encoding is similar to FM encoding in that both have constant carrier amplitude and encode the signal in the total phase $\psi(t)$ of the carrier. In PM encoding, the signal controls the instantaneous phase excursion of the carrier.

$$y(t) = A_o \cos [\omega_o t + \theta x(t)] \qquad (2\text{-}22)$$

The maximum phase excursion in PM is independent of signal frequency, while the maximum frequency excursion is proportional to the signal frequency. PM systems are better suited for encoding of high frequency signals than are FM systems. They are used mainly in high speed digital recording. Modems operating at 1200 bits-per-second use phase encoding.

2.2.9.5 Pulse Modulation

In contrast to continuous wave modulation, where the carrier wave is a sinusoid, pulse modulation encoders use a pulse train as the carrier wave. As in the case of multiplexers, pulse modulators sample the signal at certain times (usually at constant intervals) so the signal must not contain frequencies greater than half the sampling rate. The sampling rate of a single channel modulator is the pulse repetition frequency. Pulse modulation may be easily combined with multiplexing to produce a pulse train representing frames of samples from each input channel. This procedure is commonly used when a number of low frequency signals must be tape recorded in synchronization with high frequency channels; the low frequency channels can be combined in a pulse modulator/multiplexer to produce a single

channel which can then be FM modulated or recorded directly along with the high frequency signal channels.

There are three basic types of pulse encoding: pulse amplitude modulation (PAM), pulse time modulation (PTM), and pulse code modulation (PCM). Within PTM there are at least three sub-methods: pulse duration modulation (PDM), which is also known as pulse width modulation (PWM); pulse position modulation (PPM); and pulse frequency modulation (PFM). An additional descriptor of the form of pulse modulation is whether the waveform returns to zero (ground) after each pulse. Those forms that return to zero are called RZ, those which do not are NRZ (non-return-to-zero).

In pulse amplitude modulation (PAM), the signal's value at the sampling time is encoded in the height (amplitude) of the pulse. Different forms of PAM use single polarity or double polarity encoding.

Pulse time modulation (PTM) is the pulse analog of FM encoding. The pulse amplitude remains constant, reducing the influence of noise and gain variations, and the signal's value at the sample time is encoded into time parameters of the pulse. In pulse duration (PDM, also known as pulse width PWM) modulation, the width of the pulse corresponds to the signal amplitude. In pulse frequency modulation, the frequency of pulse repetition is varied around the center frequency just as in FM encoding. Pulse position modulation results in pulses of uniform width, but with the spacing between pulses varied to encode the signal.

Pulse code modulation (PCM) is fundamentally different from other forms of encoding. In addition to sampling the signal at discrete times, as do the other pulse modulations, PCM uses quantization to discretize the signal's information content. Instead of encoding the signal into pulse amplitude or timing, the signal is represented by a digital code (sequence of pulses). Because there can be only a finite number of such

codes, the dynamic range becomes limited by the encoding scheme. The advantage over other forms of encoding is that the decoder need only recognize the sequence of pulses in order to reconstruct the quantized amplitude. Interfering effects such as additive noise, gain variation, and tape speed fluctuation do not influence the decoded signal unless they are so severe that the pulse sequence cannot be recognized.

PCM encoding is especially useful when the test signal dynamic range requirements exceed the capabilities of other encoding schemes. The tradeoff for this increased dynamic range is that multiple pulses are needed to encode each sample, thereby increasing the frequency bandwidth requirements of the encoded signal. If the transmitter or recorder bandwidth is fixed, then PCM encoding requires a lower sampling rate or fewer input channels than the other pulse modulation encoders

Different types of PCM encoding are derived from the form of the pulse code used to represent the signal. One factor is the number of amplitude levels that the encoded pulses may take. The simplest and most popular choice is binary encoding, where only two levels are used - the pulse is either on or off. Systems using three levels are called ternary. The most important factor in any PCM system is the choice of quantization resolution - the number of amplitude levels of the input signal that can be encoded. In binary encoding, the number of bits in the code determines the resolution. A complete discussion of quantization effects is given in Section 3.1.2.

Another factor in PCM designs is the mapping scheme for converting input amplitudes to the output codes. The most common method is linear encoding, where the input amplitude range is divided into equal increments with one code for each level. Other types are logarithmic encoding, which uses small increments for small amplitudes and larger increments at large amplitudes,

and delta modulation, which encodes the signal's change from the previous sample rather than its current amplitude.

Pulse code modulation encoding is used extensively in telemetry systems, where adequate dynamic range may be difficult to achieve with other systems, and also in computer-based analysis systems, where the decoding step is unnecessary since the signal must be quantized to digital form in any case. Because of the importance of such systems in the application of acoustic and vibration data, the most popular devices used to accomplish PCM encoding are described in the following section.

2.2.9.6 Analog-to-Digital Converters

Analog-to-digital converters (ADC's) encode continuously varying signals into digital codes by both sampling (discretizing in time) and quantizing (discretizing in amplitude) the input signal. The ADC itself is usually an integrated circuit chip which performs the quantization function from a voltage or current level into a binary code. The converter is often preceded by a multiplexer for channel selection and a sample-and-hold amplifier that performs the sampling function. The sample-and-hold (also called track-and-hold) device acquires and holds the amplitude of the signal at the sampling time so that the ADC need only deal with a constant input during conversion. In some systems the multiplexer is preceded by a sample-and-hold amplifier for each channel so that all can be sampled at the same time, avoiding interchannel time delays. Complete systems including multiplexers, sample-and-holds, ADC's, timing clocks, and often isolation amplifiers or low-pass filters are known as data acquisition subsystems.

The two most important specifications in choosing an A/D system are resolution and speed. The system must provide enough bits of resolution to provide the dynamic range needed for the measurement (see Section 3.1.2) and it must be capable of

performing its conversions quickly enough to follow the signal variations at the required sampling rate. Dynamic range is set by the maximum allowable signal amplitude (full scale range or FSR) and the minimum resolvable signal amplitude (1 least significant bit or LSB). Accuracy specifications that may be used are monotonicity, gain error, zero error, nonlinearity, noise, and temperature stability. All such amplitude performance errors are combined in the system's absolute accuracy, which may be specified in percent of full scale or in fractions of a least significant bit (1/2 LSB is a typical accuracy specification).

Several different ADC designs are available, with the choice depending on the speed and accuracy tradeoff. Integrating ADC's are used where high accuracy and low drift are essential and conversion speed is secondary, such as in precision digital scales and in voltmeters. A dual-slope integrating converter counts clock pulses as a capacitor is first charged by the input signal and then discharged by a reference signal. The number of pulses counted is then translated into the output code. A faster type of converter which is commonly used for acoustic and vibration measurements operates on the principal of successive approximations. In this type, the input signal is successively compared to binary fractions of the full scale voltage (1/2, 1/4, ...) until its level is known to the resolution of the coding scheme. An advantage of this type of device is that the conversion may be "short-cycled", or reduced in resolution, if faster sampling is necessary. The fastest ADC's, which are needed for video digitizing, operate on some form of flash encoding, where the signal is simultaneously compared to all of the available level codes and the result of the comparisons is translated to the output code. Such devices are more complicated and expensive than successive approximation types.

Other ADC specifications refer to the interface at both ends. If the input signal can share a ground potential with the

measuring system, the input connection is single-ended. In situations where the signal must be measured between two wires that share an unwanted noise signal (common mode signal), the input connection must be differential, with isolation from ground. If the input signal contains negative as well as positive levels, the ADC must be bipolar, with a range from plus to minus full scale, as opposed to unipolar for input signals from zero to positive full scale.

The output coding must match the requirements of the following equipment. With binary coding, a unipolar input range is typically coded from 0000 to 1111, base 2, for example, in a 4-bit system. With a bipolar range, the unipolar codes may be simply shifted down (offset binary) so that minus full scale is 0000 or the most significant bit may be used as a sign bit. A more common bipolar coding that is compatible with the integer representations in most computer systems is called two's complement. In two's complement codes, negative numbers are represented by taking the positive form of the number, changing 1's to 0's and 0's to 1's and adding 1, so that minus full scale is 1000, negative 1 LSB is 1111, positive 1 LSB is 0001, and positive full scale is 0111, all in our 4 bit example. The system output must also be electrically compatible with the interface bus or the next device. Pulse levels may be specified by the type of circuitry (TTL, CMOS, etc.). The format (serial bit-by-bit or parallel all bits at once) and timing control signals are also important. Most data acquisition subsystems are available in versions for popular microcomputer buses and for industry standard instrumentation buses.

2.2.10 Cabling Requirements

As discussed in Section 2.2.3 the capacitance of the cable between the transducer and the input stage of the signal-conditioning preamplfier can affect the voltage sensitivity of the transducer. Thus, general setup is illustrated in Figure 2.18 involving the cable capacitance C_c and a capacitively coupled preamplifier with input resistance R_p and input capacitance C_p. The Norton equivalent circuit of the transducer assumes it is a current source with current $i_T = E_o/Z_T$, where E_o is the transducer open circuit voltage and Z_T the transducer impedance in parallel with the source.

The transmission of the preamplifier voltage signal E_p to the transducer open circuit voltage E_o is

$$E_p/E_o = \frac{1/Z_T}{1/Z_T + j\omega(C_c + C_p) + 1/R_p} \quad . \tag{2-23}$$

If the input resistance is sufficiently large, then

$$E_p/E_o = \frac{1}{1 + j\omega(C_c + C_p)Z_T} \quad . \tag{2-24}$$

If the transducer is a capacitive device so that Z_T is capacitive ($Z_T = 1/(j\omega C_T)$) ,

then

$$E_p/E_o = \frac{C_T}{C_T + C_c + C_p} \approx \frac{C_T}{C_T + C_c} \tag{2-25}$$

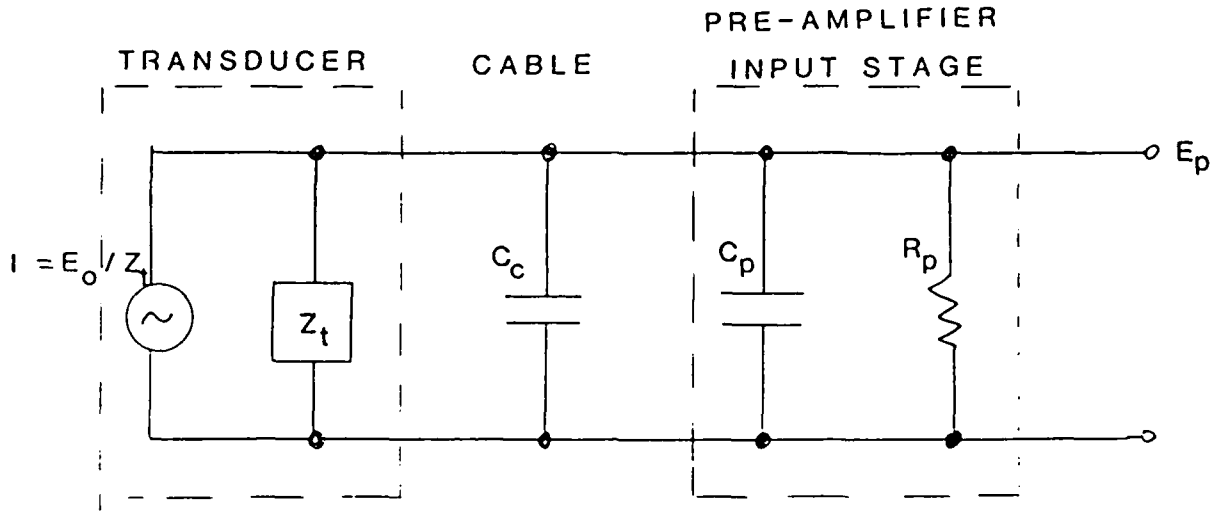


Fig. 2.18 General Circuit for Signal Conditioning

since the preamplifier input capacitance C_p is usually much less than cable capacitance. Thus, the voltage sensivity of a capacitive-type transducer is reduced by cable capacitance. Typical cable capacitane is 30 pF per foot of cable. In such a the calibration of the measurement system can become a function of cable length. In addition to reducing the voltage sensitivity the cable can also introduce noise. This is due to the high source impedance of the capacitive-type transducer, and the electromagnetic fields near the signal conditioning amplifiers. For this reason the cable length should be kept as short as possible and should be coaxial or shielded twisted pair cable with properly shielded connectors.

Cable vibration also introduces noise. The center conductor and shield of a coaxial cable are separated by a non-conducting dielectric material. Vibration of the cable can cause the shield and the dielectric to separate by a small amount, which results in the generation of a triboelectric charge. The flow of this charge on the shield through the high input impedance of the signal conditioning amplifier produces noise.

Noise due to cable vibration is generally a problem only below 200 Hz. It is typically increased by cable length, tension or stress in the cable, and vibration level. To avoid this source of noise it is recommended that the cable be tied down to avoid long unsupported lengths; that tension, small radius turns, and twisting of the cable be avoided; and that high quality cable be used. Several transducer manufacturers provide low noise cable. In addition to being treated to reduce noise due to cable vibration, these low noise cables are typically lightweight and flexible so that they do not affect the vibration being measured. They also are provided with connectors to attach to the transducer and to standard signal conditioning amplifiers. Finally, they are available to withstand a variety of temperature

conditions. Although these cables are more expensive than standard cable, they should be used whenever possible.

It is difficult to predict whether cable noise will be a problem. Although ISA has a cable noise test, the test conditions do not simulate well the conditions during an actual test. Therefore, a recommended procedure is to use a "dummy" transducer having the same capacitance as an actual transducer, but with no sensitivity to vibration. This dummy transducer is mounted at a convenient location within the test object and treated as one data channel having a cable, signal conditioning preamplifier, etc., that are identical to the other active channels. This noise monitoring channel will help to determine if cable noise is a problem when the data are processed.

Cable selection depends also on environmental factors. The broadest temperature range Teflon coated cables operate from -100°C to 200°C. Cables using flexible metal conduits with high temperature resistance operate in the range from -196°C to 480°C. These type of cables are used in very high level, sustained vibration experiments, in which considerable heat can be generated in the cable.

A key concern with miniature and subminiature accelerometers, in particular, is to use the smallest, lightest and most flexible cables available. The weight of cables can otherwise exceed the accelerator weight.

2.2.11 Transmitters

In many vibration and acoustic data measurement problems it is impossible to record the large amount of data that is to be collected on-site. A common example occurs for missile flight testing. Because of weight and space limitations it is usually impossible to include the recording device within the missile. In addition, even when space is available, recovery of the recording device can be impossible. Radio telemetry is often used for such tests.

Radio telemetry may also be used to collect data for rotating machinery. In this case the recording device cannot be located on the rotating part of the machine. Slip-rings and other similar devices may be used to transmit the electrical signal from the transducer to the recording device. However, use of radio telemetry may be a more accurate approach, particularly for collection of high frequency data.

The most common form of radio telemetry for missile and aircraft flight testing is through use of standard FM/FM telemetry techniques. In these techniques the output signals from a number of transducers are conditioned and used as inputs to subcarrier oscillators (SCO's), which serve as the first frequency modulation (FM) stage of the system. The SCO outputs are then summed and used to modulate a radio frequency (RF) oscillator, which serves as the second FM stage of the system. The use of the FM/FM system allows several channels of data to be collected with a single RF transmitter. Thus, although each channel of data could be used directly to modulate a separate RF oscillator, the cost of this simple FM approach greatly exceeds that of the FM/FM approach when many channels of data are to be collected.

To meet IRIG recommendations for data accuracy, the frequency separation of the SCO frequencies must be significantly greater than the bandwidth of the data. Thus, the greater the bandwidth of the data the fewer the number of subcarriers (channels of data) per RF transmitter. A comprehensive flight test may require several transmitters.

Due to limitations of on-board space it is often necessary to limit the number of data channels, the bandwidth of each channel, and the accuracy or dynamic range of each channel. Often this limitation compromises the over mission. For this reason much attention has been given to use of on-board data processors and pulse code modulation (PCM) techniques.

Using pulse code modulation (PCM) techniques the data signals are first sampled and converted to digital signals. These digital signals are then quantified, coded and transmitted by a single RF transmitter. An advantage of the PCM technique over the FM/FM technique is that multiple channels of data can be encoded with higher accuracy using digital techniques and transmitted without the interference between adjacent channels that occurs in FM/FM systems.

For data channels having a very wide frequency bandwidth the sampling requirements for the PCM technique result in very high sampling rates and a high bit/sec transmission rate. As for the FM/FM system the number of high frequency channels that can be transmitted over the single RF link is very limited. However, the data accuracy is improved.

A simple example will illustrate the influence of high frequency vibration and acoustic data channels on transmitter bandwidth. Assume that the test plan calls for collection of 100 channels of low frequency data and 10 channels of vibration and acoustic data. The desired sampling rate of the low frequency channels is 10 samples per second. The desired bandwidth of the vibration and acoustic channels is 5,000 Hz. To prevent aliasing

allowing a 20% increase in bit rate to account for synchronization and identification of samples results in a 1.5 x 10^6 bits/sec rate which is excessive for a typical transmitter. Note that the bit rate required by the vibration and acoustic channels is more than 100 times the rate required by the low frequency channels. The use of an FM/FM system for the above example would allow a reduction in overall bandwidth. However, the dynamic range of each channel would be limited to less than 40 dB.

2.3 RECORDING DEVICES

2.3.1 General Requirements

In many acoustic and vibration measurements it is impossible to complete the required data analysis in real time. In addition, it is usually desirable to have a permanent record of the measured data so that measurements do not have to be repeated should additional processing be needed at a later time. The general requirements for a permanent recording device include:

a) preservation of data quality
b) size and weight
c) cost
d) environmental effects, and
e) compatability of the recording format with accepted standards

The use of transient recorders in acoustic and vibration measurements is discussed in Section 2.3.3.

2.3.2 Permanent Recording Devices

Three general types of permanent recording devices are considered. These are analog magnetic tape, graphical hard copy, and digital magnetic tape/or disks. In most cases a magnetic tape device is used to obtain a permanent record of the data. A graphical hard copy is then obtained at a later time.

2.3.2.1 Analog Magnetic Tape Recorders

The analog magnetic tape recorder is a device to record an analog voltage signal. This type of recorder is commonly used because of its ability to store a large amount of data for a long period of time and then to reproduce the data with great accuracy.

The frequency response of the analog recorder depends on whether the recording method is direct or FM (frequency modulated). In both cases the bandwidth is limited by tape speed. Figure 2.19 lists the bandwidth for several types of direct and FM recording at different tape speeds. However, at a given tape speed the bandwidth of the direct recording extends to a much higher frequency than that of an FM recording. On the other hand the FM recording bandwidth can extend down to DC while that of the direct recording is limited by the playback heads of the recorder to approximately 20 Hz.

For many vibration and acoustic measurements the direct method of tape recording is used because of its inherently lower cost and higher signal-to-noise ratio over the mid-frequency range. However, FM recording is also commonly used because of its inherently better phase linearity at low frequencies and its insensitivity to tape magnetization characteristics. In the direct method of recording a magnetic flux is produced at the recording head that is proportional to the input voltage to the recorder. This flux induces a permanent magnetization on the

DIRECT

Dynamic Characteristics, 600 kHz System

Tape Speed (ips)	Bandwidth ±3 dB (kHz)	Signal rms/Noise rms (dB)[1]
120	0.3 - 600	41
60	0.3 - 300	41
30	0.15 - 150	40
15	0.1 - 75	39
7.5	0.1 - 37.5	38
3.75	0.1 - 18.7	38
1.87	0.1 - 9.3	37
0.937	0.1 - 4.7	35

Dynamic Characteristics, 2 MHz System

Tape Speed (ips)	Bandwidth ±3 dB (kHz)	Signal rms/Noise rms (dB)[1]
120	0.4 - 2000	26
60	0.4 - 1000	27
30	0.4 - 500	27
15	0.4 - 250	26
7.5	0.4 - 125	25
3.75	0.4 - 62.5	24
1.87	0.4 - 31.25	21
0.937	0.4 - 15.00	19

[1] Measured at the output of a bandpass filter having 18 dB/octave attenuation beyond bandwidth limits and using recommended tapes

FM

Dynamic Characteristics, ±40% FM (IRIG Intermediateband)

Tape Speed (ips)	Center Freq (kHz)	Data Bandwidth (kHz within 1 dB)	SNR rms (dB)	(dB)[2]
120	216	0 - 40	51	47
60	108	0 - 20	51	47
30	54	0 - 10	51	46
15	27	0 - 5	50	46
7.5	13.5	0 - 2.5	49	46
3.75	6.75	0 - 1.250	48	45
1.87	3.38	0 - 0.625	47	44
0.937	1.68	0 - 0.313	46	40

Dynamic Characteristics, ±40% FM, IRIG Wideband Group I

Tape Speed (ips)	Center Freq (kHz)	Data Bandwidth (kHz within 1 dB)	SNR rms (dB)	(dB)[2]
120	432	0 - 80	50	46
60	216	0 - 40	50	46
30	108	0 - 20	50	44
15	54	0 - 10	49	44
7.5	27	0 - 5	49	44
3.75	13.5	0 - 2.5	48	43
1.87	6.75	0 - 1.250	47	42
0.937	3.38	0 - 0.625	44	40

[2] When operated in a wideband instrument

Dynamic Characteristics, ±30% FM, IRIG Wideband Group II

Tape Speed (ips)	Center Freq (kHz)	Data Bandwidth (kHz +1, -2 dB)[3]	Data Bandwidth (kHz +1, -3 dB)[3]	SNR rms (dB)
120	900	400	500	34
60	450	200	250	34
30	225	100	125	34
15	112.5	50	62.5	34
7.5	56.25	25	31.2	33
3.75	28.125	12.5	15.16	31
1.87	14.062	6.25	7.8	30

[3] IRIG reference frequency 1 kHz

Data Courtesy of Honeywell Test Instrument Division for Model 101 Magnetic Tape Recorder

Fig. 2.19 Bandwidth and Noise For Various Tape Speeds

tape. Thus, as the tape is transported past the record head a permanent record of the input voltage time history is obtained.

The induced magnetism on the tape is not linearly proportional to the applied flux because of hysteresis. Therefore, a high frequency bias is added. This bias effectively results in linearity between the induced magnetism and the applied flux. Typically the bias frequency is 2 to 5 times higher than the upper limit of the frequency bandwidth, and, therefore, does not affect the accuracy of the recorded data. The level of the applied bias does have an effect on the frequency response and distortion of the recorder and should be set to produce the desired results for the particular tape being used. It is important to recognize that changes in bias level and/or changes in tape will alter the frequency response and can introduce distortion into the recording. The manufacturers of the recording devices often recommend particular types of tape to be used. On most high quality recorders the bias level can be changed for different types of tape.

In general a bias level that is too low for the type of tape being used will result in increased distortion and an increased response at high frequencies. A level that is too high can also introduce distortion by saturating the recording head and will reduce response at high frequencies.

The reproduce head of a magnetic tape recorder generates a voltage that depends on the rate of change of flux. This results in a frequency response characteristic that increases 20 dB/decade. Typically, this rising frequency response is corrected to a flat response by equalization circuits in the input and output amplifiers of the recorder. On many recorders the equalization can be changed to produce an overall frequency response that has desired characteristics, i.e. maximum flat frequency response or maximum bandwidth. As in the case of the bias applied at the record head, the equalization circuits must

be properly matched to the type of tape used. It must also be properly set for the tape speed being used.

The characteristics of the playback head make it impossible using the direct recording method to record DC or low frequency data. Typically, the lower frequency limit is in the range 20 to 50 Hz and may be higher for some recorders. However, it is possible on some recorders to extend low frequency response to a few Hz by recording at a slow speed and playing back at a higher speed. This approach overcomes the limitation of the reproduce head to respond to these low frequencies. A better approach, however, is to use an FM recording method.

The frequency bandwidth of a direct recorder extends to very high frequencies and depends directly on tape speed and width of the recording head gap.

A direct recording of a pure tone produces a fluctuating magnetization on the tape with a wavelength given by

$$\lambda = \frac{s}{f} \tag{2-26}$$

where s is the tape speed, and f is the frequency of the tone. At a sufficiently high frequency the half-wavelength becomes equal to the gap width. This frequency defines an upper bound to the frequency response of the recorder. If we set $\lambda = w$, where w is the gap width, in the equation above an upper limit to the frequency response is obtained

$$f_{\ell} = \frac{s}{2w} \tag{2-27}$$

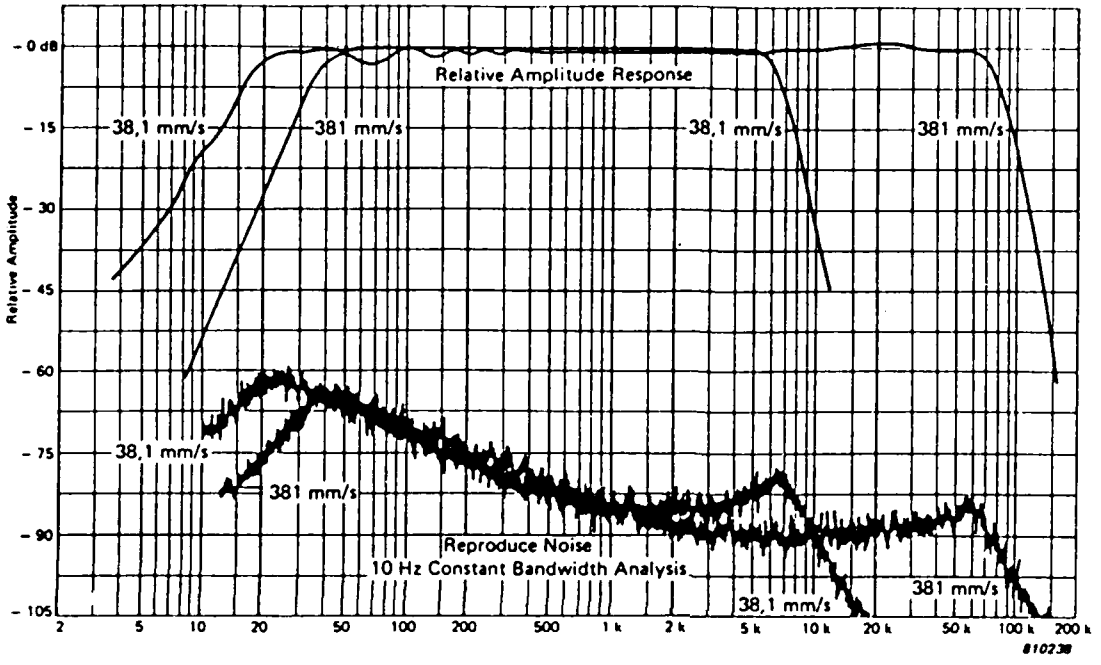
For example, at a tape speed of 120 ips with a gap width of 10^{-4} inches, the upper frequency limit is found to be 600 kHz. An actual recorder may not reach this theoretical limit to the bandwidth, but the role of tape speed and magnetic head gap width is clear.

Figure 2.19 shows the direct record bandwidths as a function of tape speed for three different designs. Note the proportional decrease in bandwidth with tape speed. Also note that at a given tape speed a higher bandwidth is obtained at the expense of reduced signal to noise.

Although a wide frequency bandwidth is required for some applications, the need in vibration and acoustic measurements is much more limited. An FM recorder can give the required bandwidth and exhibit other favorable characteristics. However, because of the inherently lower cost, the direct recording method is used in lower cost recording systems. It also can provide a somewhat greater signal-to-noise ratio than FM recording.

The frequency response of a typical high quality direct recording device for vibration and acoustic measurements is shown in Figure 2.20 with the power spectrum of the noise. At a tape speed of 381 mm/s (15 ips) the bandwidth extends from 40 Hz to 60 kHz. At the lower tape speed of 3.8.1 mm/s (1.5 ips) the bandwidth extends from 20 Hz to 6 kHz. The RMS signal-to-noise ratio is 42 dB at both speeds. However, the power spectrum of the noise is such that higher signal-to-noise ratios can be obtained by filtering out the high and low portions of the bandwidth. A visual comparison of the noise spectra for the direct and FM recording shows that at a tape speed of 381 mm/s the noise for the direct recording over the mid-frequency range from 1 to 20 kHz is up to 10 dB lower with the largest difference at high frequencies. Finally, note that the noise spectra for the direct recording is not greatly influenced by tape speed. This is not the case for the FM recording.

In the FM recording method the input voltage signal is used to frequency modulate a high frequency carrier signal which is then recorded on the tape. Since only the frequency of the signal on the tape is important and since harmonic distortion of the signal can be filtered out, the problems of hysteresis in the



Data courtesy of Bruel & Kjaer for Type 7005 Instrumentation Recorder

Fig. 2-20 Frequency Response and Noise for Direct Recording

recording process and saturation of the heads, which limit the accuracy of the direct recording, are not a factor. Although the quality of the electronics required for the modulation and demodulation has a direct influence on the quality of the recorded data the FM recording is reasonably insensitive to the magnetic recording process. The only factor of critical importance is variations in tape speed due to flutter. Since these are detected as a change in frequency of the recorded data, tape flutter results in low frequency noise on the FM recording. Generally the FM recorder requires a more precise tape transport mechanism than does the direct recorder. Typically a phase-lock, servo-feedback system is used to regulate the tape speed. In addition, most FM recorders use a capstan drive, which can be accurately controlled because of its low mass.

A typical commercial recorder uses IRIG standard 14-channel format on 1-inch tape with 10 1/2-inch reels. The circuitry allows choice of direct and FM electronics with voice-over recording capability, so that the operation can make an audible record of test conditions. Units contain built-in test equipment for diagnostic purposes, and internal shock mounting is available for air-borne and other rugged environments.

2.3.2.2 Graphical Hard Copy

Graphical hard copy devices provide a means to obtain a visual display of the data. The primary use of such devices in vibration and acoustic measurements is to provide a "quick-look" capability in reviewing recorded data. Although it may be feasible to use a hard copy device as the primary device for data storage, this is not done because of the limited frequency bandwidth of hard copy devices and the difficulty in carrying out further data analysis. Therefore, the hard copy device should be used only to support other types of data storage, such as mag-

netic tape, or as a means to display processed data results. Indeed, a picture can be worth a thousand words (or numbers).

Two basic types of hard copy recording devices exist - strip chart and X-Y recorders. High performance strip chart recorders with the capability of recording over a wide frequency bandwidth are commonly referred to as oscillographic recorders. A strip chart recorder that records the RMS level of fluctuating signals is called a Graphic Level Recorder and is used for acoustic and high frequency vibration data.

The various types of strip chart recorders produce a continuous record of the data on a strip of paper. The lower peformance and lower cost units use a stylus writing technique with either an ink pen or a hot-wire pen with thermally sensitive paper. The performance of these stylus units is limited to an upper frequency of approximately 150 Hz, with many units having a frequency response up to only 1 Hz.

The limited frequency response of the stylus recorders preclude their use for the direct recording of most vibration and acoustic data. They are used, however, in graphic level recorders, which will be discussed later in this section.

Oscillographic recorders using fiber optic writing techniques can have a frequency bandwidth extending up to 5000 Hz. Recorders of this type can be used to directly record vibration and acoustic data from many measurement programs. The visual records obtained from such a recorder provide a useful means to review the quality of recorded vibration and acoustic data and to select portions of the data for further processing.

In selecting an oscillographic recorder, factors influencing the usefulness of the chart recording should be carefully evaluated. Since the chart is only used for a visual review of the data, the accuracy of the recording may not be as important as other factors such as multi-channel recording, choice of chart speed, record width, time-interval marking, and cost.

A typical commercial fiber optic oscillographic recorder has 18 to 32 input channels with DC to 5 kHz frequency response. Recording speeds vary from 0.1 to 120 inch/sec with over 40 discrete steps, and remote drive and speed control are built-in. Such an oscillographic recorder has no moving elements, and thus all inertial-related overshoot phenomena are eliminated.

The graphic level recorder is an alternative to the oscillographic chart recorder. It is typically used to display acoustic data, although it can also be used for vibration data. The level recorder provides a chart record of the RMS value of a fluctuating signal. The averaging time for the RMS detector is adjustable so that short-duration transient events in the data can be identified. However, because of the limited frequency response of the stylus writing mechanism, the averaging time is usually too long to allow identification of impulsive events such as drop-outs in telemetered data or due to intermittent cable shorts.

The X-Y recorder serves a major role in preparing a graphical record of processed data. Both analog and digital recorders are available in a variety of sizes, frequency bandwidths, and costs. Digital X-Y recorders are often referred to as plotters. Most X-Y recorders and plotters use a stylus writing mechanism with ink pens in a variety of line widths and colors. The typical plotter can produce report quality graphs on 8 1/2 x 11" paper. Many plotters also have the capability of generating block letters for labelling graphs.

The dynamic performance of the X-Y recorder is specified in terms of a slewing speed (maximum pen speed) and a maximum pen acceleration. Both parameters enter in determining the recorder performance with the slewing speed being more important for recording large changes in signal amplitude and the pen acceleration being more important for recording low amplitude rapidly fluctuating signals.

2.3.2.3 Digital Recording Devices

Digital recording devices use a magnetic tape or disk as the recording media. Paper tape can also be used, but is obsolete because of its very low recording density - bits per inch.

Digital tape is the most common means to obtain permanent storage of digital data in a computer system. Most computer magnetic tape systems conform to industry standards. For this reason 1/2" magnetic tape is the most common form for both data and software storage. Large signal processing software programs are usually obtained in the form of an industry-standard 1/2" magnetic tape.

The recording density for a computer magnetic tape system is typically 800 or 1600 bpi (bits per inch) with 9 tracks of data on a 1/2" tape. A 10 1/2" tape reel with 2400 ft of 1/2" tape can store over 45 million characters of data (45 Mbytes). The data transfer rate from computer memory to tape is dictated by tape speed and the speed at which the tape controller of the computer can transfer data from memory. In most cases the controller will buffer the data and use a direct memory access channel so that the transfer rate of the overall system is limited by tape speed. With a recording density of 1600 bpi and a tape speed of 75 ips the transfer rate is limited to 120,000 bytes per second. This rate can be increased by increasing tape speed, but in all cases the transfer rate from computer memory to tape is well below the rate that can be achieved from a computer to a magnetic disk.

Recently a number of manufacturers have introduced digital tape recording systems for high quality analog data recording. The primary advantage of the digital recording technique for vibration and acoustic data is dynamic range. A 12 bit conversion of an analog signal provides 72 dB of dynamic range. Direct and FM recorded range of only 30 to 50 dB. In addition

the digital recording system provides higher accuracy and long-term stability. Although the cost of the digital recording system is much greater at the present time, there is no doubt that this type of data recorder will replace many analog recording systems because of its greater dynamic range.

The primary limitation of digital data recording is its limited frequency bandwidth. The digital system described above with a data recording rate of 120,000 bytes per second could record analog data from a 12 bit A to D converter at a rate of 60,000 points per second. Assuming that the sampling rate is twice the highest frequency of interest, the bandwidth would extend only up to 30 kHz. An FM recorder using 1/2" tape at a similar tape speed can achieve this frequency bandwidth for 7 channels of data instead of just one. However, the dynamic range for each channel is only 50 dB.

High density digital recorders are available which achieve data transfer rates up to 20 Mbytes/sec on a 1" tape at 240 ips. This rate allows 14 channels of data to be recorded, each with a bandwidth up to 350 kHz. This level of performance can be achieved using a direct or FM Wideband Group II analog recording, but with a dynamic range that is approximately one-half that of the high density digital system.

2.3.3 Transient Recorders

The transient recorder is a device to make a temporary record of large amounts of data, so that data can be subsequently analyzed and/or stored permanently. Transient recorders digitally sample the incoming conditioned inputs and store the digitized inputs in memory. Capabilities allow command control and data transfer to and from host computers via standard buses (including GPIB and STD). Various permanent storage options include oscillographic recorder output and storage on floppy or hard disks.

For a given number of input channels and sample rate, the main parameters for assessing transient record performance are the resolution of the analog/digital converter (ADC), the input noise spectrum and the aperture uncertainty. The resolution of the ADC is typically 8 to 12 bits. This corresponds to 1 part in 256 resolution for an 8 bit ADC, and 1 part in 4096 resolution for a 12 bit ADC.

The input noise spectrum effectively limits the low frequency resolution. If the inputs are shorted, and the measured internal noise signal occupies 4 of the 4096 levels in a 12 bit ADC, then any applied signal has the low level resolution of 4 parts in 4096 or 1 part in 1024. Thus, this noise level has reduced the resolution from 12 to 10 bits ($2^{10} = 1024$).

The high frequency resolution is degraded by aperture uncertainty. The aperture uncertainty is the variation from the nominal digitizing rate. A commercial transient analyzer with sample rates of 1 MHz has an aperture uncertainty of 0.1 ns. The aperture uncertainty reduces the resolution of the fastest changing part of the signal.

As a practical example, assume that sampling waveform V(t), being sampled, is a sinusoid of frequency f. The maximum rate of change of this signal $(dV/dt)_{max}$ is:

$$(dV/dt)_{max} = 2\pi f V_o \qquad (2\text{-}28)$$

where V_o is the full scale voltage of the ADC. The aperture uncertainty Δt leads to a corresponding voltage uncertainty ΔV as follows:

$$\Delta V = \Delta t \ (dV/dt)_{max} \qquad (2\text{-}29)$$

or

$$\Delta V/V_o = 2\pi f \Delta t \qquad (2\text{-}30)$$

Note that the voltage uncertainty increases with sampling frequency (or signal frequency). If f=1MHz and Δt=0.15 ns, then

$$\Delta V/V_o = 9.42 \times 10^{-4} \qquad (2\text{-}31)$$

corresponding to 4 parts in 4096 for a 12 bit ADC. Thus the high frequency resolution is reduced to that of a 10 bit ADC.

The total length of the sampled waveform depends on the total memory available for storage. Commercial units with 512k points in transient storage are available.

The transient recorders of today may be viewed as the front end of a data acquisition system controlled by a host computer. Transient recorders provide the capability to acquire large amounts of data and store these in memory. The host computer initiates and stops data acquisition and performs the analysis. Various types of output and permanent storage options include CRT displays, oscillographic recorder output, digital tape storage of data and results, and the alternative possibility of storage on hard disk floppy disk.

2.4 SYSTEM DESIGN

Previous sections of this chapter have described the individual components making up a measurement system. It is important to understand the performance of each component in designing an overall instrumentation system. It is equally important to understand how the different components interact and how to specify tests that determine end-to-end measurement accuracy and calibration.

The different methods for applying measured data to vibration and acoustic problems require different degrees of measurement system accuracy. Thus, in designing the measurement system it is necessary to understand the purposes for which the data are to be used. A detailed discussion of the requirements for each of the methods presented in this report is presented in Chapter 4. This section describes the procedures to be followed in designing the overall system. We assume that the reader has a basic understanding of the performance features of each individual component.

2.4.1 General Requirements

The general requirements for a measurement system can be divided into four major categories:

accuracy
cost
availability
compatibility with weight, size,
and electrical power limitations

Often these requirements conflict and compromises must be made. Therefore, the system designer must be in a position to revise the general requirements and to select alternate system components when necessary.

As a general rule it is advisable to avoid use of specially built electronic components and sensors. The length of time required to test the component operation is often beyond the available time for measurement system design. In addition, the specially built component can often have an unexpected fault which compromises the performance of the entire system.

It is also a good rule to avoid components that are useful only for a single vibration and acoustic survey. Again unexpected errors are more common for such special purpose components. In addition, errors in operation of the measurement system by field engineers are more probable if they are not familiar with the equipment. It is better to use a familiar set of system components even though the overall system accuracy is less than could be achieved with other components.

The Sections to follow will focus on measurement system accuracy. This does not imply that the other factors listed above are not important. Rather it is an indication that these factors are unique to each measurement group and it is difficult to discuss them in general terms.

2.4.2 System Design Procedures

A major factor in the design of a measurement system is system accuracy. In the design and use of a system for a particular measurement program it is necessary to consider the many sources of measurement error. These errors can cause poor data quality and in the extreme total loss of data.

A complete analysis of system errors includes the following topics:

Calibration Errors

Errors due to unknown or inaccurate transducer sensitivites;

Errors due to failure to log changes in amplifier gain between system calibration and data measurements;

Errors due to inaccurate calibration signals.

Component Errors

Errors associated with poor signal to noise ratio for a particular component;

Errors associated with insufficient frequency response for a component;

Errors due to overload - exceeding the maximum input voltage for a particular component;

Errors due to failure to correctly log component gain settings and other parameter settings;

Errors due to drift in component gain and/or frequency response due to temperature changes or other environmental factors.

Interaction Errors

Errors due to improper impedance match between components.

Cable Errors

Errors due to improper connection or mislabled cables;

Errors due to cable vibration;

Errors due to broken cables or cables with intermittent shorts or open circuits.

Interference Errors

Errors due to pickup of unwanted signals from electromagnetic fields;

Errors due to ground loops.

Operational Errors

Errors due to improper setting of amplifier gains;

Errors due to failure to record gains and other component settings;

Errors due to failure to follow calibration procedures;

Errors due to failure to properly annotate data recordings or other records.

These sources of error will be discussed in the following sections and procedures for minimizing the errors will be outlined.

2.4.2.1 Accuracy Requirements

The accuracy of a measurement system can be described in terms of a number of key descriptors. These include:

Number of Data Channels

recorder limits, high and low bandwidth signals, annotation signals, multiplexing, encoding.

Frequency Bandwidth

different for vibration/acoustic and for annotation channels, DC more difficult, high frequencies may cause aliasing, 3 dB down points, deviation within the band, amplitude and phase.

Dynamic Range

definition, upper limit linearity, lower limit noise, recorder limitations, peak factor, signal to-noise, auto-ranging amps, level estimation uncertainty, frequency range

Noise

broadband and single frequencies, bandwidth, signal to noise, biased and unbiased addition.

Linearity

definition, harmonic distortion, clipping.

Interference

interchannel, pickup

The first step in the measurement system design is to determine the required measurement accuracy. This step is part of the test design and is described in Chapter 4. For consistency Chapter 4 presents accuracy requirements for each of the different methods for applying measured data to vibration and acoustic problems that are presented in later chapters of this report. These requirements are divided into the six key descriptors listed above. Procedures for designing a system to meet these requirements are presented below.

2.4.2.2 Number of Data Channels

The first step in designing a measurement system is to determine the total number of data channels that will be required. Many measurement projects require multiple data channels. The system must accomodate the output signals from several vibration and acoustic transducers as well as the more slowly varying signals describing operating conditions such as temperature, altitude, etc. and signals containing data annotation such as time of day, amplifier gain settings, etc.

In most cases a permanent record of the data is needed. The available recording device often sets a limit on the number of channels that can be recorded. Thus, the requirements of the measurement survey must be combined with the capability of the data storage component. If magnetic tape is to be used the number of recorded channels is typically 14. The vibration and acoustic data channels require a much greater frequency bandwidth than the more slowly varying operating condition and annotation channels. Therefore, it is common to record many of these slowly varying signals on a single channel. The usual technique is to use a PCM (pulse-code-modulation) system. When available, an edge track on the tape is often used for a voice annotation. It is also common to use one channel for a time-of-day encoder. Thus, a well designed tape recording system will use 2 of the 14 channels for recording data annotation information such as amplifier gain settings and time of day so that 12 channels are available for recording data.

A single one inch tape can be set up to record more than 14 channels, but with some loss of dynamic range per channel. The loss of dynamic range is usually too great to be acceptable. Thus, if more than 12 channels of data are required, it will be necessary to select one of the following options: use of multiple recorders or use multiplexing of data channels so that more than one data channel can be placed on a single tape channel.

A time multiplexer can be used to place several dynamic signals from vibration and acoustic transducers on a single data channel. However, some data will be lost while the multiplexer is changing channels and many of the more advanced methods for applying measured data to vibration and acoustic problems cannot be used since they require correlation analysis between the signals from different transducers.

Although analog magnetic tape recording is most commonly used for field measurements, recent advances make digital tape recording worth consideration. The advantage of digital techniques is that greater data accuracy can be achieved. However, the number of available data channels cannot be increased without loss of frequency bandwidth. If a digital magnetic disk is to be used, the number of data channels to be used is limited by the total bit rate. Typically, a rate of 1 to 2 Mbits/sec can be achieved using a high speed magnetic disk. Assuming sampling rates of 2.5 times the upper frequency limit of each data channel this bit rate would allow up to 10 to 20 channels of data with a frequency bandwidth of 4,000 Hz and 10 bit data words. However, a 50 Mbyte disk would allow collection of only 2.5 minutes of data. This method of data storage is not suitable for field measurements and FM or digital tape storage methods are preferred.

2.4.2.3 Frequency Bandwidth

Measurement of vibration and acoustic signals requires that the measurement system have a sufficient bandwidth to encompass the entire range of frequencies to be measured. Many sources of vibration such as flow turbulence, aerodynamic noise, and mechanical impact and shock produce signals with a frequency content extending up to very high frequencies. At low frequencies vibration at the fundamental resonance frequencies of a test object or structure can be high. Thus, the frequency range for measurement of vibration and acoustic data covers several decades.

In some cases the higher frequencies are not of interest. Then the measurement system must filter out these high frequencies so that the data accuracy over the frequency range of interest is not compromised. In many cases the high frequency

output of vibration or acoustic transducers causes the input to an amplifier or recording device to exceed its upper limit. The high frequency output can also cause aliasing errors in digital sampling procedures.

The lower frequency limit of the measurement system must be low enough to encompass the lowest frequency of interest. Low frequency vibration can be a cause of equipment failure. However, in many measurement programs, the low frequency vibration relates more to the rigid body motion of the test object and is not of interest. In these cases, filters and/or AC amplifiers are used to eliminate the low frequencies so that they do not interfere with measurement accuracy.

Typically, the frequency bandwidth requirements for vibration and acoustic data are from 20 Hz to 20,000 Hz. However, for many programs in which data is to be collected to develop or verify equipment Qualification Test requirements the upper frequency limit is set at 4,000 Hz. On the other hand, many vibration sources including pyrotechnic devices, high velocity impacts, and aerodynamic shock oscillations result in high vibration levels up to and beyond 50,000 Hz.

The frequency bandwidth required for data annotation channels, which may have information on operating conditions, amplifier gains, and time-of-day, is significantly less than that required for an acoustic channel. Typically, the upper frequency limit can be set to 100 Hz. The lower frequency limit must usually be DC, if analog signals are to be directly recorded. This is not a problem if FM recording is used. The use of a wide bandwidth vibration/acoustic data channel for a single low frequency annotation signal is very inefficient. Therefore, it is common to combine several low frequency annotation data channels onto a single wide-bandwidth channel. The typical approach is to use a PCM system to encode these channels. The frequency bandwidth required for the encoded bit stream is approximately twice the bit rate.

The frequency response of the measurement system can be defined by the amplitude and phase response as a function of frequency. The ideal system would provide a uniform response within the desired frequency bandwidth and a zero or negligible response outside of the band. The actual system shows some variation of response within the frequency bandwidth and a rapid fall-off of response outside of the band. A variety of methods exist for determining the bandwidth. Typically, the bandwidth is determined by the frequencies at which the amplitude response is 3 dB below the average response within the band. However, several other methods exist.

The frequency response of the measurement system, which consists of several components, is determined by adding the amplitude responses in dB of each of the components. Thus, a system consisting of a vibration transducer, signal conditioning amplifier, amplifier, and tape recorder, each having a amplitude response that is -3 dB at 5,000 Hz, will have a system amplitude response that is 4 x (-3) = -12 dB at 5,000 Hz. It follows that the design of a system with an upper frequency limit of 5,000 Hz requires use of components each having a upper frequency limit greater than 5,000 Hz. System design must include the calculation of system frequency response. This step can be done quite easily if the frequency response for each component is available. If this information is not available it can be determined by measurement using a sine-sweep oscillator to generate an input signal.

The phase response of the system is important if the measurement must preserve the exact time histories of the vibration or acoustic signals. However, if the data is to be used only to obtain power spectral densities the phase response is not a major concern. In addition, for many of the correlation techniques the exact phase response for the different channels is not important as long as it is the same for all channels.

If the phase response of a system is a linear function of frequency, the time history of the signal is preserved but delayed in time. This may not reduce the accuracy of the data. Therefore, a linear phase response is often adequate. Typically, the frequency bandwidth over which the phase response is linear is more limited than the bandwidth for the amplitude response.

The phase response of the system is the sum of the degrees of phase lag for each component. As in the case of the amplitude response a system design meeting a requirement for phase response at an upper frequency limit of 5,000 Hz requires components with a upper frequency limit above this frequency.

2.4.2.4 Dynamic Range

The dynamic range of the measurement system defines the range of signal amplitudes that can be accurately measured. At the lower limit the dynamic range is set by the noise (noise floor) of the system. Signals with amplitudes less than the noise have poor signal-to-noise ratios and cannot be accurately determined. At the upper limit the dynamic range is set by the linearity of the system. Signals with a sufficiently high amplitude will be distorted resulting in harmonic distortion and in the extreme clipping of signal amplitude. A discussion of noise and linearity will be presented in Sections 2.4.2.5 and 2.4.2.6 below.

A high dynamic range is required for most measurement procedures. Although it is possible in some cases to obtain useful measurements with a 20 dB dynamic range, a range of 40 dB for each vibration and acoustic channel will provide data of much higher quality and reduces the risk of obtaining no data due to the measured signals being above or below the available range of signals that can be measured. It is common to require a 10 dB margin at the lower and upper limits of the dynamic range - that is to design the system so that the minimum expected signal is 10

dB above the noise floor and the maximum signal is 10 dB below the limit of system linearity. A system with 20 dB of dynamic range can meet this requirement only if the amplitude of the signal to be measured is known exactly. This is rarely the case unless amplifier gains can be precisely adjusted for each channel during the course of the measurements.

A system with a 40 dB dynamic range allows measurements with sufficient accuracy to determine vibration and acoustic signal levels and power spectral densities for most applications. Thus, although a wider dynamic range is useful, a system with 40 dB of range per channel is often sufficient.

Many of the techniques for applying measured data to vibration and acoustic problems require data processing beyond the determination of signal levels and power spectral densities. For these techniques a dynamic range of 40 dB is often not adequate. Several of the techniques require the signal amplitudes to be more than 20 dB above the noise. Other techniques require a high degree of linearity so that the signal amplitudes must be less than 20 dB below the upper limit of the dynamic range. For these techniques a system having a dynamic range of 60 dB or more is needed.

The dynamic range of the measurement system is typically limited by the recording device. FM tape recorders operated under IRIG Intermediate band or Wideband Group I specifications provide a dynamic range of approximately 45 dB over a frequency bandwidth that is typically required for vibation and acoustic measurements. A dynamic range greater than 50 dB cannot be achieved without severely limiting the frequency bandwidth.

Two approachs can be used to achieve system dynamic ranges greater then 50 dB. In the first, a digital recording technique is used. The signals are encoded according to one of several methods and then recorded. Use of a 12-bit encoding provides a dynamic range of 72 dB with some loss of frequency bandwidth.

However, by using a wide bandwidth recording device a measurement bandwidth adequate for most vibration and acoustic applications can be obtained.

The second technique to achieve large system dynamic ranges is to use auto-ranging or auto-gain amplifiers. These units adjust the gain according the level of the signal being measured. Although the dynamic range at an given time is limited to the range of the recording device, the amplifier gain will be set to optimize the use of this range. A system of this type reduces the need to predict the expected signal levels with great accuracy in setting up the system gain. A negative feature of the system is that some data is lost while the gain is being changed. Typically, the performance of the auto-ranging amplifiers is inadequate for measuring short duration events or shocks.

Although the dynamic range of the measurement system is typically controlled by the range of the recording device, the system design should evaluate the range component by component. This involves two steps. First, calculation of the system noise and second calculation of the system linearity. These calculations will be discussed in the following Sections.

2.4.2.5 Noise

The lower limit of the dynamic range is determined by the system noise. In well designed systems the noise can be represented by a stationary broadband random time history with a power spectral density varying inversely with frequency. It is common to define the minimum signal that can be measured on the basis of the ratio of the root mean square value (rms) of the signal within the frequency band of interest to the rms value of the noise within the same frequency band. A system with a wide bandwidth will have a higher rms noise and therefore will require a higher value for the minimum signal amplitude. In many cases,

the system noise can be reduced by filtering during later signal processing. It is advisable, therefore, to collect the data with a bandwidth sufficiently wide as to encompass all frequencies of interest in the signal to be measured.

A primary goal in designing a measurement system is to minimize the system noise within the frequency bandwidth of interest. Typically, the design process can be carried out on the basis of the rms noise within the band. However, in special cases it will be desireable to calculate the noise for various frequency bands within the overall bandwidth. The need for such special consideration is dictated by the application for which the data is intended.

The system noise is the additive combination of the noise from each component. The noise for a particular component can be represented by two noise inputs - one at the input to the component and one at the output. For the purposes of system design both noise signals can be assumed to be independent of the component gain. In addition, the two noise signals can be assumed to be statistically independent and to be independent of the input signals to the component. Thus, rms noise amplitudes are additive.

As a general rule, the noise is minimized by using as much gain as possible on the input stages of the system so that the noise generated by these input stages is not amplified by later components. However, great caution is required. Typically, the data recorder will require an rms input signal of approximately 1 volt. The transducer output is typically on the order of a few micro-volts. Therefore, a total channel gain of 90 to 120 dB is often needed. To achieve such high levels of gain a multiple stage instrumentation amplifier is needed. The size, weight, and electrical power needed for such an amplifier usually require it to be located remotely from the transducer with a long cable. Cable noise and pick-up can easily become a

problem thereby eliminating the advantages of having all the gain in the first amplifier of the instrumentation chain. A better approach is to include some gain in a signal conditioning amplifier located near the transducer.

2.4.2.6 Linearity

The upper limit of the dynamic range is governed by system linearity. The ideal system would provide perfect linearity up to a maximum signal value above which the measured signal would be constant - a hard clip. The ideal system would also recover instantly from an overload. The actual performance of many instrumentation components approximates this ideal performance. Component linearity is typically 2% of full scale or less.

In the usual case the system designer need consider only the requirement that the maximum input and output signals for each component be less than the limiting values for that component. The maximum signals are obtained by multiplying the maximum expected rms signal by a peak factor. For random signals it is common to set the peak factor as a factor of three. A higher peak factor is needed for signals containing transient or shock events.

The maximum output signal of a component is typically limited by the component's power supply voltage, which is usually 5 to 15 volts. Thus, many components have an upper limit on the output of +/- 5 volts peak. As a general rule it is good practice to design the system so that the maximum output of each component is under one volt rms. A higher value may be required at the input to the data recorder.

Certain applications require a high degree of linearity. For these cases it is good practice to keep component gains as low as possible and to use special transducers with high linearity and usually a low sensitivity. The signal to noise ratio is decreased by these actions, so that a trade-off must be made between linearity and signal-to-noise.

2.4.2.7 Interference

Measurement system design must consider two types of interference. In the first, the signals to be measured are contaminated by pick-up of extraneous signals due to electromagnetic fields. Typically, problems occur at the system level due to the existence of ground loops. A ground loop is created when the interconnected components are grounded at more than one point. The electromagnetic fields create a current in the ground loop which is converted to a large amplitude voltage signal at the input to high impedance components. Battery operated components can help to eliminate ground loop problems. Optical isolators can also be used to eliminate a ground loop.

The second type of interference is between different channels of a multichannel measurement system. Typically, problems occur when different signals share a single channel on a recording system.

2.4.2.8 System Design Summary

1) Number of Sensors and Data Channels
   Low Frequency Number, bandwidth less than 100 Hz.
   High Frequency Number, bandwidth more than 100 Hz.
2) Dynamic Range Required for each Sensor
   If greater than 40 dB must use special storage techniques or gain ranging amplifiers.
   If less than 40 dB use magnetic tape.
3) Number of Recording Devices
   If number of High Frequency Sensors is less than 12, use single 14 track magnetic tape recorder.
   If more than 12 choose:
   a) time multiplexing using switch box
   b) multiple recorders
   c) on-board data processing

4) Frequency Response

Minimum and Maximum frequencies of interest.

5) Sensor Selection

Variables to be measured

Sensitivity required

Size, weight and cable attachment

Environmental

6) Signal Conditioning

Impedance match

Gain

Filtering

Noise

Amplifier requirements for signal level and impedance

Cable lengths

7) Amplifier Selection

AC or DC

Gain control - manual or coded

Power requirements

Recorder signal levels required

8) Cable Requirements

Lengths

Connectors

Labeling

Bundling and tie-downs

Environmental

9) System Performance

Gain

Frequency Response

Overload

Noise

10) System Calibration

Calculation

Signal insertion

Continuity check

REFERENCES

2.1 Brown, R.V., "Analysis of Shock Indicator Devices for Packing Applications," Wright-Patterson AFB, Rept. 73-52, July, 1973.

2.2 Harris, C.M. and Crede, C.E., Shock and Vibration Handbook, 2nd Edition, McGraw-Hill Book Company, Inc., New York, 1976.

2.3 Doebelin, E.O., Measurement Systems Application and Design, Revised Edition, McGraw-Hill Book Company, Inc., New York, 1976.

CHAPTER 3. DIGITAL DATA ANALYSIS TECHNOLOGY

By

R. G. DEJONG
G. M. GLASS

CHAPTER 3. DIGITAL DATA ANALYSIS TECHNOLOGY

TABLE OF CONTENTS

		Page
	LIST OF FIGURES	3-v
	LIST OF SYMBOLS	3-vii
3.1	CONVERTING ANALOG SIGNALS TO (USEFUL) DIGITAL DATA	3-1
	3.1.1 Sampling Analog Signals	3-3
	3.1.1.1 Sample Rate	3-3
	3.1.1.2 Multichannel Sampling	3-6
	3.1.1.3 Record Length	3-8
	3.1.2 Quantization Process	3-8
	3.1.2.1 Rounding	3-9
	3.1.2.2 Truncating	3-10
	3.1.2.3 Signal to Noise Ratio	3-10
3.2	SINGLE TIME SERIES ANALYSIS	3-11
	3.2.1 Statistical Measures of Digital Data	3-11
	3.2.1.1 Mean Value	3-12
	3.2.1.2 Mean Square Value	3-12
	3.2.1.3 Variance	3-12
	3.2.1.4 Probability Density Function	3-13
	3.2.1.5 Moments of the PDF	3-16
	3.2.1.6 Cumulative Distribution Function	3-17
	3.2.1.7 Accuracy of Statistical Measures	3-17
	3.2.1.8 Confidence Intervals	3-18
	3.2.1.9 Significance Levels	3-26

3.2.1.10 Time Averaging vs. Ensemble Averaging..... 3-27
3.2.1.11 Tests for Stationarity..... 3-28
3.2.2 Autocorrelation Function..... 3-31
3.2.2.1 Statistical Errors in the ACF Estimate..... 3-34
3.2.3 Power Spectral Density Function..... 3-40
3.2.3.1 Direct Method..... 3-41
3.2.3.2 Indirect Method..... 3-52
3.2.3.3 Calibration and Units..... 3-54
3.2.3.4. Statistical Errors in the PSD Estimate..... 3-55
3.2.4 Digital Filtering..... 3-58
3.2.4.1 Nonrecursive Filters..... 3-60
3.2.4.2 Recursive Filters..... 3-64
3.2.4.3 Integration..... 3-68
3.2.4.4 Differentiation..... 3-70
3.2.4.5 Interpolation..... 3-71
3.2.4.6 Statistical Errors in Digital Filtering..... 3-73
3.2.5 Special Analysis Considerations..... 3-74
3.2.5.1 Periodic Signals..... 3-74
3.2.5.2 Non-stationary Signals..... 3-76
3.3 DUAL TIME SERIES..... 3-79
3.3.1 Joint Statistical Measures..... 3-79
3.3.1.1 Joint Probability Density Function.... 3-79
3.3.1.2 Joint Cumulative Distribution Function..... 3-80
3.3.1.3 Conditional Probability Density Function..... 3-81
3.3.1.4 Covariance..... 3-81

3.3.1.5 Regression Analysis..... 3-82
3.3.1.6 Accuracy of Joint Statistical Measures..... 3-83
3.3.2 Cross Correlation Function..... 3-85
3.3.2.1 Statistical Errors in the CCF Estimate..... 3-88
3.3.3 Cross Spectral Density Function..... 3-89
3.3.3.1 Direct Method..... 3-91
3.3.3.2 Indirect Method..... 3-94
3.3.3.3 Time Delay Corrections..... 3-96
3.3.3.4 Calibration and Units..... 3-96
3.3.3.5 Coherence Function..... 3-97
3.3.3.6 Frequency Response Functions..... 3-100
3.3.3.7 Statistical Errors in the CSD Estimate..... 3-101
3.3.4 Cepstral Analysis..... 3-102
3.4 MULTIPLE TIME SERIES..... 3-105
3.4.1 Multiple Statistical Measures..... 3-105
3.4.1.1 Joint Probability Density Function.... 3-106
3.4.1.2 Joint Cumulative Distribution Function..... 3-107
3.4.1.3 Conditional Probability Density Function..... 3-107
3.4.1.4 Multiple Regression Analysis..... 3-108
3.4.1.5 Multiple Correlation Coefficient..... 3-110
3.4.2 Multiple Correlation Relationships..... 3-110
3.4.2.1 Multiple Correlation Function..... 3-111
3.4.2.2 Partial Correlation Function..... 3-112

3.4.3 Multiple Spectral Analysis..... 3-114

3.4.3.1 Multiple Coherence Function..... 3-115

3.4.3.2 Partial Coherence Function..... 3-115

3.4.3.3 Multiple Frequency Response Function..... 3-117

3.4.3.4 Statistical Errors in Multiple Time Series Analysis..... 3-118

REFERENCES..... 3-119

LIST OF FIGURES

Figure		Page
3-1.	Analog to Digital Conversion.....	3-2
3-2.	Aliasing in Digital Sampling of Components from Signal in Figure 3-1a. Components a, b, and c are below sample rate, while d is above sample rate, resulting in the apparent signal component d'.....	3-5
3-3.	Multiplex Sampling of Three Signals with a Sampling Time Interval of τ and a Time Shift Between Signals of t_s.....	3-7
3-4.	Probability Density Function of Sampled Time Histories Plotted as Histograms of Sampled Values..	3-14
3-5.	Probability Density Function of Some Typical Signals.....	3-16a
3-6.	Cumulative Distribution Function of Some Typical Signals.....	3-17a
3-7.	Characteristic of a Random Variable x with a Normal Distribution Having a Mean Value μ and a Variance σ^2.....	3-21
3-8.	Confidence Intervals for the Sampled Values of a Random Variable $T_n = \sqrt{N}(\bar{x} - \mu)/s$ with a Student's T Distribution Having $\nu = N - 1$ Degrees of Freedom...	3-22
3-9.	Confidence Intervals for the Sampled Values of a Random Variable $\chi^2_\nu = \nu s^2/\sigma^2$ with a Chi-Squared Distribution Having $\nu = N - 1$ Degrees of Freedom...	3-24
3-10.	Confidence Intervals for the Sampled Values of the Random Variable $F_{\nu_1\nu_2}$ with a Fisher's Distribution Having $\nu = N - 1$ Degrees of Freedom.....	3-30
3-11.	Autocorrelation Function of Some Typical Signals...	3-33
3-12.	Normalized Mean Square Error Bound of Autocorrelation Function Estimate.....	3-38
3-13.	Power Spectral Density of Some Typical Signals.....	3-44

3-14. Characteristics of Selected Window Functions..... 3-48

3-15. Example of Errors in the Power Spectral Density Estimate of a Signal with a Non-Zero Mean Value and Slope..... 3-51

3-16. Rectangular Filter Shapes..... 3-61

3-17. Cross Correlation Function of Some Typical Signal Pairs..... 3-87

3-18. Cross Spectral Density of Some Typical Signal Pairs..... 3-92

3-19. Coherence Function of Some Typical Signal Pairs.... 3-99

LIST OF SYMBOLS

A	Constant
$\underline{A}$	Matrix of multiple regression coefficients
A/D	Analog to digital
A(t), B(t)	Non-stationary time functions
acf	Autocorrelation function
a_3	Skewness
a_4	Kurtosis
$a_k(c)$	Moment of order k about axis c
acvf	Autocovariance function
B	Bias error
b	Number of binary digits
b_m	Digital filter parameter
$\underline{C}$	Matrix of cross-covariance functions
$C_{ij \mid l}$	Partial covariance function (effects of l removed)
C_x	Sample acvf of x
$C_{xy}(f)$	Coincident spectral density function
$C_{xy}(t)$	Sample ccvf
c	Axis of moment
ccf	Cross correlation function
ccvf	Cross covariance function
cdf	Cumulative distribution function
cos	Cosine function
cov	Covariance
csd	Cross spectral density
$\underline{C}_{xx}$	Covariance matrix
c_{xy}	Sample covariance
DFT	Discrete Fourier transform
dB	Decibel

$d(t)$	Signal component
d_n	Signal component at time sample n
$E[x]$	Expected value of x
E_x	Total energy of a transient signal x
E_r	Rounding error
E_t	Truncation error
e	Natural antilog of 1
F	Ratio of sample variances
FT	Fourier transform
FFT	Fast fourier transform
FIR	Finite impulse response
f_b	Effective filter bandwidth (Hertz)
f_c	Center frequency (Hertz)
f_l	Lower cut-off frequency (Hertz)
f_m	Maximum signal frequency (Hertz)
f_{min}	Lowest resolvable frequency (Hertz)
f_q	Nyquist or folding frequency (Hertz)
f_s	Sample rate (Hertz)
f_u	Upper cut-off frequency (Hertz)
$G_x(f)$	Psd of signal x at frequency
$\underline{G}_{xx(f)}$	Matrix of cross spectral density functions
$G_{xy}(f)$	Cross spectral density at frequency f
$H(f)$	Transfer function
$H_{y\|x}(f)$	Frequency response function
$h(t)$	Impulse response
IIR	Infinite impulse response
Im[ ]	Imaginary part of complex quantity
Int	Truncation function (integer value)
i	Summation index
j	$\sqrt{-1}$
K	Number of adjacent frequency samples averaged, also, the order of a Butterworth filter
$K_H(f)$	Natural log (complex) of H(f)

$k_H(q)$	Complex cepstrum of H(f)
$K_y(f)$	Natural log of $G_y(f)$
k	Order of moment, also, digital frequency index
$k_y(q)$	Cepstrum of $G_y(f)$
L	Number of α_l digital filter parameters
lim	Limit
M	number of intervals in pdf, also Number of records for averaging , also Number of time signals
m	Digital delay index
mse	Mean square error
N	Record length of sampled signal, number of samples
n	Digital sample index
$P_{ij \mid l}$	Partial correlation function (effects of l removed)
$P(X_i \mid Y_j)$	Conditional pdf of x given by y
$P(X_i)$	Sample pdf
$P(X_k)$	cdf for X_k
P_{xy}	Normalized cross-correlation function
p	Power of two for FFT
pdf	Probability density function
psd	Power spectral density function
$Q_{xy}(f)$	Quadrature spectral density function
q	Time delay (seconds)
Re[ ]	Real part of complex quantity
R_{xy}	Sample ccf
r_{xy}	Correlation coefficient
$r_{y \mid x}$	Multiple correlation coefficient
SNR	Signal to noise ratio
SR	Sample rate (Hertz)
s	Sample standard deviation
sin	Sine function

T	Time period of sampling (seconds), also matrix transpose as superscript
tan	Tangent function
T_p	Lag time corresponding to an integer number of periods
t_n	Time at time sample n (seconds)
t_s	Time shift between signals (seconds)
U	Unit of measure
V	Full scale range of A/D converter (volts)
$V[x]$	True variance of x
W_x	Width of x interval for pdf
w	Value range interval
w_T	Sampling window of length T
X_i	Center of i'th x interval
$X_T(f_k)$	DFT of signal x at frequency f_k
x_a	Actual signal level
x_q	(Digital) quantized signal level
$x(t)$, $y(t)$	Signals
Y_n	Filter output
Z_k	Complex array element for FFT
$>$	Greater than
$<$	Less than
π	Ratio of a circle's circumference to its diameter (pi)
Δf	Difference in frequency (Hertz)
τ	Sampling time interval (seconds)
ε	Quantization error
μ	Mean value
σ	Standard deviation
ε_r	Rounding error
ε_t	Truncation error

ν	Number of degrees of freedom
T_ν	Student's T distribution
χ^2_ν	Chi-squared distribution
$F_{\nu 1,\nu 2}$	Fisher's F distribution
∞	Infinity
ρ_{xx}	Normalized correlation function of x
α	Exponential decay constant
δf	Filter bandwidth (Hertz)
$\|\ \|$	Magnitude of a complex quantity
$\neq$	Not equal
Π	Product
$<<$	Much less than
$>>$	Much greater than
*	Convolution operator, also, complex conjugate as superscript
$\Phi_{xy}(f)$	Phase angle between x and y at frequency f
$\gamma^2_{xy}(f)$	Coherence function
$\gamma^2_{y\|x}(f)$	Multiple coherence function
$\gamma^2_{ij\|1}(f)$	Partial coherence function (effects of 1 removed)
'(apostrophe)	Differentiation with respect to functional variable

CHAPTER 3. DIGITAL DATA ANALYSIS TECHNOLOGY

In this chapter we present various methods of analyzing measured data. We assume here that the analysis will be done digitally, that is, the measured signals are quantified within some range of numbers. Analog processing methods (such as frequency band processing) are included in the discussions of measurement technology in Chapter 2 under the heading of signal conditioning. This division between analog processing and digital analysis is made for two reasons. First, in almost all uses of measured data the measurements are at some point quantified, if merely to evaluate an average or peak level, for example. Second, this enables us to differentiate between the effects of the analog instrumentation used to condition and process the measured signals and the effects of the manipulation of the measured data in digital form (which are assumed to be independent of the instrument on which the manipulation is done). The major exception to this distinction is considered in the next section on converting analog signals to digital data.

3.1 CONVERTING ANALOG SIGNALS TO (USEFUL) DIGITAL DATA

The methods of analog processing and digital analysis overlap in the process of converting analog signals to digital data. Analog to digital (A/D) conversion involves two processes, sampling and quantization, as shown in Figure 3-1. Sampling is the process of measuring the value of a continuous signal at a finite number of points in time. Quantization is the process of rounding or truncating the measured value so it can be represented by a finite number of digits.

We will first consider sampling exclusive of quantization effects. We will restrict our discussion to periodic sampling, or sampling at equally spaced time intervals. We will then consider quantization.

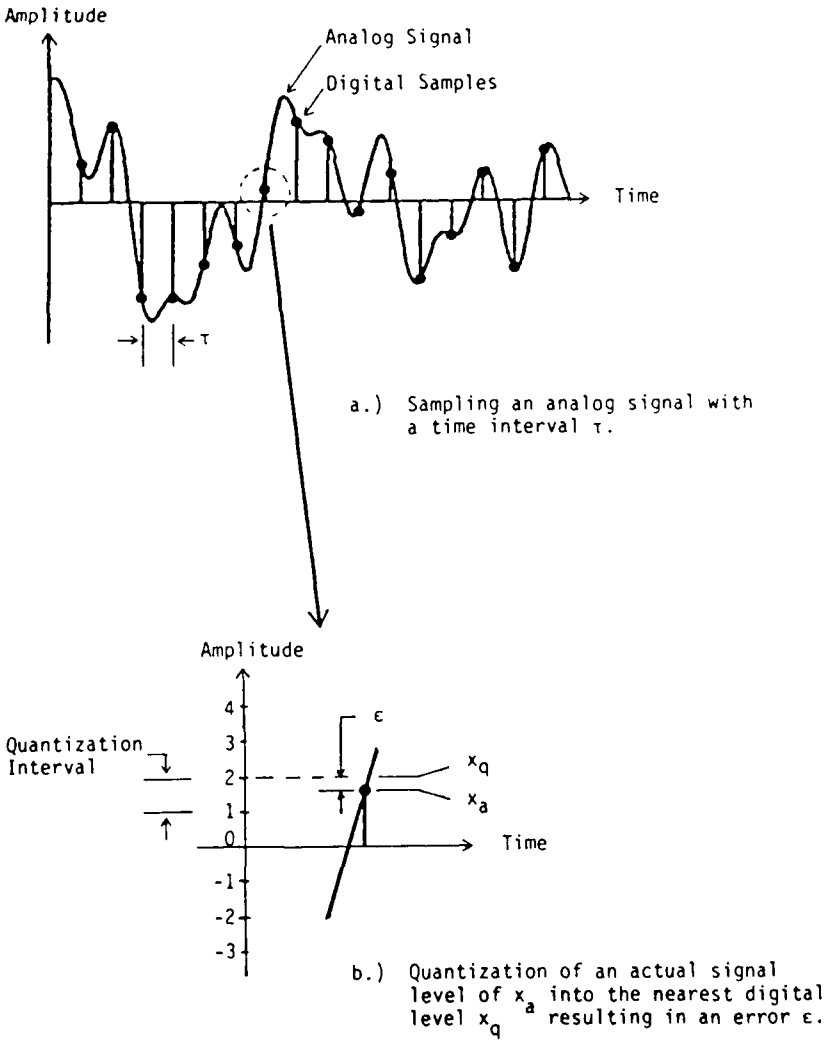


Figure 3-1. Analog to Digital Conversion

3.1.1 Sampling Analog Signals

The relationship between a finite sequence of numbers obtained by periodic sampling and the original analog signal depends on the frequency content of the signal, or the rate at which it is changing in value. In general terms the rate at which the signal is sampled must be faster than the fastest rate at which the signal is changing in value if accurate frequency information is required. In addition, the duration of the sampling (called the record length) must be longer than the longest period of fluctuation in the value of the signal if consistent time averaged information is required.

The following paragraphs present useful guidelines for the proper selection of sampling parameters. The detailed manner in which these parameters affect the various types of analysis to be performed is discussed in the individual sections of this chapter which outline these analysis procedures.

3.1.1.1 Sample Rate

In order to preserve accurate frequency information when digitally sampling an analog signal the sample rate SR (in samples per second) must be greater than twice the maximum frequency f_m (in cycles per second, or Hertz) which is present in the signal, or

$$SR > 2\ f_m. \qquad (3\text{-}1)$$

If this condition is not met the signal content outside the frequency range $0 \leq f < SR/2$ will be confused with frequencies within this range. This phenomenon is called aliasing. The frequency $f_q = SR/2$ is called the Nyquist frequency.

If the maximum frequency content of the signal is not known it is necessary to filter out any undesired frequencies with an anti-aliasing filter before the signal is sampled. (For detailed information on analog filtering see Section 2.2.8.) A typical choice of such a filter is a low-pass filter with a high

frequency roll-off rate of approximately 24 dB per doubling of frequency. In most cases [3-1] it is sufficient to set the upper cut-off frequency f_u (defined as the 3 dB attenuation frequency) by

$$f_u < 0.4 \text{ SR}. \tag{3-2}$$

This will give an attenuation of approximately 10 dB at the Nyquist frequency and will usually avoid contamination of the data below the filter cut-off frequency. However, the filter attenuation characteristics must be accounted for in the frequency range of the filter cut-off.

A more detailed look at the phenomenon of aliasing can be made using the illustration in Figure 3-1a. This signal consists of the sum of four sinusoids with different frequencies as shown in Figure 3-2. Components a, b, and c have frequencies below the Nyquist frequency and therefore are accurately preserved by the sampling. However, component d has a frequency greater than the Nyquist frequency and the sampled data cannot distinguish it from the lower frequency sinusoid d' shown by the dashed line superimposed over component d.

The phenomenon of aliasing can be described mathematically by representing the signal component d by a sine wave of frequency $(f_q + \Delta f)$ with $0 < \Delta f < f_q$:

$$d(t) = \sin [2\pi t(f_q + \Delta f)] \tag{3-3}$$

when this signal is sampled at times $t_n = n\tau$; $n = 1, 2..., N$; the resultir data can be represented by:

$$d_n = \sin [2\pi n\tau(f_q + \Delta f)]. \tag{3-4}$$

Since $f_q = \frac{1}{2} \text{ SR} = \frac{1}{2\tau}$:

$$d_n = \sin [\pi n + 2\pi n\tau\Delta f]. \tag{3-5}$$

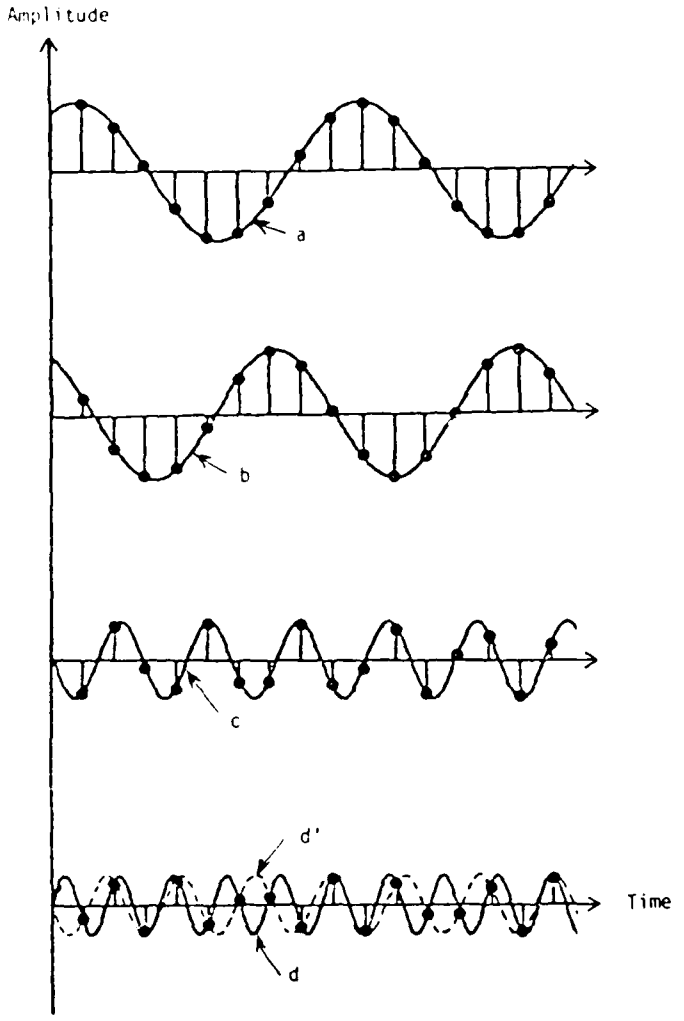


Figure 3-2. Aliasing in Digital Sampling of Components from Signal in Figure 3-1a. Components a,b, and c are below sample rate, while d is above sample rate, resulting in the apparent signal component d'.

Since the sine function is periodic with 2π:

$$d_n = \sin [-\pi n + 2\pi n\tau\Delta f]$$

$$= \sin [-2\pi n\tau(f_q - \Delta f)]$$

$$= -\sin [2\pi n\tau(f_q - \Delta f)] \qquad (3\text{-}6)$$

Therefore, when d(t) is sampled it looks identical to a signal with frequency $(f_q - \Delta f)$ which is the frequency of the aliased signal d'(t) shown in Figure 3-2.

If a signal is sampled without aliasing, it is theoretically possible to reconstruct from the sampled data the value of the analog signal at any time between samples. This process is called interpolation and is described in Section 3.2.4.5.

3.1.1.2 Multichannel Sampling

If it is required to sample more than one analog signal for purposes of relating their phase or time characteristics it is necessary to sample the various signals over the same time period. However, it is not necessary that they be sampled at the exact same instants. It is possible to sample multiple signals with a known time shift between corresponding samples, as shown in Figure 3-3, and still preserve accurate time and phase information. This process is called multiplexing and allows for multiple signals to be processed through a single A/D converter.

The time shift introduced by multiplexing can be removed in two ways. The easiest and most frequently used method involves a phase shift in the frequency spectrum if the signals are analyzed by Fourier Transform. This is discussed in Section 3.3.3.3. A second method involves interpolating the signals after they are sampled to obtain sample values at exactly the same times. Interpolation methods are described in Section 3.2.4.5.

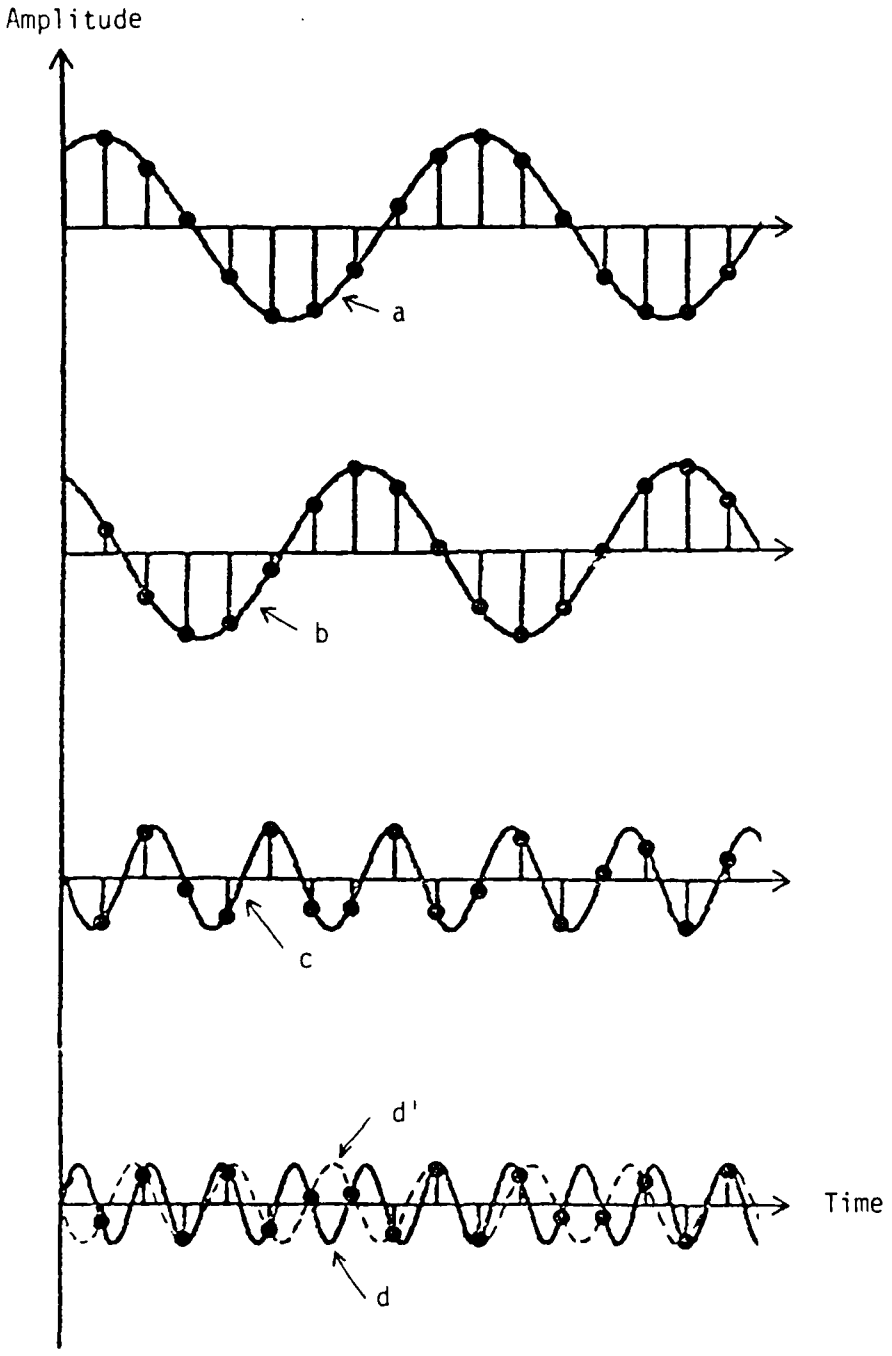


Figure 3-2. Aliasing in Digital Sampling of Components from Signal in Figure 3-1a. Components a, b, and c are below sample rate, while d is above sample rate, resulting in the apparent signal component d'.

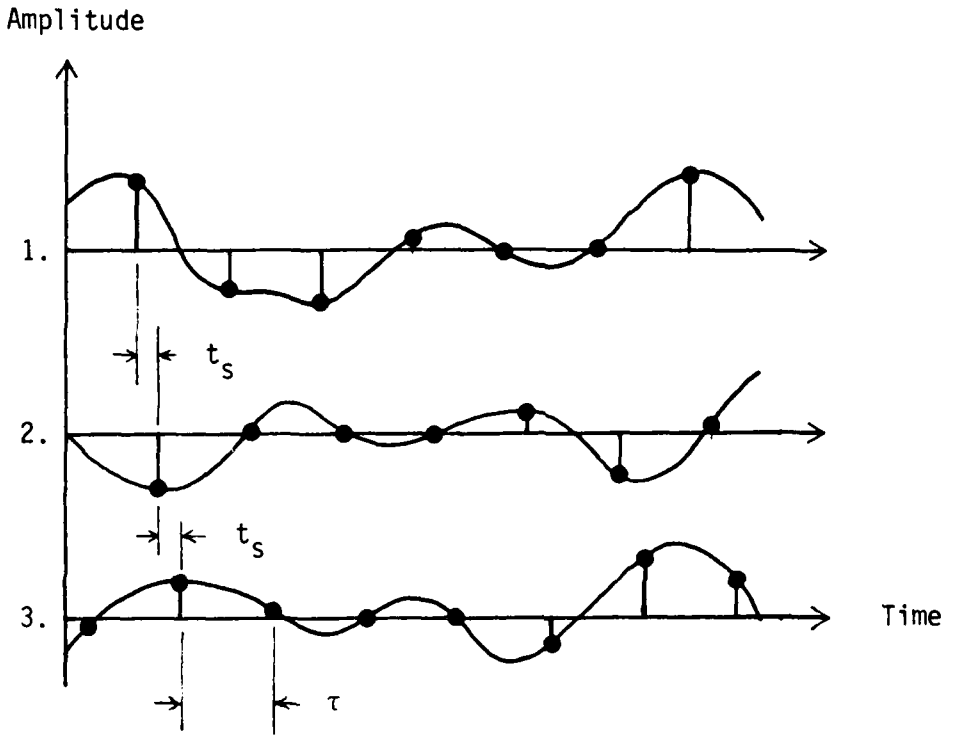


Figure 3-3. Multiplex Sampling of Three Signals with a Sampling Time Interval of τ and a Time Shift Between Signals of t_s

Data from Ref. 3-6

3.1.1.3 Record Length

The record length is important when time averaging is required. It also controls the lower frequency limit in a frequency analysis of the data. If a signal is sampled at regular intervals over a time period T (in seconds), then the lowest frequency f_{min} that can be resolved by standard frequency analysis is

$$f_{min} = 1/T. \tag{3-7}$$

If lower frequencies are present, the frequency analysis of the sampled data may be contaminated. This is discussed in more detail in Section 3.2.3.

The requirements on the record length for time averaging are less precise because they involve statistical sampling criteria. The relationship between the record length and the statistical accuracy of a particular analysis method is included in the Section of this Chapter which outlines that procedure. In many cases the record length is limited by the number of samples which can be stored in the digital processing hardware. Averages over longer periods of time can then be accomplished by repeating the sampling process a number of times and averaging the results of each sample record.

Averaging over a number of sample records to obtain an accurate time average assumes that the signal is stationary in time (its statistical properties do not change with time). Tests to determine whether or not a signal is stationary are given in Section 3.2.1.11. Special considerations for analyzing transient or non-stationary signals are given in Section 3.2.5.

3.1.2 Quantization Process

We now consider the effects of quantization on the analog to digital (A/D) conversion process. As noted above, quantization is the process of rounding or truncating a signal level so

that it can be recorded by a finite number of digits (see Figure 3-1b). In general the process of rounding is more desirable than truncating. The mean error introduced by rounding is zero, while the process of truncating introduces a bias error. Quantization also introduces noise into the signal which is of equal magnitude for both rounding and truncating. The following paragraphs give a more detailed description of these errors.

The quantized signal level x_q can be modeled as the sum of the actual level x_a and a quantization error ε:

$$x_q = x_a + \varepsilon \tag{3-8}$$

Since ε is not known, it is necessary to describe it statistically. The following assumptions have been shown to be valid for most practical cases [3-2]:

1) The properties of ε do not vary with time.

2) The values of x_a and ε are linearly independent.

3) The probability that ε takes on a particular value is constant over the entire range of ε.

4) The value of ε for one sample is linearly independent with that of another sample.

We will now consider the effects of both rounding and truncating using these assumptions. For this analysis we will assume that a binary A/D converter is used with b digits and a dynamic range/voltage range of -V to V, so that the difference between successive quantization levels is $2V/2^b$.

3.1.2.1 Rounding

The error introduced by rounding, ε_r, may be either positive or negative and its magnitude is limited by one half the distance between quantization levels:

$$-V/2^b < \varepsilon_r \leq V/2^b \tag{3-9}$$

It can be shown [3-3] that the mean value of the rounding error μ_{ε_r} is

$$\mu_{\varepsilon_r} = 0 \tag{3-10}$$

and the variance $\sigma^2_{\varepsilon_r}$ is

$$\sigma^2_{\varepsilon_r} = \frac{1}{3} (V/2^b)^2. \tag{3-11}$$

3.1.2.2 Truncating

The error introduced by truncating, ε_t, is always negative and its magnitude is limited by the distance between quantization levels,

$$-2V/2^b < \varepsilon_t \leq 0. \tag{3-12}$$

It can be shown [3-3] that the mean value of the truncating error μ_{ε_t} is

$$\mu_{\varepsilon_t} = -V/2^b \tag{3-13}$$

and the variance $\sigma^2_{\varepsilon_r}$ is

$$\sigma^2_{\varepsilon_t} = \frac{1}{3} (V/2^b)^2. \tag{3-14}$$

3.1.2.3 Signal to Noise Ratio

The signal to noise ratio (SNR) for both rounding and truncating can be found by comparing the variance of the actual signal with that of the quantization error. If we assume x_a is a

random signal which has been scaled so that its variance $\sigma^2_{x_a}$ is

$$\sigma^2_{x_a} = (V/4)^2 \tag{3-15}$$

then

$$\begin{aligned} SNR &= 10 \log (\sigma^2_{x_a} / \sigma^2_\varepsilon) \\ &= 6.0b - 7.3 \text{ dB} \end{aligned} \tag{3-16}$$

where b is the number of quantizer "bits."

The actual SNR will, of course, depend on the scaling of the input signal. However, this example indicates the importance of using the full dynamic range of the A/D converter whenever possible.

3.2 SINGLE TIME SERIES ANALYSIS

In this section we present various methods of analyzing digital data obtained from a single time series. We begin with statistical measures, then move to correlation and spectral analysis, digital filtering, and finish with some specialized analysis methods.

3.2.1 Statistical Measures of Digital Data

In this section we outline the procedures for computing various statistical measures of digital data. We begin without assuming anything about the origin of the digital data, considering them to be merely a set of sampled values to be analyzed. The data could be periodic samples of an analog signal or they could be an ensemble of successive estimates of some signal

characteristic, such as the magnitude of the frequency spectrum at a particular frequency.

A statistical measure computed from a subset of the available data is denoted as a "sample" measure in order to indicate that it is an estimate of the actual measure obtained from a sample of the data.

3.2.1.1 Mean Value

The sample mean value $\bar{x}$ of a set of N values x_n is computed by

$$\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n \qquad (3\text{-}17)$$

The sample mean value is a measure of the arithmetic average of the set of values x_n and is an estimate of the expected value of the data denoted by $\mu_x = E[x]$.

3.2.1.2 Mean Square Value

The sample mean square value $\overline{x^2}$ of a set of N values x_n is computed by:

$$\overline{x^2} = \frac{1}{N} \sum_{n=1}^{N} x_n^2 \qquad (3\text{-}18)$$

3.2.1.3 Variance

The sample variance s^2 of a set of N values x_n is computed by:

$$s^2 = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \bar{x})^2 = \frac{1}{N-1} \left[\sum_{n=1}^{N} x_n^2 - N(\bar{x})^2 \right] \qquad (3\text{-}19)$$

The square root of the sample variance is called the standard deviation, s. The sample variance is a measure of the spread of the set of values x_n about the sample mean value. It is an estimate of true variance of the data about its expected value which is defined by $\sigma^2 = V[x] = E[(x-\mu_x)^2]$.

Note that the factor $1/(N-1)$ is used in Equation (3-19) instead of $1/N$. This is necessary in order to make the estimation of the variance an unbiased one [3-4]; that is, the statistical errors in the computation are equally likely to be positive or negative.

3.2.1.4 Probability Density Function

The mean and variance are parameters of a more general statistical descriptor called the probability density function (pdf). The sample pdf of a set of N values x_n is related to the number of values in the set that fall within specific value intervals. If the range of x_n values is divided into intervals of width w_x, centered at values X_i, then an array N_i can be defined as the number of sample values that lie in each value interval. This array can be plotted as a histogram as shown by the examples in Figure 3-4. The sample pdf, denoted by $p(X_i)$, is computed by dividing each value N_i by the width of the value interval w_x and by N:

$$p(X_i) = \frac{N_i}{w_x N}. \qquad (3\text{-}20)$$

The magnitude of the computed pdf will depend somewhat on the size of w_x. However, the value of $p(X_i)$ will converge to a consistent level as w_x is made smaller as long as N is sufficiently large.

A procedure for computing a constant interval pdf from a set of N values x_n is given in Reference [3-5] and is summarized in the following table with slight modifications.

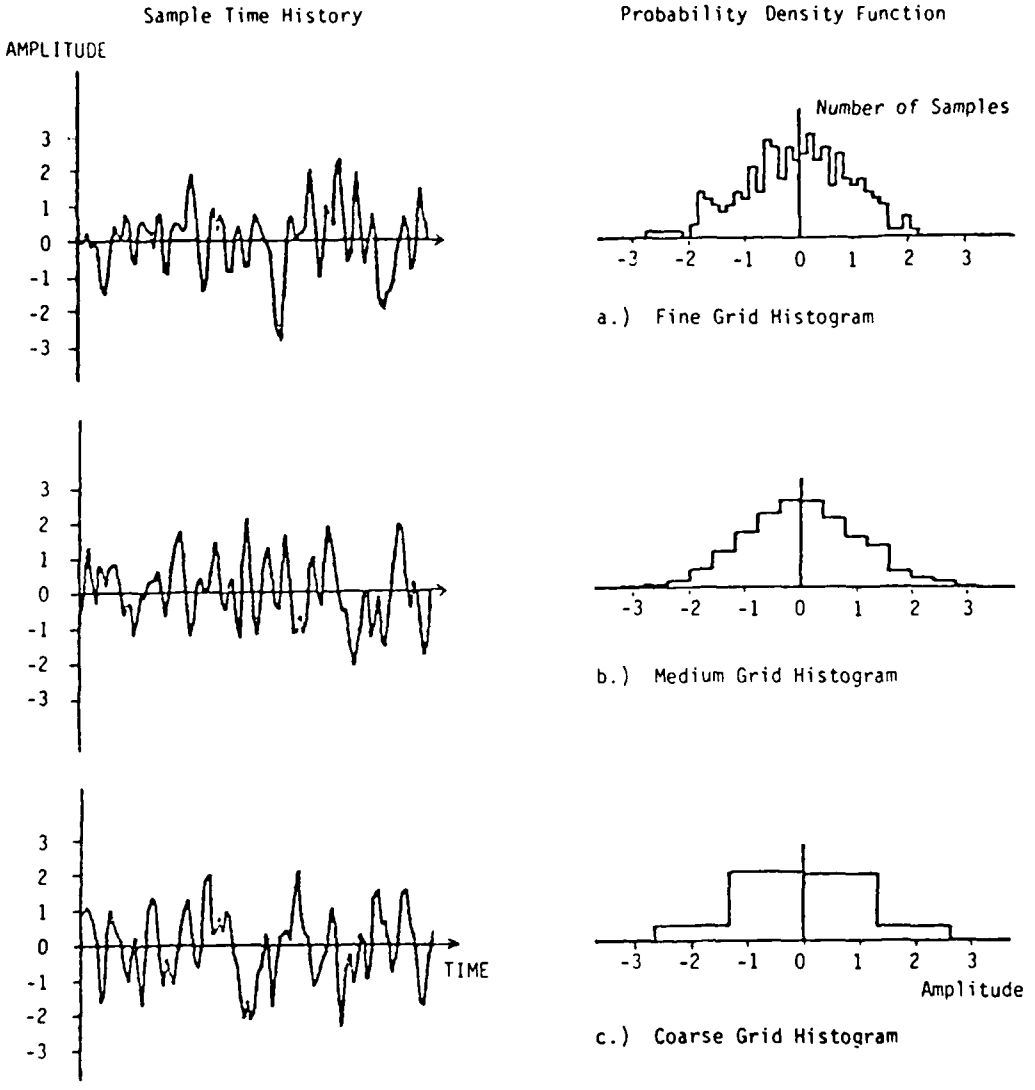


Figure 3-4. Probability Density Function of Sample Time Histories Plotted as Histograms of Sampled Values

Table 3-1.

1. Determine the range of values, $x_n = a$ to $x_n = b$, to be covered by the pdf and the number of value intervals M in the pdf. The width of each interval is given by $w_x = \frac{b-a}{M-1}$. The center point of each interval, X_i; i = 1, 2, ..., M; is given by $X_i = a + (i-1)w_x$.

2. Define a sequence of M+2 numbers, N_i; i = 0, 1, ..., M+1; where:

N_0 = number of x_n such that $x_n \leq a - \frac{w_x}{2}$

N_i = number of x_n such that $X_i - \frac{w_x}{2} < x_n \leq X_i + \frac{w_x}{2}$

N_{M+1} = number of x_n such that $x_n > b + \frac{w_x}{2}$

Initially set the values of all $N_i = 0$.

3. Compute the integer $i = M + 1 - \text{Integer } [\frac{b-x_n}{w_x} + 1.5]$

- if $i < 1$, add the integer one to N_0
- if $i > M$, add the integer one to N_{M+1}
- otherwise, add the integer one to N_i

4. Use the resulting sequence of numbers N_i to compute the pdf using Equation (3-20).

The resulting pdf meets the following constraint:

$$\sum_{i=0}^{M+1} w_x \, p(X_i) = 1. \qquad (3\text{-}21)$$

The accuracy of the pdf will depend on the choice of w_x for a given number N samples. A good choice for most applications is

$$w_x = \frac{20s}{\sqrt{N}} \qquad (3\text{-}22)$$

where s is the standard deviation of the data. If s is not known, it can be approximated by assuming that a range of 8s will encompass the dynamic range of the data. As w_x is made smaller the resolution of the pdf increases while the variance of the magnitudes increases, causing the pdf to look noisy. As w_x is made larger the pdf becomes smoother but with a sacrifice in resolution. This effect is seen in the examples of histograms shown in Figure 3-4.

The pdf computed from a sample record can be compared with those of known functions to identify the characteristics of the sampled data. Figure 3-5 gives examples of some typically encountered signals and their computed pdf. More information on these and other types of probability density functions can be found in Reference [3-6].

3.2.1.5 Moments of the PDF

The pdf of a set of N values x_n can be described by a series of moments defined by the average value of $(x_n - c)^k$, where k is the order of the moment and c is the axis of the moment. The moments $\alpha_k(c)$ can be computed by:

$$\alpha_k(c) = \frac{1}{N} \sum_{n=1}^{N} (x_n - c)^k \simeq \sum_{i=0}^{M+1} (X_i - c)^k \, p(X_i) \qquad (3\text{-}23)$$

The first moment about zero is the mean value as given in Equation (3-17). The second moment about zero is the mean square value as given in Equation (3-18). The second moment about the mean value is an estimation of the variance for large N as given in Equation (3-19).

Higher order moments are usually computed for the normalized value $(x_n - \bar{x})/s$. The third order normalized moment is called skewness, a_3:

$$a_3 = \frac{1}{N} \sum_{n=1}^{N} \left(\frac{x_n - \bar{x}}{s}\right)^3 \qquad (3\text{-}24)$$

The fourth order normalized moment is called the kurtosis, a_4:

$$a_4 = \frac{1}{N} \sum_{n=1}^{N} \left(\frac{x_n - \bar{x}}{s}\right)^4 \qquad (3\text{-}25)$$

These normalized moments of some common pdf's are given in Reference [3-6] and can be used to help identify the characteristics of the data being analyzed.

3.2.1.6 Cumulative Distribution Function

The cumulative distribution function (cdf) of a set of N values x_n is a measure of the probability that the value of x_n is less than or equal to a certain value X_k. The cdf, denoted by $P(X_k)$, can be computed from the pdf by summing up the values of the pdf at all intervals where $X_i \leq X_k$:

$$P(X_k) = \sum_{i=0}^{k} p(X_i) \qquad (3\text{-}26)$$

Figure 3-6 gives examples of the cdf of some typically encountered signals.

3.2.1.7 Accuracy of Statistical Measures

When a statistical measure is evaluated for a set of data

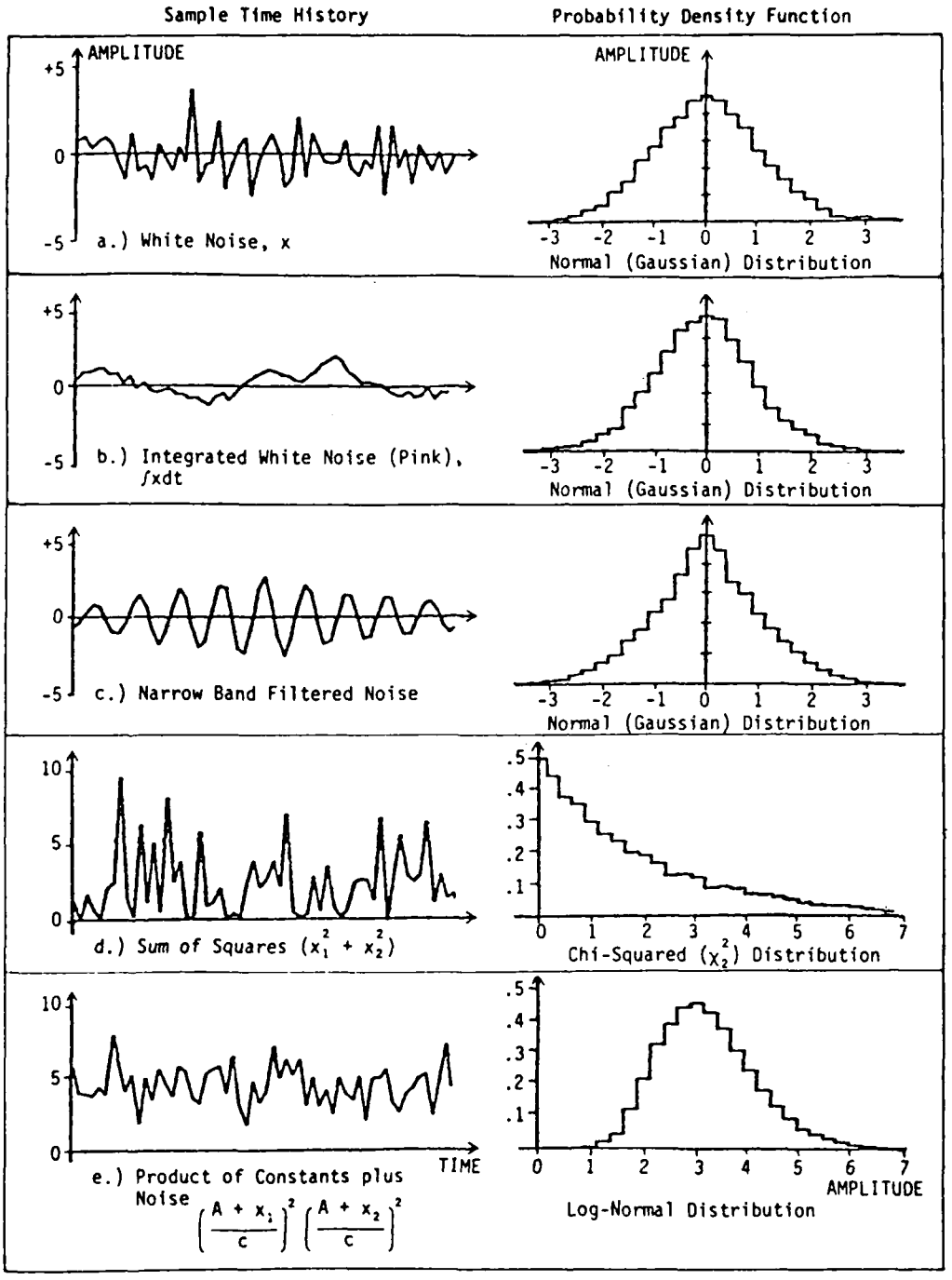


Figure 3-5. Probability Density Function of Some Typical Signals.

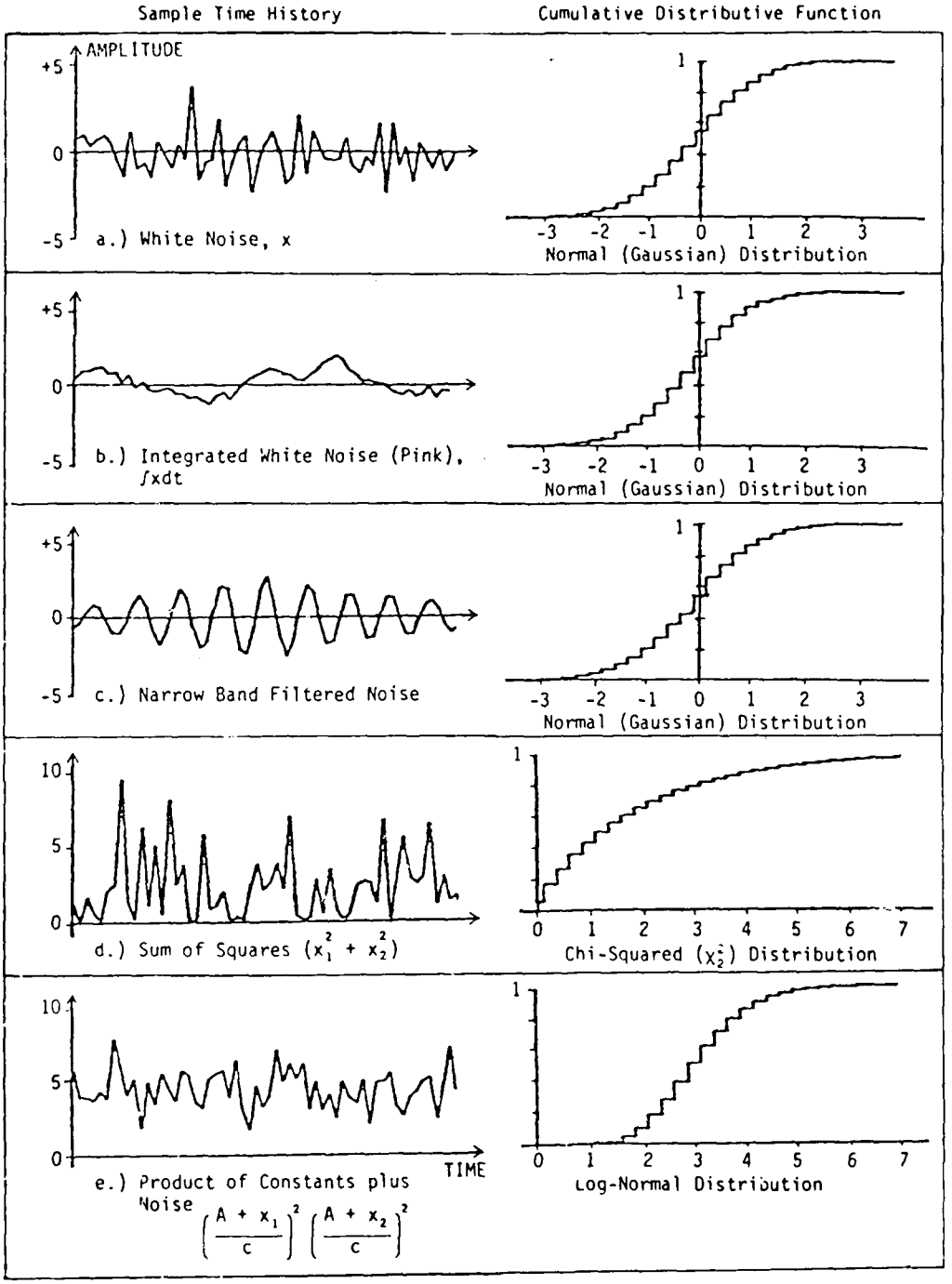


Figure 3-6. Cumulative Distribution Function of Some Typical Signals

that is only a subset of a longer signal, it is desirable to have some information about the accuracy of the measure in estimating the true value of the desired parameter of the signal. Two approaches can be taken to specify the accuracy of a statistical measure. One approach is to assume the signal has a certain statistical distribution and then evaluate the potential for errors in calculating the desired parameter of the distribution from a finite set of data. This approach leads to the specification of confidence intervals around the estimated statistical measures. The second approach is to assume the signal has a certain statistical distribution with known parameters and then develop a test for determining whether or not the finite set of data confirms this assumption. This approach leads to the specification of the levels of significance for the test.

These two approaches have similarities in the basic principles from which they are derived, but they have different applications. In the following section we present summaries of their application to the analysis of digital data and refer the reader to other sources for a more detailed development of the principles being used.

3.2.1.8 Confidence Intervals

The confidence intervals of a statistical measure are based on the probability distribution of that measure when it is computed for a number of different sets of data taken from the same process. For instance, if a continuous random signal is digitally sampled and the mean value of a finite number of samples is evaluated, this mean value $\bar{x}$ will be a random variable with a certain pdf. If a certain form of the pdf is assumed, such as the Normal distribution, then confidence intervals can be set around the value of $\bar{x}$ based on the probable distribution of all possible values. The following paragraphs present formulas for determining the confidence intervals for various statistical measures.

SAMPLE MEAN VALUE: The sample mean value of a random signal can usually be assumed to have a pdf like the Normal distribution. This is true even if the random signal does not have a Normal pdf [3-7]. If a random signal x has a mean value μ and a variance σ^2, then the sample mean value calculated using Equation (3-17) has an expected value given by:

$$E[\bar{x}] = \mu. \tag{3-27}$$

The variance of the estimated mean is

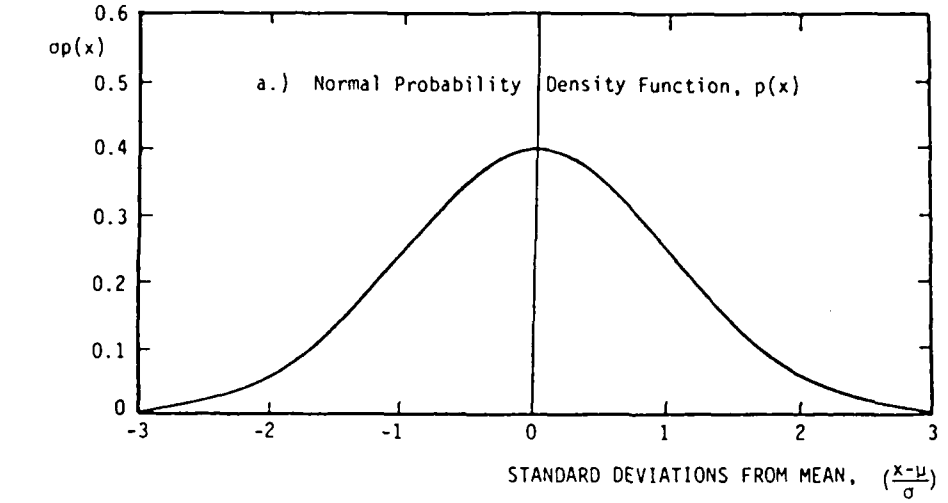
$$V[\bar{x}] = \frac{\sigma^2}{N}. \tag{3-28}$$

where N is the number of samples of x used to calculate the mean value.

The characteristics of the Normal pdf can then be used to evaluate confidence intervals for the sample mean value $\bar{x}$ as shown in Figure 3-7. For example, the confidence interval at $\pm$ one standard deviation from the mean has a 68% confidence level. Therefore, based on the sample mean value $\bar{x}$ of N numbers (which has a standard deviation of $\sigma/\sqrt{N}$) it can be stated that the true mean μ is within $\pm\ \sigma/\sqrt{N}$ of $\bar{x}$ with 68% confidence. The interval $\bar{x} \pm \sigma/\sqrt{N}$ is then called the 68% confidence interval for the sample mean value.

One practical limitation in applying the confidence intervals based on the Normal pdf is that the true variance σ^2 is usually not known. In order to use the sample variance s^2, calculated from the same N values, it is necessary to use a slightly different pdf called the Student's T distribution [3-6]. This is the distribution of the variable $T_\nu = \sqrt{N}\ (\bar{x}-\mu)/s$ where $\nu = N-1$ is the number of degrees of freedom in the computation of s. As ν becomes large the T distribution converges to the Normal pdf. The use of the T distribution to evaluate the confidence inter-

vals for an estimation of the mean is shown graphically in Figure 3-8, where the limits on the value $(\mu-\bar{x})\frac{\sqrt{N}}{s}$ are plotted versus ν for different confidence levels. The ordinate in Figure 3-8 is a measure of the number of standard deviations in the estimate of the mean that are encompassed by each confidence interval. It can be seen that confidence intervals grow dramatically for $\nu < 10$.



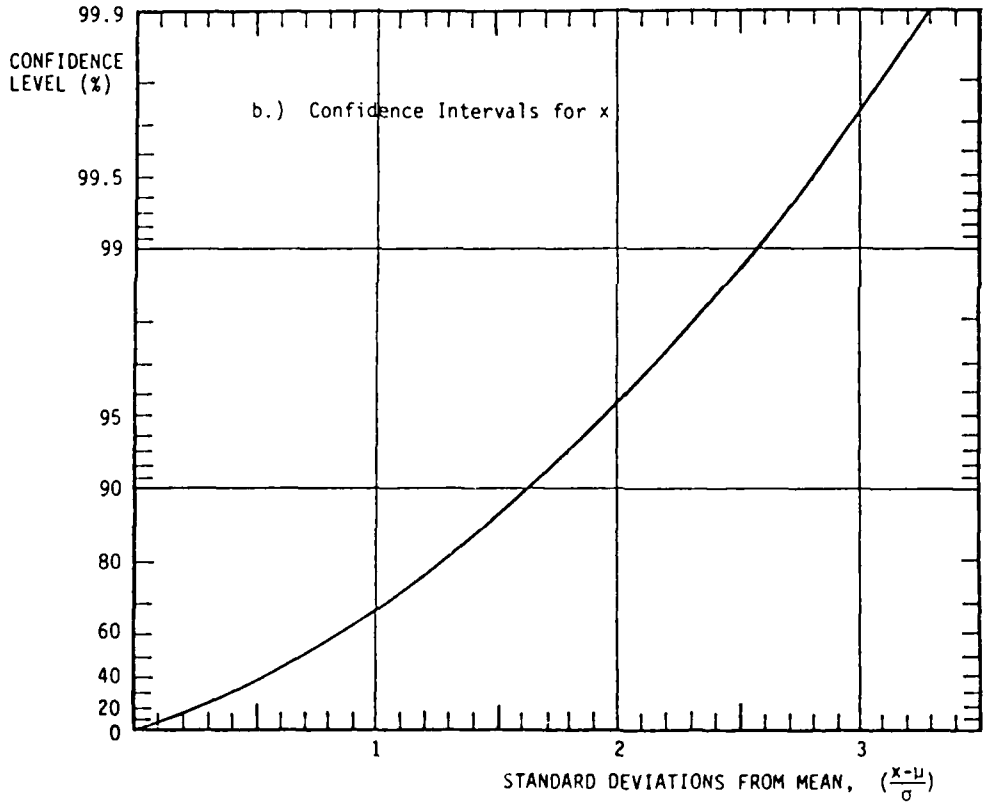


Figure 3-7 Characteristic of a Random Variable x with a Normal Distribution Having a Mean Value μ and a Variance σ^2

Data from Ref. 3-6

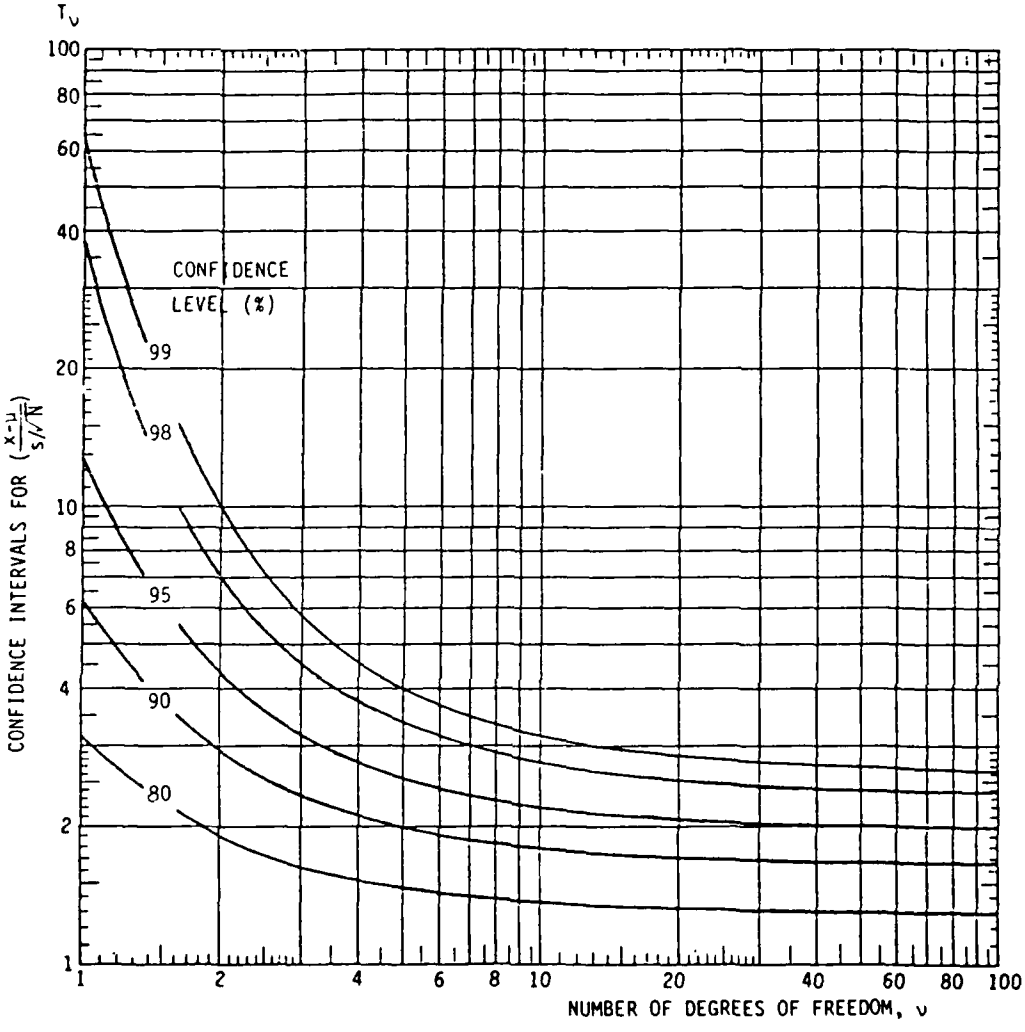


Figure 3-8. Confidence Intervals for the Sampled Values of a Random Variable $T_\nu = \sqrt{N}(\bar{x} - \mu)/s$ with a Student's T Distribution Having $\nu = N - 1$ Degrees of Freedom

Data from Ref. 3-6

SAMPLE VARIANCE: The distribution of the sample variance s^2 obtained from Equation (3-19) is related to the Chi-squared distribution denoted by χ^2_ν where ν is the number of independent samples used to obtain s^2. The sample variance of a set of N values x_n is computed using N samples of $(x_n-\bar{x})^2$. However, only N-1 of these samples are independent if $\bar{x}$ is also computed from the same data because of the constraint:

$$\sum_{n=1}^{N} (x_n - \bar{x}) = 0. \tag{3-29}$$

Therefore a value of $\nu = N-1$ is used.

If a random signal has a variance σ^2 then the sample variance s^2 will have an expected value given by:

$$E[s^2] = \sigma^2 \tag{3-30}$$

and a variance

$$V[s^2] = \frac{2\sigma^4}{N-1} \tag{3-31}$$

The use of the Chi-squared distribution to obtain confidence intervals for the estimation of the variance is shown in Figure 3-9 where the error bounds on σ^2/s^2 are plotted versus ν for different percentages of confidence. As ν becomes large, the Chi-squared distribution also converges to the Normal pdf, however, at a slower rate than the Student's T distribution. For $\nu > 100$ the variable $\sqrt{2N-2}\ (\frac{s}{\sigma} - 1)$ has a Normal distribution with zero mean and unity variance. Then the confidence intervals for this variable can be obtained from the Normal distribution and the confidence intervals for σ^2/s^2 can be computed from these.

SAMPLE PDF: The distribution of the sample pdf is related to the Binomial distribution which is a measure of the probability of a given number of samples N_i in a certain value range w being selected out of a total of N samples. For given value

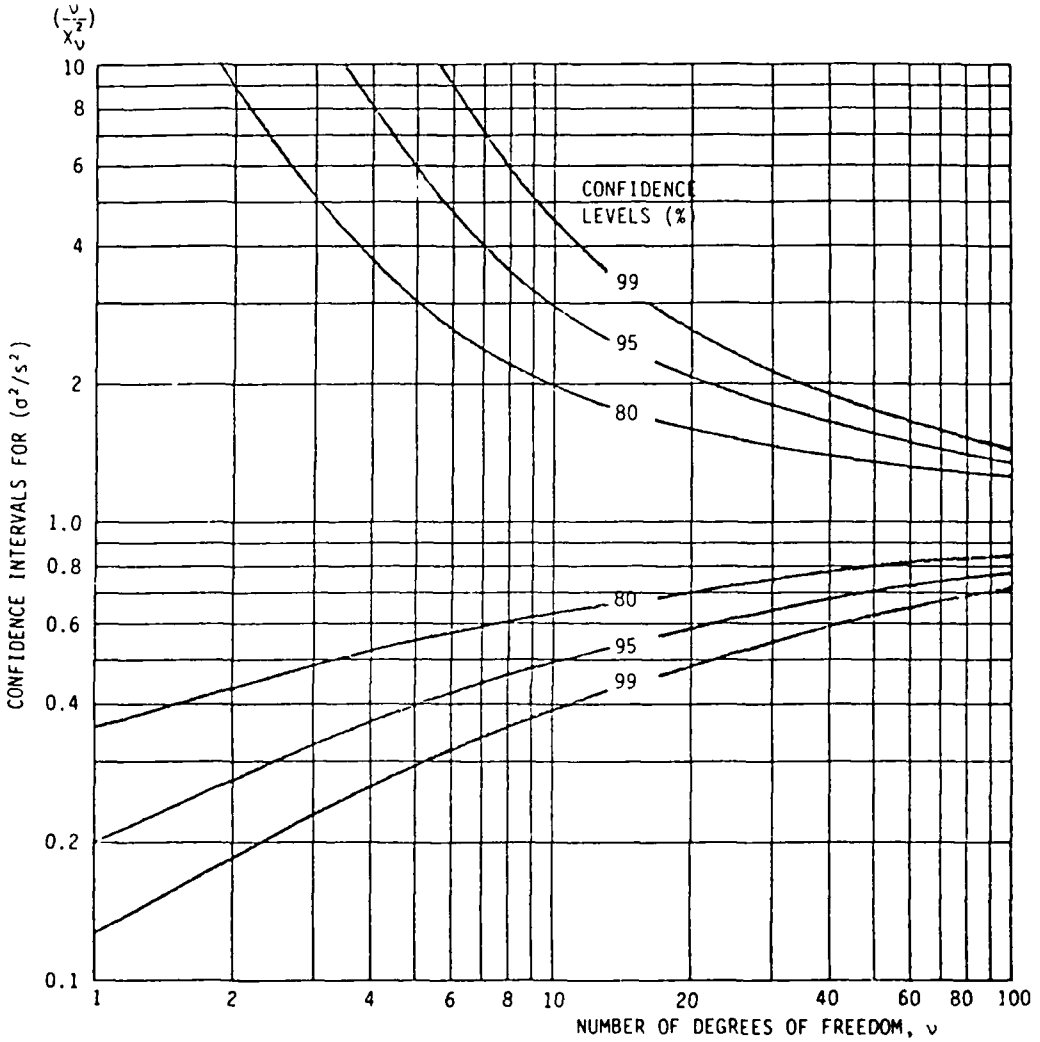


Figure 3-9. Confidence Intervals for the Sampled Values of a Random Variable $\chi^2_\nu = \nu s^2/\sigma^2$ with a Chi-Squared Distribution Having $\nu = N - 1$ Degrees of Freedom

range w the expected value of the pdf magnitude can be approximated by [3-5]:

$$E\left[\frac{N_i}{wN}\right] \approx p(x_i) + \frac{w^2}{24} p''(x_i) \qquad (3\text{-}32)$$

where $p(x_i)$ is the true pdf and $p''(x_i)$ is the second derivative of $p(x_i)$ with respect to x. For the Normal distribution with a variance σ^2 the maximum value of $p''(x_i)$ is approximately $-0.4/\sigma^3$. Equation (3-32) indicates that a bias error exists in the estimation of the pdf which grows proportional to the square of the value interval.

The variance of the pdf magnitude is computed from the variance of the Binomial distribution using the probability N_i/N so that

$$V\left[\frac{N_i}{wN}\right] = \frac{N_i}{w^2N^2}\left(1 - \frac{N_i}{N}\right) \approx \frac{p(x_i)}{wN} \qquad (3\text{-}33)$$

where the approximation is valid for $\frac{N_i}{N} << 1$ which will be true for sufficiently small values of w. This indicates that the variance (or random error) is reduced by increasing the value interval w which conflicts with the reduction of the bias error. Therefore it is necessary to arrive at a compromise between a small variance and a small bias.

One approach to determine the size of the value interval w which will result in reasonably good pdf estimate is to set the sum of the variances of the pdf to a constant value:

$$\sum_{i=1}^{M+1} V\left[\frac{N_i}{wN}\right] = \text{constant.} \qquad (3\text{-}34)$$

This can be evaluated by

$$\sum_{i=1}^{M+1} \frac{p(x_i)}{Nw} = \frac{1}{Nw^2} \qquad (3\text{-}35)$$

From experience it has been found that a good choice for the constant is the squared value of the pdf at two standard deviations from the mean, which for the Normal distribution is $(0.05/\sigma)^2$. Using this result gives

$$\frac{1}{Nw^2} \simeq \left(\frac{0.05}{\sigma}\right)^2$$

or

$$w \simeq \frac{20\sigma}{\sqrt{N}}. \qquad (3\text{-}36)$$

3.2.1.9 Significance Levels

As mentioned above, significance levels are used to determine quantitatively the validity of a hypothesis or assumption about the data being analyzed. For example, suppose it is assumed that a signal x has a mean value of zero. To test this assumption this signal is sampled N times and the sample mean $\bar{x}$ and standard deviation s_x are computed. For a signal that does have a zero mean the sample means will be Normally distributed around zero with $s_{\bar{x}} = \frac{s_x}{\sqrt{N}}$. Then it could be stated that 95% of the sample means will be in the interval $\left[-\frac{2s_x}{\sqrt{N}}, \frac{2s_x}{\sqrt{N}}\right]$.

Now, if the computed value of $\bar{x}$ falls outside this interval it can be stated with a 95% confidence that the signal does <u>not</u> have a zero mean. However, if the computed value of $\bar{x}$ falls inside this interval it cannot be stated with any great certainty that the signal has a zero mean since there are many possible values in the interval. Historically, the level of significance of such a test has been defined by the percentage of the time a true hypothesis is rejected (which is also a measure of the ability of the test to reject a false hypothesis). The level of significance of the example test above is then 5%.

3.2.1.10 Time Averaging vs. Ensemble Averaging

Up to this point in the discussion of the statistical measures of digital data we have not assumed anything about the origin of the data. They could be a set of periodically sampled values of a time series or they could be successive estimations of a statistical parameter evaluated on an ensemble of data sets. Most of the statistical analysis methods discussed in this chapter are based on the principle that the data is an ensemble of independent samples.

The assumption that these same methods apply to a sequence of periodically sampled values of a time series is that of stationarity. A stationary signal is one for which the statistical measures of interest do not change with time. It is possible to define different types of stationarity. For example, a signal can be stationary in its mean value but not in its variance; or a signal can be stationary in its variance but not in its mean value. In general, if the first k moments of the pdf of a signal are stationary, then the signal is said to be stationary to the k^{th} order.

The determination of whether a signal is stationary or not may depend on the length of the time interval in which the signal is observed. For example, if a signal consists of white noise added to a low frequency sine wave, the signal will not be stationary over a time interval that is on the order of the period of the sine wave. However, for time intervals either much shorter or much longer than the period of the sine wave, the signal will appear to be stationary.

Tests for determining whether a signal is stationary over the observed time interval are given in the next Section. Special methods of analyzing non-stationary data are discussed in Section 3.2.5.2.

3.2.1.11 Tests for Stationarity

Two successive sample records x_1 and x_2 of the same signal can be tested for stationarity as follows.

1. Compute the sample means $\overline{x_1}$, $\overline{x_2}$ and the sample variances s_1^2, s_2^2.

2. Compute the value $F = s_1^2/s_2^2$ or $F = s_2^2/s_1^2$, whichever is larger.

3. Compare the value F to one of Fisher's $F_{\nu 1, \nu 2}$ distributions shown in Figure 3-10 where $\nu_1 = N_1 - 1$; $\nu_2 = N_2 - 1$; N_1 and N_2 are the respective sample sizes.

4. If $F > F_{\nu 1, \nu 2}$ for a given confidence level, then a significant difference exists in the variances and the signal has been determined to be non-stationary in its variance with that confidence level.

5. If $F < F_{\nu 1, \nu 2}$ then compute

$$T = \frac{x_1 - x_2}{s_t \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \qquad (3\text{-}37)$$

where

$$s_t = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2} \qquad (3\text{-}38)$$

6. Compare the value of T with one of the Student's T_ν distributions shown in Figure 3-8 where $\nu = N_1 + N_2 - 2$.

7. If $T > T_\nu$ for a given confidence level then a significant difference exists in the mean and the signal has been determined to be non-stationary in its mean with that level of confidence.

8. If $T < T_\nu$ and $F < F_{\nu 1, \nu 2}$ the signal can be assumed to be stationary over the duration of the two sample records. The level of significance of such a test is (100% - percent confidence level).

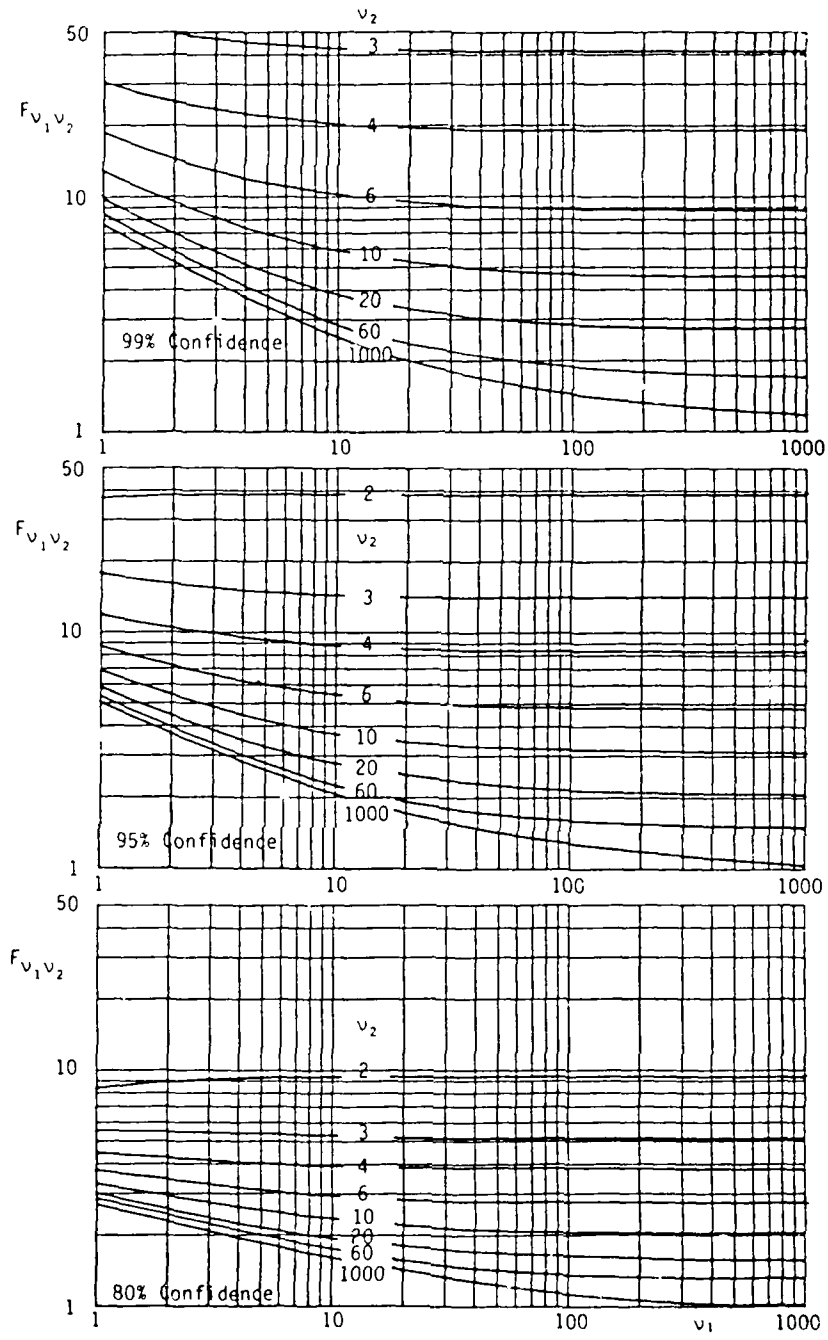


Figure 3-10. Confidence Intervals for the Sampled Values of the Random Variable $F_{\nu_1\nu_2}$ with a Fisher's Distribution Having ν_1 and ν_2 Degrees of Freedom

Data from Ref. [illegible]

3.2.2 Autocorrelation Function

In this section we present a direct method of estimating the autocorrelation function (acf) of a single time series (or signal). The acf is a measure of the relationship between the values of a signal at two different times, t_1 and t_2. This relationship is measured by the average of the product of the two signal values, $x(t_1)$ and $x(t_2)$.

If the signal is stationary the acf will depend only on the time difference (t_2-t_1), and not on the absolute times t_1 and t_2. In this case the acf can be averaged over time within one sample record. If the signal is non-stationary, the acf may also depend on the absolute time. In this case time averaging is not possible. For repetitive, non-stationary signals the acf can be averaged over an ensemble of sample records which have been obtained during the same time interval relative to some repeated characteristic in the repeated signal. Accurate synchronization of the sample records is important.

The time averaged acf of a signal x(t) can be estimated by the sample acf, denoted R_x, of a set of N values x_n sampled periodically with a time interval τ as follows:

$$R_x(m\tau) = \frac{1}{N-m} \sum_{n=1}^{N-m} x_n x_{n+m} \quad ; \quad m < n \qquad (3\text{-}39)$$

This direct method of estimating the acf is generally used only for small values of m, because it requires a large number of computations. When m is large a more general procedure can be used which is based on the relationship of the acf to the frequency spectrum of a signal. This procedure is outlined in Section 3.2.3.2.

The following are some general properties of the acf.

1. For stationary signals the acf for a positive time delay is equal to that for a negative time delay.

$$R_x(m\tau) = R_x(-m\tau) \tag{3-40}$$

2. The acf at zero time delay is equal to the mean square value of the signal.

$$R_x(0) = \overline{x^2} \tag{3-41}$$

3. The acf at zero time delay is greater or equal in magnitude to the acf at any time delay.

$$R_x(0) \geq |R_x(m\tau)| \tag{3-42}$$

4. For non-periodic signals the limit of the acf for large time delays is the square of the mean value.

$$\lim_{m\to\infty} [R_x(m\tau)] = (\overline{x})^2 \tag{3-43}$$

Figure 3-11 presents the acf of some typical signals. The acf of a periodic signal is also periodic. The acf of Gaussian (white) noise with zero mean value is non-zero only at m = 0 (meaning there is no relationship between signal values at different times). However, for band limited white noise, the acf is non-zero over a time period which is approximately equal to the inverse of the frequency bandwidth.

When the acf is to be evaluated for a signal having a non-zero mean, it is computationally desirable to remove the mean value first before evaluating Equation (3-39). The resulting quantity is called the autocovariance function (acvf), denoted by C_x, which is defined by:

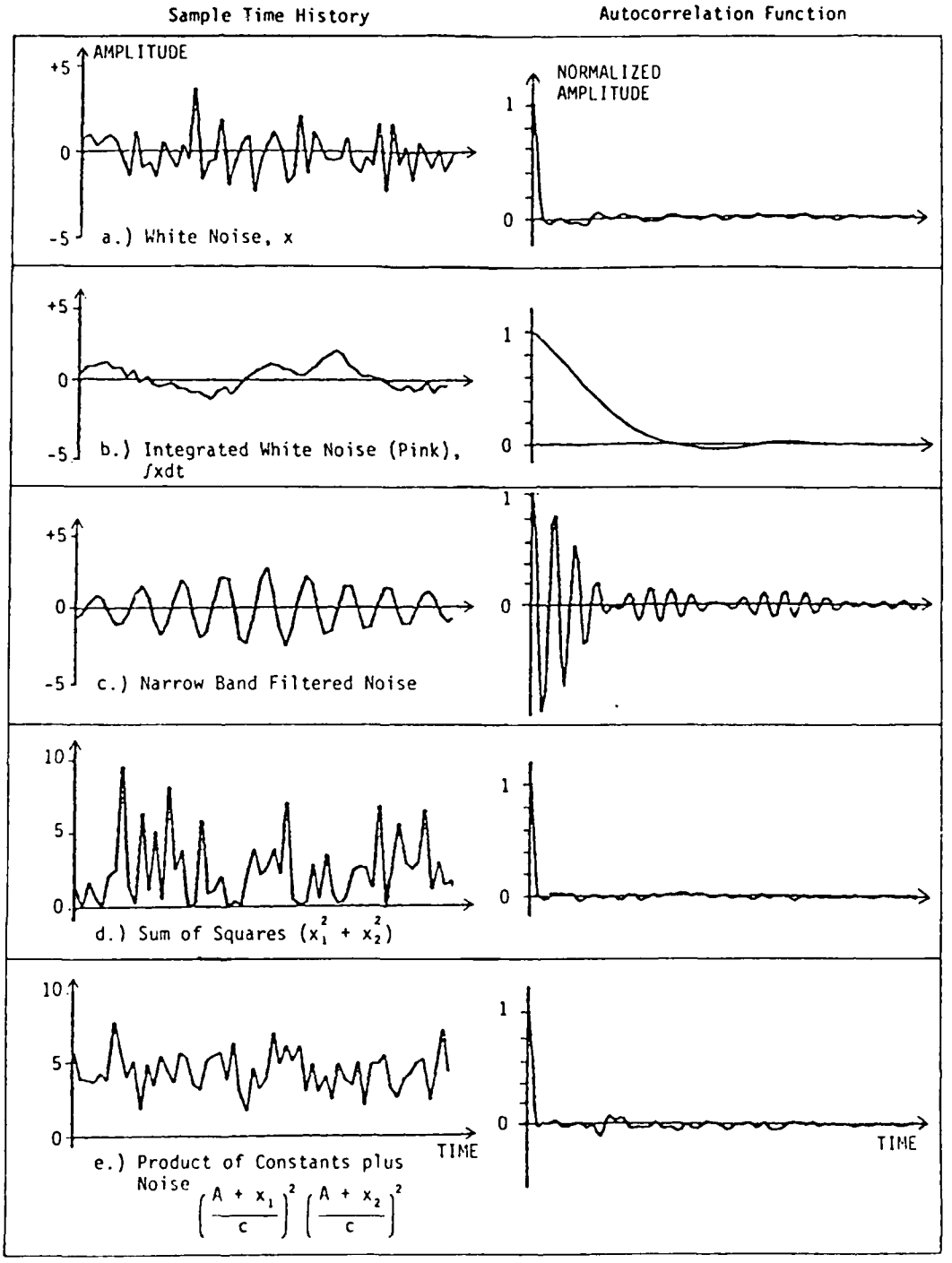


Figure 3-11. Autocorrelation Function of Some Typical Signals

$$C_x(m\tau) = \frac{1}{N-m} \sum_{n=1}^{N-m} (x_n - \bar{x})(x_{n+m} - \bar{x}) \; ; \quad m < N \qquad (3\text{-}44)$$

The acf and acvf are related by:

$$R_x(m\tau) = C_x(m\tau) + (\bar{x})^2 \qquad (3\text{-}45)$$

The acf is often represented by the normalized correlation function ρ which is defined by:

$$\rho_{xx}(m\tau) = \frac{C_x(m\tau)}{C_x(0)} \qquad (3\text{-}46)$$

The magnitude of the normalized correlation function is independent of the mean value and variance of the data and is a direct measure of the correlation of the signal over a time interval. It follows from Equation (3-42) that

$$\rho_x(0) = 1 \qquad (3\text{-}47)$$

$$|\rho_x(m\tau)| \leq 1. \qquad (3\text{-}48)$$

3.2.2.1 Statistical Errors in the ACF Estimate

Equation (3-39) is an unbiased estimate of the acf in that the expected value of the computation is equal to the true value of the acf [3-4]. The difference between the expected value and the true value of a measure is called the bias error, B. Therefore,

$$B[R_x(m\tau)] = 0 \tag{3-49}$$

The variance of the acf estimate around the expected value is often called the random error, V. The variance in the acf computation is difficult to evaluate in general because it depends on the specific form of the true acf [3-4]:

$$V[R_x(m\tau)] = \frac{1}{(N-m)^2} \sum_{i=-N+m}^{N-m} \left[(N-m-|i|) \left(R_x^2(i\tau) + R_x([m+i]\tau)\, R_x([m-i]\tau) \right) \right] ; \quad m \geq 0 \tag{3-50}$$

The values of the acf on the right hand side of Equation (3-50) should be those of the true acf for the equation to be theoretically correct. However, the variance can be estimated by using the computed acf.

Equation (3-50) indicates that as $m \to N$ the variance becomes very large. This has led some statisticians to suggest a modified estimate of the acf given by:

$$R_x^b(m\tau) = \frac{1}{N} \sum_{n=1}^{N-m} x_n x_{n+m} \tag{3-51}$$

This is a biased estimate of the acf since

$$B[R_x^b(m\tau)] = \left(\frac{m}{N}\right) R_x(m\tau) \tag{3-52}$$

However, the variance of the estimate is reduced in value to the form of:

$$V[R_x^b(m\tau)] = \frac{1}{(N)^2} \sum_{i=-N+m}^{N-m} [(N-m-|i|)\ (R_x^2(i\tau) + R_x([m+i]\tau)\ R_x([m-i]\tau))] \quad ; \quad m > 0 \tag{3-53}$$

The overall statistical error of the biased and unbiased estimates of the acf can be compared using the mean square error (mse) which is the sum of the variance and the squared bias error

$$mse = V + B^2 \tag{3-54}$$

In order to evaluate the mse of the acf estimate it is necessary to assume a form of the true acf. A convenient form to assume is a decaying exponent of the form $Ae^{-\alpha m/N}$ $(m > 0)$ where α determines the rate of decay as m increases. This form of the acf is also useful for estimating an upper bound on the error of an acf estimate where α and A are chosen such that

$$R_x(m\tau) \leq Ae^{-\alpha m/N} \quad ; \quad m > 0 \tag{3-55}$$

Using this form of the acf Reference [3-4] derives expressions for the mse of the biased and unbiased acf estimates as follows:

$$mse[R_x(m\tau)] \simeq \frac{A^2}{\alpha} \left(\frac{N}{N-m}\right) \left[1 + e^{-2\alpha m/N}\left(1 + \frac{2\alpha m}{N}\right)\right]$$

$$mse[R_x^b(m\tau)] \simeq \frac{A^2}{\alpha} \left[1 + e^{-2\alpha m/N}\left(1 + \frac{2\alpha m}{N}\right)\right] + A^2 \left(\frac{m}{N}\right)^2 e^{-2\alpha m/N} \quad ; \quad m > 0 \tag{3-56}$$

The values of these mse's are plotted in Figure 3-12a for $\alpha = 4$. The mse of the biased estimate of the acf is considerably smaller for all values of $m > 0$.

This result suggests that the biased acf estimate using Equation (3-51) is to be preferred. This assumes that all of the sampled signal values are available at the same time when the acf is computed. However, there are cases when the number of data samples is larger than that which can be stored at one time (this is especially true when the processing is done in real time). When the data storage size is the limiting factor, then the acf can be computed by dividing the total number of samples into M records of length N each, and averaging the ensemble of acf's computed from the individual records. The mse expression for the averaged acf, denoted by $\overline{R_x}(m_T)$, are given by:

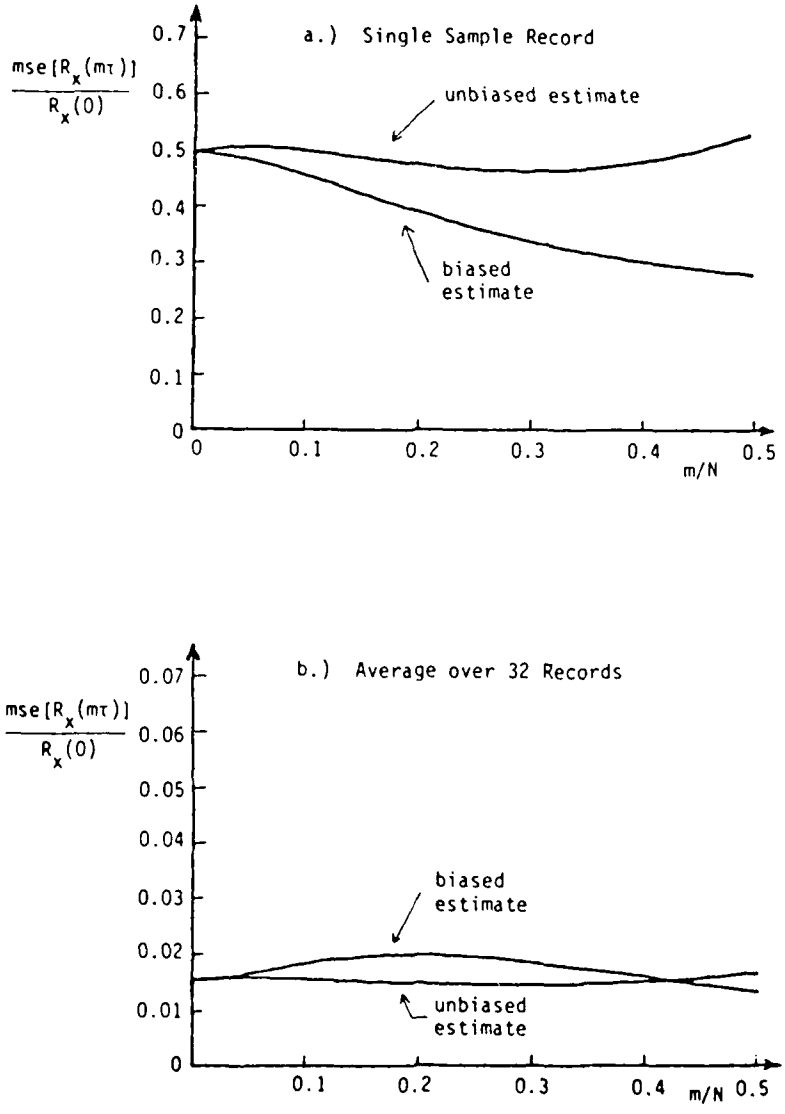


Figure 3-12. Normalized Mean Square Error Bound of Autocorrelation Function Estimate

$$\text{mse}\left(\overline{R_x}(m\tau)\right) = \frac{A^2}{M\alpha}\left(\frac{N}{N-m}\right)\left[1 + e^{-2\alpha m/N}\left(1 + \frac{2\alpha m}{N}\right)\right]$$

$$\text{mse}\left[\overline{R_x^b}(m\tau)\right] = \frac{A^2}{M\alpha}\left[1 + e^{-2\alpha m/N}\left(1 + \frac{2\alpha m}{N}\right)\right] + A^2\left(\frac{m}{N}\right)^2 e^{-2\alpha m/N} \quad ; \quad m > 0 \tag{3-57}$$

By averaging over M records, the variances of the acf estimates are reduced but the bias is not. The values of the mse's for $\alpha = 4$ and $M = 32$ are plotted in Figure 3-12b. Here it is seen that the unbiased acf estimate is preferred.

This is an illustration of a more general principle that can be followed in selecting the best statistical estimator. If it is possible to perform ensemble averages over a number of individual time records, then an unbiased estimator should be used to compute the desired parameter. However, when only a single time record is available, the unbiased estimator may not be the one which has the smallest mean square error. Unfortunately, at this time, there is no method available for deriving the best acf estimator which minimizes the mean square error using a single data record.

3.2.3 Power Spectral Density Function

In this section we present two methods of calculating the power spectral density function (psd) for a set of periodically sampled values x_n obtained from a continuous signal $x(t)$. The psd is a measure of the frequency distribution of the mean square value of the signal. The psd $G_x(f)$ for a continuous signal can be defined in terms of analog filtering as the mean-square output of a narrow band filter divided by the bandwidth of the filter in the limit as the bandwidth goes to zero [3-5]:

$$G_x(f) = \lim_{\delta f \to 0} \left[\frac{\overline{x^2(t,f,\delta f)}}{\delta f}\right] \qquad (3\text{-}58)$$

where $x(t,f,\delta f)$ is the output of a bandpass filter of width δf centered at frequency f (cycles per second, or Hertz).

There are two approaches to calculating the psd. The first, which will be called the direct method, involves the computation of the Fourier Transform of the sampled data, from which the psd is estimated. In general, this approach attempts to minimize the number of necessary computations and is typically used in digital signal processing equipment which does spectral analysis in "real time," that is, the data are processed while they are received without the need of temporary bulk storage of the data. The speed of computation is achieved at the expense of introducing some statistical errors. Therefore, the direct method is more applicable to stationary signals of sufficient duration to allow for averaging over a large number of sample records.

The second approach to calculating the psd, which will be called the indirect method, is based on the relationship of the psd to the autocorrelation function (acf). This approach involves a preliminary calculation of the acf from which the psd is computed. Although this requires some additional computation

relative to the direct method, the statistical properties of the indirect method are better understood from a theoretical point of view. Therefore, the indirect method is more applicable to signals of limited duration from which it is desired to obtain the psd with the minimum statistical error.

3.2.3.1 Direct Method

The direct method of calculating the psd of a finite number of sample values x_n is based on the principles of Fourier Transforms (FT) of periodic signals. A signal with period T (seconds) can be constructed from the sum of a set of sinusoidal signals at evenly spaced frequency intervals $\Delta f = 1/T$. The number of sinusoidal signals required depends on the maximum frequency content of the periodic signal. A digital representation of the periodic signal is obtained by sampling it N times at intervals τ (seconds), such that $T = N\tau$. It is assumed that proper conditioning of the signal has been done so that the maximum frequency content f_{max} is less than one half the sample rate SR (see Section 3.1.1.1):

$$f_{max} < \frac{1}{2} SR = \frac{1}{2\tau} = \frac{N}{2T} \qquad (3\text{-}59)$$

Therefore, the signal has at most $(\frac{N}{2} - 1)$ frequency components.

In this case the sampled values x_n can be constructed by the following series:

$$x_n = \bar{x} + \sum_{k=1}^{K} A_k \cos\left(\frac{2\pi kn}{N}\right) + \sum_{k=1}^{K} B_k \sin\left(\frac{2\pi kn}{N}\right) \qquad (3\text{-}60)$$

where $K = \frac{N}{2} - 1$,

$$\bar{x} = \frac{1}{N} \sum_{n=1}^{N} x_n ; \text{ the mean value,}$$

$$A_k = \frac{2}{N} \sum_{n=1}^{N} x_n \cos\left(\frac{2\pi kn}{N}\right),$$

$$B_k = \frac{2}{N} \sum_{n=1}^{N} x_n \sin\left(\frac{2\pi kn}{N}\right). \qquad (3\text{-}61)$$

The Discrete Fourier Transform (DFT) of the sample values x_n is defined at frequency $f_k = k/T$ by the complex function,

$$X_T(f_k) = \frac{T}{2}(A_k - jB_k) \; ; \quad k = 1, 2, \ldots \frac{N}{2} - 1 \qquad (3\text{-}62)$$

$$X_T(0) = T\bar{x}$$

where $j=\sqrt{-1}$. The psd $G_x(f_k)$ is given by

$$G_x(f_k) = \frac{2}{T} |X_T(f_k)|^2 \; ; \quad k = 0, 1, \ldots \frac{N}{2} - 1. \qquad (3\text{-}63)$$

Outside the time interval [0,T] the series in Equation (3-60) will repeat in value with period T. If x(t) is periodic with period T, the psd in Equation (3-63) will be exact. However, if x(t) is not periodic (which is generally the case), Equation (3-63) can still be used to estimate the psd of the signal in the time interval [0,T]. This estimate can then be averaged over an ensemble of sample records to improve the statistical accuracy of the estimate.

Figure 3-13 presents the psd of some typical signals. The psd of a periodic signal is a line spectrum consisting only of those frequencies which are integer multiples of the inverse of the signal period. The psd of Gaussian (white) noise is flat with equal value at all frequencies. However, for band limited white noise, the psd has the shape of the transfer function of the band limiting filter.

The direct method of calculating the psd has become useful in practice because of the development of a computationally efficient algorithm for evaluating the DFT, called the Fast Fourier Transform (FFT), which significantly reduces the required computation time. The implementation of this method is outlined in the following paragraphs.

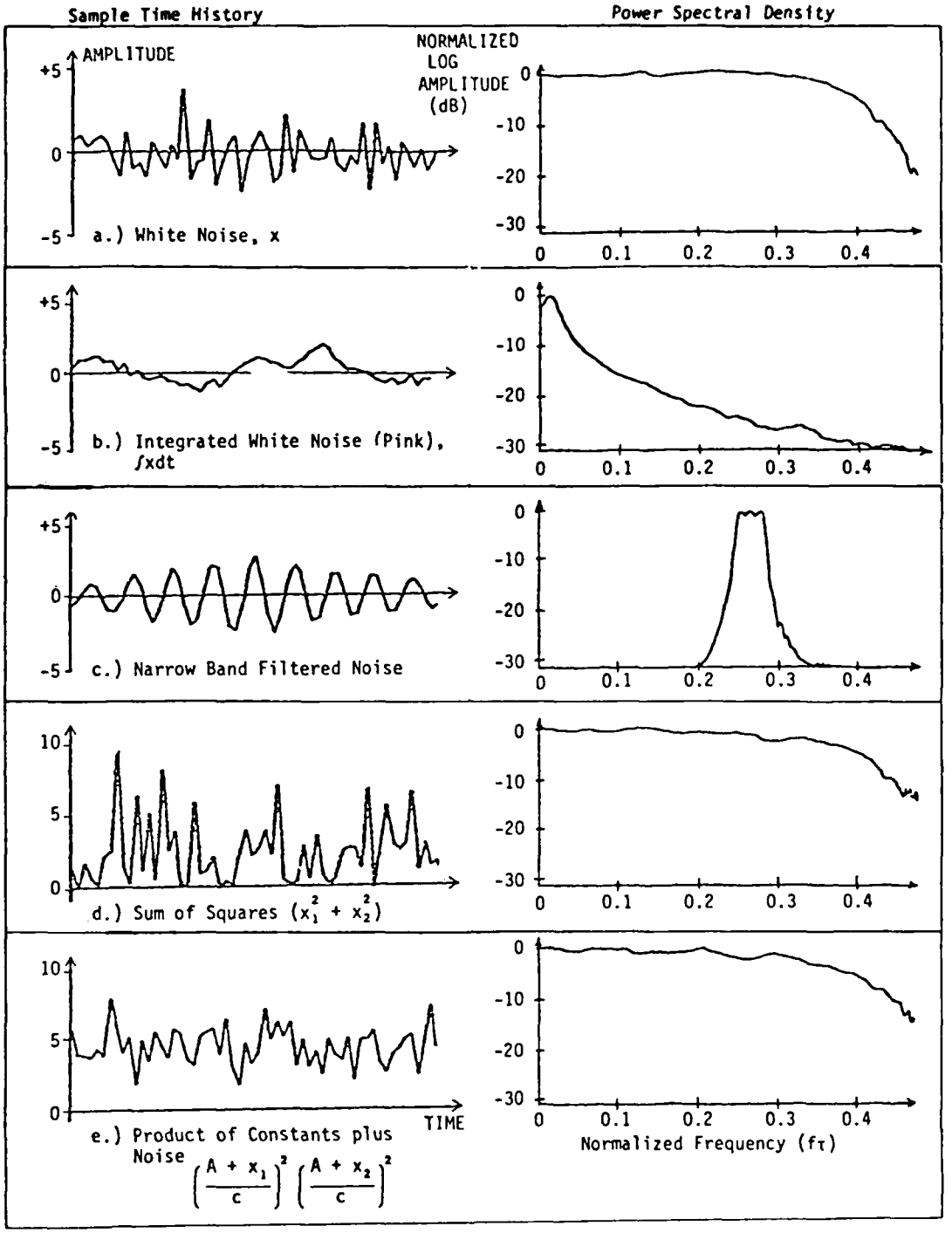


Figure 3-13. Power Spectral Density of Some Typical Signals

FAST FOURIER TRANSFORM: The evaluation of the Fourier Transform of a signal by the procedure indicated in Equation (3-60) is not frequently used in practice because it requires approximately N^2 mathematical operations. However, a more efficient procedure of evaluating the FT is presented in a paper by Cooley and Tukey [3-8] which has become known as the Fast Fourier Transform (FFT). The direct evaluation of the psd using the FFT was first proposed by Welch [3-9]. The purpose of these paragraphs is not to derive this procedure in detail, but to indicate the manner in which it can be used to compute the psd. More detailed information on the FFT and related procedures can be found in References [3-4, 3-5, 3-8, 3-9].

The FT of a sample record can be efficiently evaluated if the number of samples N is a power of 2; that is,

$$N = 2^p. \tag{3-64}$$

This is because the FT of a set of N numbers can be evaluated by dividing the data in half, computing the FT of each half, and summing the results in a particular manner. This will require only $N^2/2$ mathematical operations. However, if $N = 2^p$ this subdivision of the data can be repeated p times until the data is divided into N/2 pairs. Then the FT of each pair requiring 4 operations each can be evaluated for the N/2 pairs. By repeating this process for each of the p divisions, the FT of the complete data set can be evaluated with approximately 2Np operations. Therefore, a considerable time saving is achieved if N is large.

The format in which many FFT computer routines are written uses a complex array z_n; n = 1, 2, ..., N; to store the data. The sampled values x_n are stored in the array for input to the FFT in the format:

$$z_n = x_n + j0 \; ; \quad n = 1, 2, \ldots, N. \tag{3-65}$$

After the FFT operates on the input data the FT $X_T(f_K)$ is stored for output in the format:

$$Z_k = \frac{1}{\tau} X_T(f_{k-1}) \; ; \quad k = 2, 3, \ldots, \frac{N}{2} \tag{3-66}$$

with $Z_1 = N\bar{x} + j0,$

$$Z_{\frac{N}{2}+1} = \sum_{n=1}^{N} (-1)^n x_n,$$

$$Z_k = \frac{1}{\tau} X_T^* (f_{N-k+1}) \; ; \quad k = \frac{N}{2} + 2, \ldots, N. \tag{3-67}$$

where X* refers to the complex conjugate (sign of imaginary part reversed).

This format is used because it is convenient for the simultaneous computation with two signals which is discussed in Section 3.3.3. Other FFT formats do exist, such as one in which the FT is also scaled by N, or one in which the sign of the imaginary part is reversed, and therefore caution must be used when first implementing an analysis program which uses an FFT to insure that the proper calibration has been made (discussed later in this section).

Assuming the data are in the format of Equation (3-65), the psd $G_x(f_k)$ at frequency $f_k = k/T$ can then be computed by

$$G_x(f_k) = \frac{2\tau}{N} |Z_{k+1}|^2 \; ; \quad k = 1, 2, \ldots, (\frac{N}{2} - 1) \tag{3-68}$$

and

$$G_x(0) = \frac{2\tau}{N} (Z_1)^2. \tag{3-69}$$

DATA PREPARATION: Before using the FFT for the computation of the psd of a sample record, some considerations must be made concerning the preparation of the data in order to minimize the errors of the computation. Section 3.1.1.1 considers the effects of aliasing on the frequency estimation of a signal sampled at time interval τ if the maximum frequency content f_{max} is not limited by $f_{max} < 1/2\tau$. It is assumed here that this condition has been met.

An additional source of error is introduced by the use of a finite duration sample record of length T in the FFT computation. As mentioned above, the FFT is based on the series representation of a periodic signal with period T. Whether or not the signal being analyzed is periodic, the FFT computes the spectrum of a signal which consists of the periodic sequence of the sample record. In general there is a discontinuity at the end of one sequence and the beginning of the next sequence. This discontinuity has the effect of adding noise to the psd since it represents a very high frequency component which is aliased throughout the frequency range of the psd.

Another way of looking at this phenomenon is to consider the sampled record as the product of the continuous signal x(t) and a sampling "window" $w_T(t)$ defined by:

$$w_T(t \leq 0) = 0,$$

$$w_T(0 < t \leq T) = 1,$$

$$w_T(t > T) = 0. \qquad (3\text{-}70)$$

The FT of the window is the effective filter shape of each frequency calculation in the FFT. The filter shape obtained using the window $w_T(t)$ is shown in Figure 3-14. It is obvious that

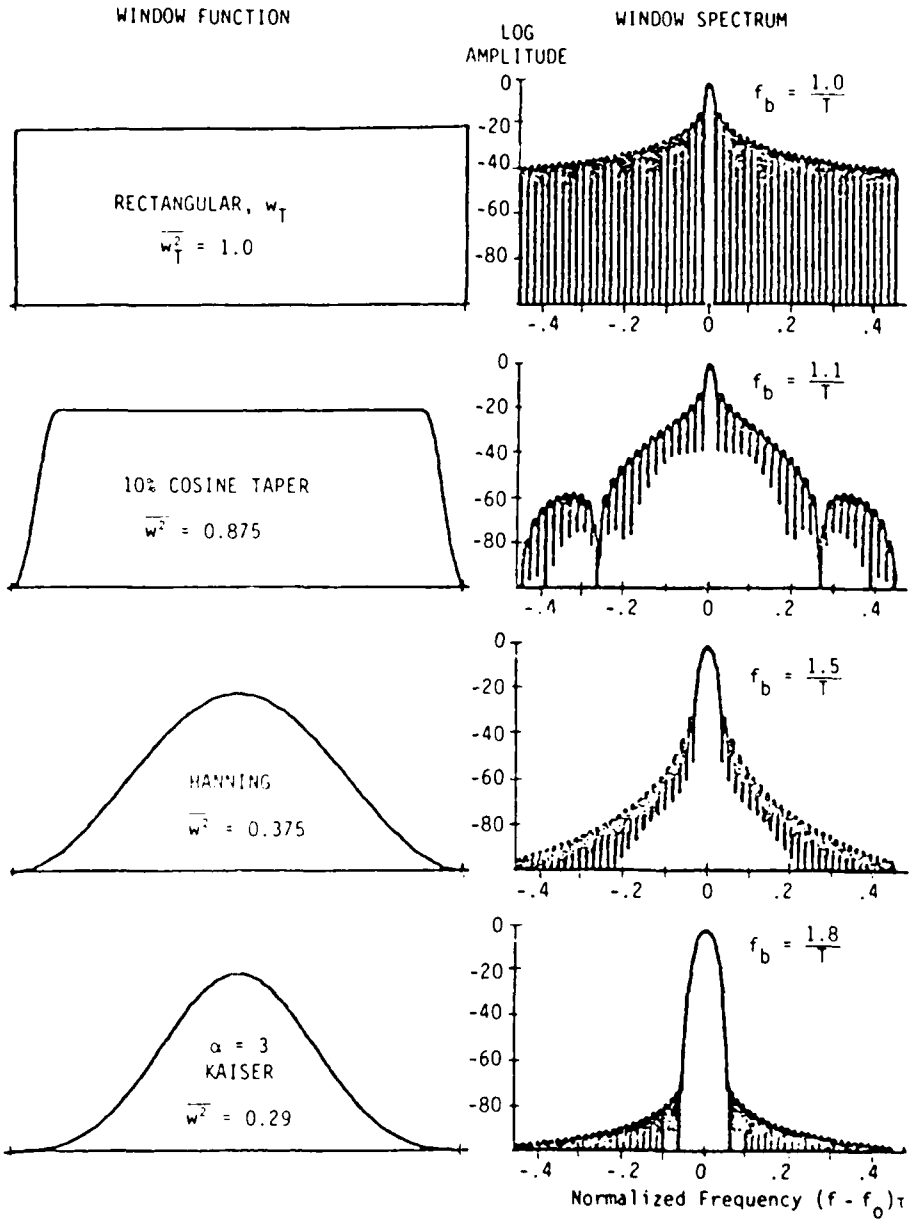


Figure 3-14. Characteristics of Selected Window Functions

this is not a satisfactory window since the relatively large amplitudes of the side lobes result in the contamination of the computed psd.

In order to overcome this problem, the sample record can be multiplied by a smoother window function. Several commonly used window functions are also shown in Figure 3-14, along with their associated filter shapes [3-18]. The mean square value $\overline{w^2}$ for each window function is the average amount the psd amplitude is reduced by the use of the window. Therefore, the computed psd must be multiplied by the inverse of this value in order to maintain proper scaling of the results.

Before applying a smoothed window to the sample record, it is necessary to remove any significant signal content below the frequency $f_1 = 1/T$. In particular, a large mean value or slope in the data can affect the low frequency end of the FFT. Examples of this are shown in Figure 3-15. If possible these low frequency components in the signal should be removed by high pass filtering the analog signal before it is sampled. This filtering can be done in conjunction with the anti-aliasing filtering. In this case it is also necessary to be sure that the A/D converter does not introduce a dc bias in the signal amplitude (that is, the signal ground is converted to the number zero).

The low frequency content of the signal can also be removed digitally, although this increases the computation time of the psd. Several methods of doing this are suggested in the literature, such as linear regression analysis [3-5] as discussed in Section 3.3.1.5, and digital filtering [3-3] as discussed in Section 3.2.4. However, in keeping with the intent of this section, which is to keep the required computation time to a minimum, the following procedure for removing the mean value and average slope of a sample record is recommended:

1. Divide the sample record in four segments of equal length and compute the mean value of each quarter record; $\bar{x}_1$, $\bar{x}_2$, $\bar{x}_3$, $\bar{x}_4$.

2. Compute the coefficients of a best-fit straight line through these mean values represented by

$$y_n = an + b \; ; \quad n = 1, 2, \ldots, N \tag{3-71}$$

where

$$a = \frac{2}{N}\left[\frac{2}{5}\sum_{i=1}^{4} i\,\bar{x}_i - \sum_{i=1}^{4}\bar{x}_i\right],$$

$$b = \frac{1}{4}\sum_{i=1}^{4}\bar{x}_i - \frac{3}{2}a. \tag{3-72}$$

3. Generate a new sample record z_n by subtracting the value of this line from each sample value:

$$z_n = x_n - an - b \; ; \quad n = 1, 2, \ldots, N. \tag{3-73}$$

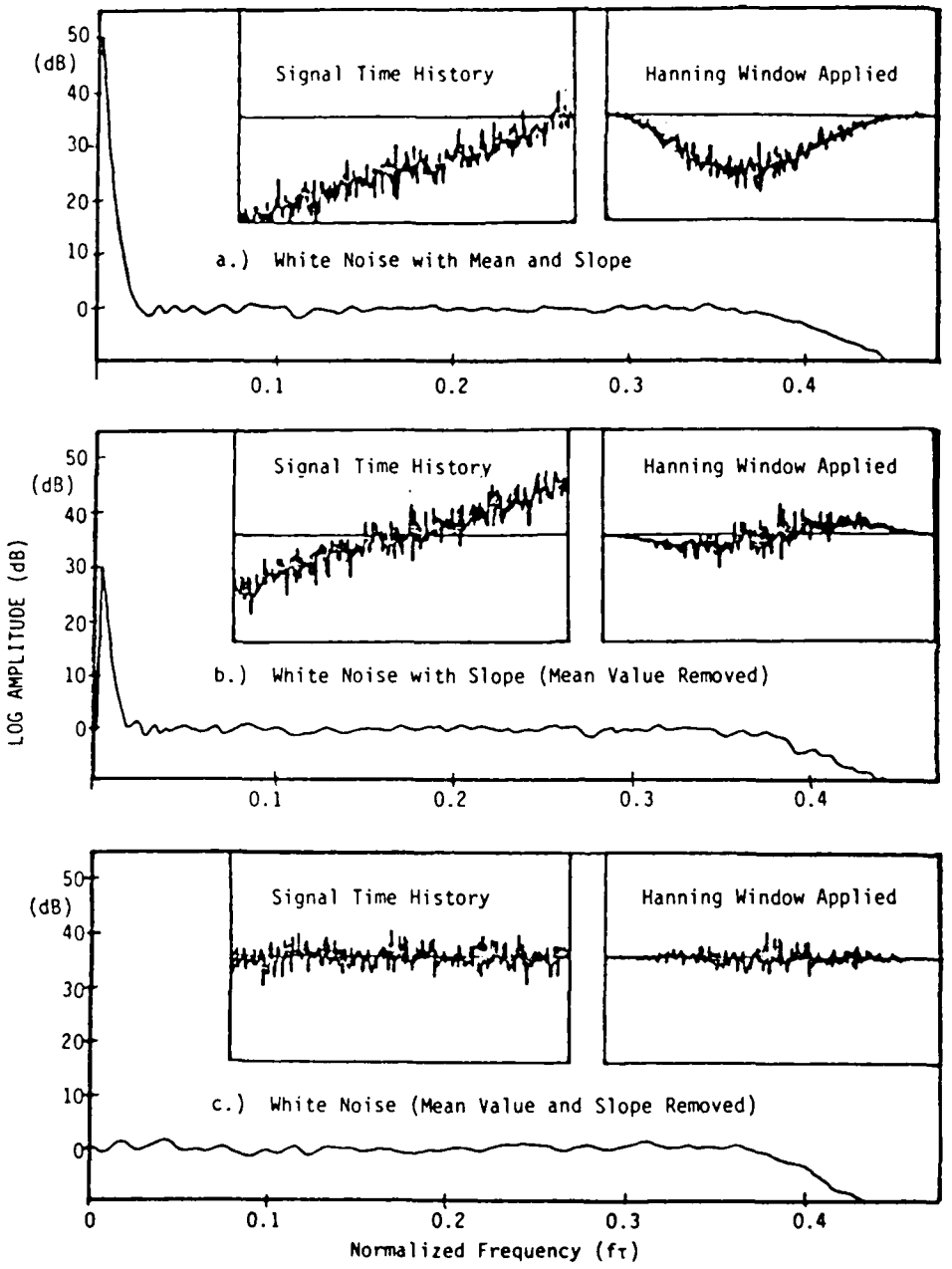


Figure 3-15. Example of Errors in the Power Spectral Density Estimate of a Signal with a Non-Zero Mean Value and Slope

3.2.3.2 Indirect Method

The indirect method of computing the psd is based on its relationship to the autocorrelation function (acf). The psd $G_x(f_k)$ and the acf $R_x(m\tau)$ are related by the Fourier series:

$$G_x(f_k) = 2\tau\, R_x(0) + 4\tau \sum_{m=1}^{M} R_x(m\tau) \cos\left(2\pi \frac{km}{N}\right)$$

$$R_x(m\tau) = \frac{1}{2N\tau} G_x(0) + \frac{1}{N\tau} \sum_{k=1}^{K} G_x(f_k) \cos\left(2\pi \frac{km}{N}\right) \qquad (3\text{-}74)$$

where N is the number of data values sampled with period τ, $f_k = k/N\tau$, and the summation limits are $M = \frac{N}{2} - 1$ and $K = \frac{N}{2} - 1$.

The acf can be evaluated by one of two methods. The first is the direct summation method outlined in Section 3.2.2, Equation (3-39). This method is computationally slow because it requires approximately $\frac{1}{2}$ MN operations. The second method is to use the FFT to compute the acf as follows:

1. Extend the sample record to a length of 2N values by adding N zeroes to the end of the sampled data. Do not window the data.

2. Compute the sample psd with a 2N point FFT giving a frequency interval $\Delta f = \frac{1}{2N\tau}$.

3. Average the sample psd over ensembles of sample records.

4. Compute the acf with a 2N point inverse FFT of the average psd giving correlation time delays from $-N\tau$ to $(N-1)\tau$.

This acf will not, in general, go to zero at $t = (N-1)\tau$ so that a window function must be applied to it in order to compute an accurate psd of the signal. The same choice of window functions exist as shown in Figure 3-14 for sample windows, except that the window function must be centered at $t = 0$. A window of length $T = N\tau$ will then compress the acf to time delays from $-\frac{N}{2}\tau$ to $(\frac{N}{2} - 1)\tau$. The psd can then be computed with a N point FT of the windowed acf, giving a frequency interval $\Delta f = \frac{1}{N\tau}$.

The windowing of the acf is essentially a smoothing operation on the psd. If the Hanning window is used on the acf, a simple smoothing process can be done directly on the sample psd to achieve the same result with fewer calculations. If a sample psd is computed from a 2N point FFT then the first N points represent the sample psd values $G_k = G_x(\frac{k-1}{2T})$; $k = 1, 2, \ldots N$. A smoothed psd, denoted by $\overline{G_x}$ can be computed from these psd values by

$$\overline{G_x}(f_k) = \frac{1}{4} G_{2k} + \frac{1}{2} G_{2k+1} + \frac{1}{4} G_{2k+2}$$

$$\overline{G_x}(0) = G_1$$

$$f_k = k/T \; ; \quad k = 1, 2, \ldots, \frac{N}{2} - 1 \qquad (3\text{-}75)$$

3.2.3.3 Calibration and Units of the PSD

If a signal has a unit of measure U, then the psd has the units of U^2/Hz (Hz = cycles/sec) since it is a measure of the signal power per unit of frequency. This differs from the mean-square output of a filter of bandwidth δf which is given by the product of the average psd and the bandwidth of the filter according to Equation (3-58).

A convenient method of calibrating the psd computed from an FFT of the sampled signal is to use a white noise voltage signal of known bandwidth and known mean square value. This can be done by analog filtering the signal from a white noise generator with a low pass filter having a roll-off rate of at least 24 dB/octave above the upper cutoff frequency f_u. The mean square value $\overline{x^2}$ can be either measured with an analog rms meter or computed from the sampled data using Equation (3-18). (Better results are obtained if the mean value is first taken out of the data.) The psd is then approximately

$$G_x(f) \simeq \frac{\overline{x}^2}{f_u} \; ; \quad f < f_u. \qquad (3\text{-}76)$$

The computed psd can be calibrated to this value with units of $(volts)^2/Hz$. The units of the psd of a measured signal can be obtained by multiplying the computed psd by the square of the sensitivity of the instrumentation system in (units/volt).

It is a good idea to repeat the calibration test for different sample rates, frequency ranges, and sample record lengths to be sure that the analysis system is properly scaled by those parameters. The value of the psd should not change when the sample rate, frequency range, or record length is changed as long as the input signal is properly low pass filtered to prevent aliasing.

3.2.3.4 Statistical Errors in the PSD Estimate

The computation of the sample psd from the FFT of a digitally sampled signal involves the summation of squared quantities in the form of Equation (3-63). In most cases the FT coefficients can be considered to be random variables with a Normal distribution [3-4]. It can then be shown that the variable $[G_x(f_k)/\tau\sigma_x^2]$ has a Chi-squared distribution χ_2^2 with two degrees of freedom, where σ_x^2 is the variance of the signal and τ is the sample interval. The characteristics of the χ_2^2 distribution can then be used to determine the statistical errors in the psd estimation for both the direct and indirect methods.

For a white noise signal, the FFT will give an unbiased estimate of the psd. However, if the signal is not white noise (that is the spectrum is not flat) the psd will have a bias which is of the form [3-4, 3-5]:

$$B[G_x(f_k)] = \frac{f_b^2 \, G_x''(f_k)}{24} \qquad (3\text{-}77)$$

where f_b is the effective filter bandwidtn of the FFT (related to the time window used) and G_x'' is the second derivative of the psd with respect to frequency (proportional to the curvature of the spectrum). This bias error results from the fact that the psd estimate at a particular frequency is an average of the true psd

over the filter bandwidth. Values of f_b are given in Figure 3-14 for typical time window functions. The value of G_x'' can be approximated from the sample psd by

$$G_x''(f_k) \approx \frac{G_x(f_{k-1}) - 2G_x(f_k) + G_x(f_{k+1})}{(1/T)^2} \qquad (3\text{-}78)$$

where T is the length of the sample record in seconds. The bias error is the largest near peaks and dips in the spectrum.

The variance of the psd derived from the FFT is given by [3-4]:

$$V[G_x(f_k)] = G_x^2(f_k). \qquad (3\text{-}79)$$

This result indicates that the record length of data sample does not affect the variance. Therefore the variance of the psd cannot be reduced by increasing the number of samples N used in the FFT. This is in contrast to the evaluation of the acf where increasing N reduces the variance. When the record length of the psd estimation is increased the reduction in variance that would be achieved from the increased number of samples used is counteracted by the increase in variance resulting from the narrower filter bandwidth of the FFT.

This indicates two methods of reducing the variance of the psd estimate. One method is to average the psd computed for M ensembles of data samples, thereby increasing the effective record length without reducing the bandwidth of the FFT. The second method is to average the psd values at K adjacent frequency points, thereby increasing the effective bandwidth of FFT without reducing the record length.

The variance of an averaged psd estimate, denoted by $\overline{G_x(f_k)}$, is given by:

$$V[\overline{G_x}(f_k)] = \frac{G_x^2(f_k)}{MK} \tag{3-80}$$

where M is the number of ensemble averages and K is the number of averages at adjacent frequencies [3-5].

Additional reductions in the variance of the psd estimate using the indirect method are obtained from windowing the acf. The window function increases the effective filter bandwidth of the FFT, thereby reducing the variance by a factor equal to the mean square value of the window function, $\overline{w^2}$ [3-4]:

$$V[\overline{G_x}(f_k)]_{INDIRECT} = \frac{1}{\overline{w^2}} \frac{G_x^2(f_k)}{MK} \tag{3-81}$$

Values of $\overline{w^2}$ for typical window functions are given in Figure 3-14.

In the direct method of estimating the psd the time signal is windowed, which also increases the effective filter bandwidth. However, the effective record length of the sample is reduced by the factor $\overline{w^2}$ as a result of the windowing and the net result is that the variance is unchanged.

A comparison of the mean square errors (defined by mse = $V + B^2$) for the direct and indirect methods of computing the averaged sample psd is as follows:

$$mse[\overline{G_x}(f_k)]_{DIRECT} = \frac{G_x^2(f_k)}{MK} + \left[\frac{f_b^2\, G_x''(f_k)}{24}\right]^2$$

$$mse[\overline{G_x}(f_k)]_{INDIRECT} = \frac{1}{\overline{w^2}} \frac{G_x^2(f_k)}{MK} + \left[f_b^2 \frac{G_x''(f_k)}{24}\right]^2 \tag{3-82}$$

3.2.4 Digital Filtering

In this section we present several methods of digital filtering. In general, filtering is the process of transforming a signal x(t) to obtain a new signal y(t). The filtering process can be characterized in the frequency domain by the transfer function H(f) defined by

$$H(f) = \frac{Y(f)}{X(f)} \tag{3-83}$$

where X(f) and Y(f) are the Fourier Transforms of x(t) and y(t), respectively.

For a set of values x_n; n = 1, 2, ..., N; sampled periodically with time interval τ, the digital filtering process can be described in general as the generation of a new set of values y_n by a summation of the previous values of x_n and y_n:

$$y_n = \sum_{\ell=1}^{L} a_\ell y_{n-\ell} + \sum_{m=-M}^{M} b_m x_{n-m} \tag{3-84}$$

The filter transfer function is then given by:

$$H(f) = \frac{\sum_{m=-M}^{M} b_m e^{-j2\pi fm\tau}}{1 - \sum_{\ell=1}^{L} a_\ell e^{-j2\pi f\ell\tau}} ; \quad f < \frac{1}{2\tau} \tag{3-85}$$

where the form $e^{j\theta}$ is the exponential notation for the complex number defined by

$$e^{j\theta} = \cos\theta + j \sin\theta. \tag{3-86}$$

In principle, the values of a_ℓ and b_m can be determined by choosing M + L - 1 values of the desired H(f) and solving the set of simultaneous equations of the form of Equation (3-85). This is

usually not done directly in practice because of the difficulty in solving the set of equations and because there is no way to control the value of the transfer function between the chosen values.

There is a large variety of alternative methods which are presented in the literature for obtaining the parameters of digital filters [3-3, 3-10, 3-11, 3-12, 3-13, 3-14]. It is not the purpose here to present detailed descriptions of these methods. However, the parameters of some of the more generally applicable filters will be presented and the reader is referred to other references to obtain more detailed information.

The various types of digital filters can be grouped into two classifications based on the form of Equation (3-84). When the output y_n of the filter depends only on previous values of the input x_n ($a_\ell = 0$ for all ℓ) the filter is said to be nonrecursive and has a finite impulse response (FIR). In this case the impulse response is the set of values b_m which is the output of the filter when the input is a unit impulse ($x_1 = 1$ and $x_n = 0$ for all $n > 1$).

When the filter output y_n also depends on the previous values of y_n ($a_\ell \neq 0$ for some ℓ) the filter is said to be recursive and usually has an infinite impulse response (IIR).

The two types of filters have different advantages. Recursive filters can be used to approximate different types of analog filters with a relatively small number of parameters a_ℓ and b_m. Nonrecursive filters can be used to obtain filter characteristics which cannot be achieved with analog filtering. In particular, they can be designed so as not to introduce any phase shifts. Nonrecursive filters generally require a much larger number of parameters b_m than do recursive filters. However, nonrecursive filters can be implemented with the use of the FFT in order to shorten the required computation time.

The following paragraphs present some generally useful procedures for implementing various types of digital filters.

3.2.4.1 Nonrecursive Filters

A nonrecursive digital filter uses the form of Equation (3-84) with the values of $a_\ell = 0$. Then:

$$y_n = \sum_{m=-M}^{M} b_m x_{n-m}. \qquad (3\text{-}87)$$

The values b_m are then the finite impulse response of the filter. If the filter is symmetrical around m = 0 such that $b_{-m} = b_m$, the filter induces no phase shift on the signal. This type of filter is non-causal since the output values y_n depend on input values x_n which come at a later time. However, this is not a problem if the input values are stored in a data record. A symmetric filter can be made causal by retarding each coefficient M positions resulting in a constant time delay of $M\tau$ and a linear phase shift of $(2\pi fM\tau)$ radians.

LOWPASS FILTER: One type of a lowpass digital filter is obtained by approximating a perfectly rectangular lowpass filter $H_{LP}(f)$ shown in Figure 3-16a, where f_u is the upper cutoff frequency. The impulse response of this filter is given by:

$$h_{LP}(m\tau) = 2\tau \int_0^{f_u} \cos(2\pi fm\tau)df = \frac{\sin(2\pi f_u m\tau)}{\pi m} \qquad (3\text{-}88)$$

where $h_{LP}(0) = 2f_u\tau$.

This impulse response is infinitely long. However, the filter can be approximated by a finite impulse response if Equation (3-88) is truncated with a window function similar to those shown in Figure 3-14. As an example, consider the use of the Hanning window defined by:

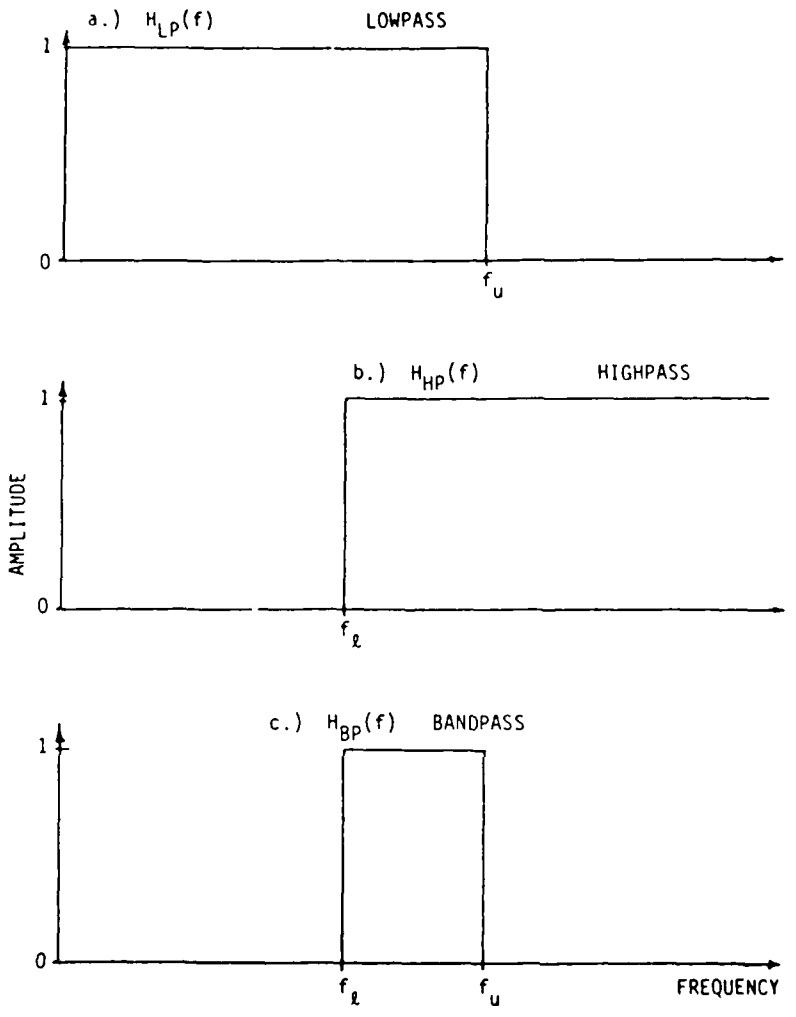


Figure 3-16. Rectangular Filter Shapes

$$w_m = \frac{1}{2}\left(1 + \cos\frac{\pi m}{M+1}\right) \qquad |m| \leq M$$
$$= 0 \qquad |m| > M \tag{3-89}$$

Then the nonrecursive lowpass filter coefficients are given by:

$$b_0 = 2f_u\tau$$
$$b_{-m} = b_m = w_m h_m = \frac{1}{2\pi m}\left(1 + \cos\frac{\pi m}{M+1}\right)\sin(2\pi f_u m\tau) \; ; \quad |m| \leq M \tag{3-90}$$

The larger the value of M, the sharper will be the cutoff of the filter.

HIGHPASS FILTER: A nonrecursive, highpass digital filter which approximates the perfectly rectangular highpass filter $H_{HP}(f)$ with a lower cutoff frequency f_ℓ, shown in Figure 3-16b, can be obtained by noting that:

$$H_{HP}(f) = 1 - H_{LP}(f) \tag{3-91}$$

when $f_\ell = f_u$. The impulse response of this filter is then given by:

$$h_{HP}(m\tau) = 2\tau \int_0^{\frac{1}{2\tau}} \left(1 - H_{LP}(f)\right)\cos(2\pi f m\tau)\,df$$
$$= \delta(m\tau) - h_{HP}(m\tau) \tag{3-92}$$

where $\delta(m\tau) = 1$ for $m = 0$

$= 0$ for $m \neq 0$. (3-93)

This indicates that the highpass filter can be obtained directly from the corresponding lowpass filter coefficients. Using the coefficients in Equation (3-90), the nonrecursive highpass filter coefficients are given by:

$$b_0 = 1 - 2f_\ell\tau$$

$$b_{-m} = b_m = \frac{-1}{2\pi m}\left(1 + \cos\frac{\pi m}{M+1}\right)\sin(2\pi f_\ell m\tau) \;; \quad |m| \leq M \qquad (3\text{-}94)$$

BANDPASS FILTER: A nonrecursive, bandpass digital filter which approximates the perfectly rectangular bandpass filter $H_{BP}(f)$ with a lower cutoff frequency f_ℓ and an upper cutoff frequency f_u, shown in Figure 3-16c, can be obtained from the impulse response:

$$h_{BP}(m\tau) = 2\tau\int_{f_\ell}^{f_u}\cos(\pi f m\tau)\,df$$

$$= \frac{1}{\pi m}\left(\sin(2\pi f_u m\tau) - \sin(2\pi f_\ell m\tau)\right) \qquad (3\text{-}95)$$

where $h_{BP}(0) = 2\tau(f_u - f_\ell)$.

Using the same Hanning window as before on the impulse response results in a nonrecursive bandpass filter with coefficients given by:

$$b_0 = 2\tau(f_u - f_\ell) \qquad (3\text{-}96)$$

$$b_{-m} = b_m = \frac{1}{2\pi m}\left(1 + \cos\frac{\pi m}{M+1}\right)\left(\sin(2\pi f_u m\tau) - \sin(2\pi f_\ell m\tau)\right) \;; \quad |m| \leq M$$

More information on these and many other types of nonrecursive filters can be found in References [3-3, 3-10, 3-11, 3-12].

3.2.4.2 Recursive Filters

A recursive digital filter uses the form of Equation (3-84) with some values of $a_\ell \neq 0$. Usually these types of filters are made causal by setting $b_m = 0$ for $m < 0$. Then:

$$y_n = \sum_{\ell=1}^{L} a_\ell y_{n-1} - \sum_{m=0}^{M} b_m x_{n-m}. \qquad (3\text{-}97)$$

In this form a recursive filter can be made to approximate the transfer function of an analog filter. The general approach used to approximate an analog filter by a digital recursion relationship in the form of Equation (3-97) is first to represent the filter transfer function in the form of Equation (3-85) and then to match one-to-one the coefficients of each term with the appropriate a_ℓ and b_m terms.

LOWPASS FILTER: One of the most common types of analog lowpass filters is the Butterworth filter with the transfer function given by:

$$|H_{LP}(f)|^2 = \frac{1}{1 + \left(\frac{f}{f_u}\right)^{2K}} \qquad (3\text{-}98)$$

where f_u is the upper cutoff frequency and K is the order of the filter. This filter has an attenuation of 3 dB at $f = f_u$ and rolls off at a rate of 6 K dB/octave above this frequency.

One method of approximating this filter (called the bilinear transformation) represents the filter transfer function by the product of a set of transfer functions in the form (for K even):

$$H_{LP}(f) = \prod_{k=1}^{K/2} H_k(f) \qquad (3\text{-}99)$$

where

$$H_k(f) = \frac{S_u^2(1 + 2z + z^2)}{(S_u^2 - 2C_kS_u + 1) + 2(S_u^2 - 1)z_z + (S_u^2 + 2C_kS_u + 1)z^2}$$

$$S_u = \tan(\pi f_u \tau)$$

$$C_k = \cos\left(\pi\left(\frac{1}{2} + \frac{2k-1}{2K}\right)\right)$$

$$z = e^{-j2\pi f\tau} \qquad (3\text{-}100)$$

This filter can then be implemented by a sequence of $\frac{K}{2}$ recursive filters in the form of Equation (3-97) with M = L = 2 and coefficients of each given by:

$$b_{0,k} = \frac{S_u^2}{S^2 - 2C_kS_u + 1}$$

$$b_{1,k} = 2b_{0,k}$$

$$b_{2,k} = b_{0,k}$$

$$a_{1,k} = \frac{2(1 - S_u^2)}{S_u^2 - 2C_kS_u + 1}$$

$$a_{2,k} = -\left(\frac{S_u^2 + 2C_kS_u + 1}{S_u^2 - 2C_kS_u + 1}\right) \qquad (3\text{-}101)$$

When K is odd the transfer function can be represented by:

$$H_{LP}(f) = H_0(f) \prod_{k=1}^{\frac{K-1}{2}} H_k(f) \qquad (3\text{-}102)$$

where

$$H_0(f) = \frac{S_u(1 + z)}{(S_u + 1) + (S_u - 1)z} \qquad (3\text{-}103)$$

and the $H_k(f)$ are the same as above.

This filter can then be implemented by a sequence of $\frac{K-1}{2}$ recursive filters in the form given above followed by an additional recursive filter with M = L = 1 and coefficients given by:

$$b_0 = \frac{S_u}{1 + S_u}$$

$$b_1 = b_0$$

$$a_1 = \frac{1 - S_u}{1 + S_u} \qquad (3\text{-}104)$$

HIGHPASS FILTER: A recursive, highpass digital filter can be obtained by transforming an appropriate lowpass filter of the same type. The transformation can be made most easily if the lowpass cutoff frequency f_u is related to the highpass cutoff frequency f_ℓ by:

$$f_u + f_\ell = \frac{1}{2\tau} \qquad (3\text{-}105)$$

Then by changing z to -z in the lowpass filter transfer function it is transformed into a highpass filter.

A Butterworth highpass filter of the form:

$$|H_{HP}(f)|^2 = \frac{1}{1 - \left(\frac{f_\ell}{f}\right)^{2K}} \qquad (3\text{-}106)$$

can be approximated as a digital filter by transforming Equation (3-100) and (for K odd) Equation (3-103). This filter can then be implemented by a sequence of $\frac{K}{2}$ recursive filters in the form of Equation (3-97) with M = L = 2 and coefficients of each given by:

$$b_{0,k} = \frac{S_\ell^2}{S_\ell^2 - 2C_k S_\ell + 1}$$

$$b_{1,k} = -2b_{o,k}$$

$$b_{2,k} = b_{0,k}$$

$$a_{1,k} = \frac{2(S_\ell^2 - 1)}{S_\ell^2 - 2C_k S_\ell + 1}$$

$$a_{2,k} = -\left(\frac{S_\ell^2 + 2C_k S_\ell + 1}{S_\ell^2 - 2C_k S_\ell + 1}\right) \qquad (3\text{-}107)$$

where $S_\ell = \tan\left(\frac{\pi}{2} - \pi f_\ell \tau\right) = \cot(\pi f_\ell \tau)$

$$C_k = \cos\left(\pi\left(\frac{1}{2} + \frac{2k-1}{2K}\right)\right) \qquad (3\text{-}108)$$

When K is odd an additional filter is required with M = L = 1 and coefficients given by:

$$b_0 = \frac{S_\ell}{S_\ell + 1}$$

$$b_1 = -b_0$$

$$a_1 = \frac{S_\ell - 1}{S_\pi + 1} \qquad (3\text{-}109)$$

BANDPASS FILTER: A recursive, bandpass digital filter can also be obtained by transforming an appropriate lowpass filter of the same type. However, if the bandpass filter can be represented as a product of a lowpass and a highpass filter, then the sequential implementation of a set of filters (half lowpass and half highpass) can be used to obtain the desired results.

More information on these and many other types of recursive filters can be found in References [3-3, 3-10, 3-13, 3-14].

3.2.4.3 Integration

An integrator is a filter with a transfer function given by:

$$H(f) = \frac{1}{j2\pi f}. \tag{3-110}$$

One difficulty in implementing an integrator is that the transfer function goes to infinity at f = 0. This means that low frequency components of a signal are greatly amplified by the integration. Therefore, in practice an integrator must have a low frequency limit, below which the transfer function diverges from a true integrator.

One type of integrator can be obtained from a first order Butterworth lowpass filter with a transfer function:

$$|H(f)|^2 = \frac{1}{1 + (\frac{f}{f_c})^2} \cdot \frac{1}{(2\pi f_c)^2} \tag{3-111}$$

where $f_c << \frac{1}{\tau}$.

This filter will behave like an integrator for frequencies $f >> f_c$. This can be implemented using a recursive filter of the form of Equation (3-97) with M = L = 1 and coefficients:

$$b_0 = \frac{\tau}{2}$$

$$b_1 = \frac{\tau}{2}$$

$$a_1 = \frac{1 - \pi f_c \tau}{1 + \pi f_c \tau} \qquad (3\text{-}112)$$

This filter has an impulse response that approximates the function $h(m\tau) = e^{-2\pi f_c m\tau}$.

An alternative method of integrating a signal is to perform the operation in the frequency domain using the following procedure:

1. Sample the signal N times with period τ.

2. Compute the Fourier Transform of the signal using an N point FFT (see Section 3.2.3.1).

3. Multiply the output Z_k of the FFT by the complex transfer function of the integrator $H(f) = (j2\pi f)^{-1}$ to obtain a new array Y_k given by:

$$Y_1 = Z_1$$

$$Y_k = \frac{N\tau}{j2\pi k} Z_k \; ; \quad k = 2, 3, \ldots, \frac{N}{2}$$

$$Y_{\frac{N}{2}+1} = \frac{N\tau}{\pi(N+2)} Z_{\frac{N}{2}+1}$$

$$Y_k = -\frac{N\tau}{j2\pi k} Z_k \; ; \quad k = \frac{N}{2}+2, \ldots, N \qquad (3\text{-}113)$$

4. Compute the inverse FFT of the array Y_k to obtain the integrated signal.

This method is particularly useful if the desired quantity is the power spectrum of the integrated signal, in which case the last step is not needed. This method is also applicable to dual and multiple signal analysis where the cross spectral density of two signals can be multiplied by the integrator transfer function in order to achieve an integration of one of the signals. As in the case of all FFT operations, care must be taken to insure the proper sign is used on the imaginary part of the complex arrays in order to avoid phase shifts by factors of $\pi/2$.

3.2.4.4 Differentiation

A differentiator is a filter with a transfer function given by:

$$H(f) = j2\pi f. \tag{3-114}$$

Several methods are available for implementing a differentiator. One method is to differentiate in the frequency domain as discussed above for integrators. In this case the same procedure is used as for the integrator except that the transfer function is inverted.

A second method is to develop a recursive filter which approximates the transfer function of a differentiator. In Reference [3-15], an optimal second order recursive differentiator was developed in the form of Equation (3-97) with M = L = 2 and coefficients given by:

$$b_0 = 1.15099674/\tau$$

$$b_1 = -0.37887796/\tau$$

$$b_2 = -(b_0 + b_1)$$

$$a_1 = -0.85938970/\tau$$

$$a_2 = -0.10210106/\tau \tag{3-115}$$

More information on these and other types of recursive differentiators can be found in Reference [3-15].

3.2.4.5 Interpolation

Interpolation is a form of digital filtering which changes the sampling characteristics of digital data. If a signal x(t) is sampled N times with period τ, then the value of the signal at anytime $t = n\tau + t_s$; $0 \leq n \leq N$, $0 \leq t_s \leq \tau$; can be approximated by:

$$y_n = x(n\tau + t_s) \approx \sum_{m=n-1}^{n-N} \frac{\sin\left[\frac{\pi}{\tau}(m\tau - t_s)\right]}{\frac{\pi}{\tau}(m\tau - t_s)} x_{n-m} \qquad (3\text{-}116)$$

This formula can be used to change the effective sampling rate of the signal or shift the data by the time t_s to correct for time delays due to multiplexing. NOTE: Before reducing the effective sampling rate of a signal, the digital data must be lowpass filtered to prevent aliasing.

The direct application of Equation (3-116) involves a large number of multiplications in order to generate a complete set of new data. Therefore, it is desirable to use an approximation to Equation (3-116) for each particular application.

TIME SHIFTS: In order to shift the sample time of a set of digital data by a fraction of the sample period while not changing the sample rate, an approximation to Equation (3-116) can be made using a nonrecursive, lowpass filter of the form of Equation (3-87) with the coefficients given by:

$$b_m = \frac{-\tau(-1)^m}{2\pi(m\tau - t_s)}\left(1 + \cos\frac{\pi m}{M+1}\right)\sin(\pi t_s/\tau); \qquad 0 < t_s < \tau, \quad |m| \leq M \qquad (3\text{-}117)$$

where the larger the value of M, the more accurate is the interpolation.

SAMPLE RATE CHANGES: A change in the effective sample rate can be most easily accomplished if the new sample period is given by $\tau' = \frac{K}{L}\tau$ where K and L are integers. This can be implemented conceptually by first increasing the sample rate by a factor of L and then decreasing the sample rate by a factor of K (after lowpass filtering). An efficient method of implementing this interpolator is given in Reference [3-16] and is summarized below with slight modifications:

1. Define a nonrecursive lowpass filter of length M = QL where Q is a positive integer which has an upper cutoff frequency defined by

$$f_u = \min\left[\frac{1}{2L\tau}, \frac{1}{2K\tau}\right]. \tag{3-118}$$

For example, using Equation (3-90) gives:

$$b_0 = 2f_u\tau$$

$$b_{-m} = b_m = \frac{1}{2\pi m}\left(1 + \cos\frac{\pi m}{QL+1}\right) \sin(2\pi f_u m\tau) \; ; \quad m \leq QL. \tag{3-119}$$

2. Compute a new set of data y_n from the original set of data x_n by the recursion relationship:

$$y_n = \sum_{q=1}^{Q} b_m x_{i-q} \tag{3-120}$$

where $i = \text{Int}\left[\frac{(n-1)K}{L} + 2\right]$

$$m = (q-1)L + k$$

$$k = \ell K - \text{Int}\left[\frac{\ell K}{L}\right] \cdot L$$

$$\ell = n - 1 - \text{Int}\left[\frac{n-1}{L}\right] \cdot L \tag{3-121}$$

This will result in a change in the number of samples by a factor of $\frac{L}{K}$. The larger the value of Q, the greater will be the accuracy of the interpolation.

3.2.4.6 Statistical Errors in Digital Filtering

One approach to evaluating the statistical error of a digital filter is to compute the mean square variation of the filter transfer function from that of an ideal filter. This can be done by the following procedure.

1. Compute N values of the filter impulse response by filtering the sequence consisting of the first value equal to unity and the following N-1 values equal to zero.

2. Compute the FFT of the filter impulse response to obtain the frequency transfer function denoted by $H_N(f_k)$; $f_k = k/N$, $k = 0, 1, \ldots, N-1$.

3. Compute the mean square error (mse) of the filter response $H_N(f_k)$ from an ideal filter $H_I(f_k)$ by

$$mse = \frac{1}{N} \sum_{k=0}^{N-1} [H_N(f_k) - H_I(f_k)]^2. \qquad (3\text{-}122)$$

3.2.5 Special Analysis Considerations

The digital data analysis procedures outlined in this chapter are designed primarily for random, stationary data. When the data contain significant periodic components or are non-stationary over the analysis time period, special considerations must be made in the analysis. These considerations are discussed in the following sections.

3.2.5.1 Periodic Signals

Although the procedures using the Fourier Transform to obtain spectral information of a signal are based on the theory of periodic signals, the use of the FFT to analyze periodic signals can lead to erroneous results. This is because the period of the signal will not in general be equal to that of the FFT. The FFT period T (in seconds) is determined by the data sample rate f_s (in samples per second) and the number of samples N by the relationship:

$$T = \frac{N}{f_s}.$$

If the FFT period is not an integer multiple of the signal period, then the frequency components of the signal will not correspond to the set of FFT frequencies $f_k = k/T$; $k = 0, 1, \ldots, N-1$. Even if the signal period is known, it is often difficult to match the FFT period to it because of the constraints imposed by the analysis system on the values of f_s and N.

In addition, the fact that the signal is sampled for only a finite time interval imposes an inherent constraint on the accuracy with which the frequencies of the periodic signal can be resolved. The uncertainty in frequency Δf is in general inversely proportional to T, which can be stated by:

$$\Delta f T = \text{constant}. \qquad (3\text{-}123)$$

For FFT analysis this constant has a value of 1.

One method to minimize the errors in using the FFT on periodic data is based on the indirect method of computing the power spectral density (see Section 3.2.3.2). The autocorrelation of a periodic signal will also be periodic. In general, the finite length autocorrelation obtained from the indirect method will not contain an integer number of periods. However, if the periodicity can be identified, the autocorrelation can be resampled using an interpolation method (see Section 3.2.4.5) in order to obtain an integer number of periods in the sample record.

This method can also be used for signals with a broad band random component (such as noise) added to a periodic signal. This is illustrated by the following procedure:

1. Obtain N samples of the signal with a sample record which includes several signal periods and extend the sample record with N zeroes. Do not apply any smoothing window.

2. Compute the sample psd with a 2N point FFT and average the psd over ensembles of sample records.

3. Compute the acf with a 2N point inverse FFT of the average psd. Truncate the acf at ± N/2 lag points.

4. Identify the time lag T_p of the last local maximum in the acf by interpolating between two adjacent maxima in the acf sample points. The time T_p then corresponds to an integer number of periods in the signal.

5. Resample the acf using an interpolation method such that N points cover the time interval ± T_p.

6. Compute the N point FFT of the resampled acf to obtain the periodic components of the signal plus the psd of the random component.

3.2.5.2 Non-stationary Signals

Non-stationary signals are characterized by variations in their statistical measures as a function of time. Therefore, in general, all statistical measures of non-stationary signals must be identified by the particular time interval over which they are obtained [3-4]. However, there are two types of non-stationary signals for which some simplifications can be made. One type is a signal which is non-stationary only in its mean and variance. The second is a transient signal which has a nonzero value only over a finite time interval.

NON-STATIONARY IN MEAN AND VARIANCE: Many non-stationary signals $x(t)$ arising from measurements of physical systems with time varying conditions can be modeled by:

$$x(t) = A(t)\ y(t) + B(t) \tag{3-124}$$

where $y(t)$ is a stationary signal with zero mean and unity variance. $A(t)$ and $B(t)$ are non-stationary time functions with slowly varying amplitudes (i.e., they have a maximum frequency component much less than that of $y(t)$).

The first two moments of $x(t)$ can then be approximated by:

$$E[x(t)] \simeq B(t) \tag{3-125}$$

$$V[x(t)] \simeq A^2(t) \tag{3-126}$$

The functions $A(t)$ and $B(t)$ can be estimated from an average over a series of N samples $x_n = x(t + n\tau)$ sampled periodically with an interval τ as follows:

$$B(t) \simeq \overline{B_T}(t) = \frac{1}{N} \sum_{n=\frac{N}{2}+1}^{N/2} x_n \tag{3-127}$$

$$A^2(t) \approx \overline{A_T^2}(t) = \frac{1}{N-1} \sum_{n=-\frac{N}{2}+1}^{N/2} [x_n - \overline{B_T}(t)]^2 \qquad (3\text{-}128)$$

where $T = N\tau$ is used as a subscript to denote the length of time over which the average is taken.

The stationary signal y(t) can then be approximated by:

$$y_n \approx \frac{x_n - \overline{B_T}(t)}{\sqrt{\overline{A_T^2}(t)}} \qquad (3\text{-}129)$$

This signal can be analyzed using the various methods discussed in this chapter for stationary signals. Average statistical measures of y(t) can be obtained by averaging over multiple sample records of x(t) where new values of $\overline{B_T}(t)$ and $\overline{A_T^2}(t)$ are computed for each sample record. NOTE: The best estimates of y(t) are obtained from sample records with a large variance. If $\overline{A_T^2}(t)$ becomes very small, the background noise in x(t) will contaminate the estimate of y(t). Therefore, it is best to set a lower limit on the value of $\overline{A_T^2}(t)$ (such as 10 times the variance of the background noise) and reject sample records with a smaller variance than this.

The values of $\overline{B_T}$ and $\overline{A_T^2}$ obtained from successive sample records can also be used to analyze the functions B(t) and A(t), especially if the sample records are spaced in time by a uniform time interval. In this manner a fairly complete description of the non-stationary signal x(t) can be obtained.

TRANSIENT SIGNALS: A transient signal is one which is non-zero for only a finite length of time. Transient signals often occur in systems that are excited by impulsive loads such as mechanical impacts and explosions. Because of the limited duration of the signal, time averages of statistical measures are of little value. If the averaging time is longer than the duration

of the signal then the average value will decrease as the averaging time increases. Therefore, statistical measures of the total contribution of transient signals are more appropriate than time averages. Some examples are given in the following paragraphs.

Suppose that a transient signal x(t) has a value of zero outside the time interval [0,T], and inside the time interval N samples x_n are obtained with a sampling period τ such that $N\tau = T$. The total energy E_x of the signal can be computed by:

$$E_x = \tau \sum_{n=1}^{N} x_n^2. \qquad (3\text{-}130)$$

The value of E_x is equal to the mean square value (Equation (3-18)) multiplied by T.

The frequency spectrum of the transient signal x(t) can be computed using the FFT in a manner similar to that for a continuous signal (Section 3.2.3). However, it is more appropriate in this case to define an energy spectral density ε_x given by:

$$\varepsilon_x(f_k) = 2|X_T(f_k)|^2 \; ; \quad k = 0, 1, \ldots, \frac{N}{2}-1. \qquad (3\text{-}131)$$

where X_T is the Fourier Transform of x. The value of ε_x is equal to the value of the power spectral density (Equation (3-63)) multiplied by T.

The statistical errors in estimating the energy spectral density can be evaluated in a similar manner as for the power spectral density which is outlined in Section 3.2.3.4. The statistical errors can be reduced by averaging adjacent frequency points or by averaging over an ensemble of repeated transients, if possible.

3.3 DUAL TIME SERIES

In this section we present various methods of analyzing the interrelationship of two signals sampled over the same time period. As discussed in Section 3.1.1.2 it is not necessary that the two signals be sampled at exactly the same instants. There may exist a small time delay between the sample times for the two signals which remains fixed for all samples. This process is called multiplexing and can be accounted for in the analysis procedures.

3.3.1 Joint Statistical Measures

This section on joint statistical measures is an extension of Section 3.2.1 which is limited to single channel analysis. All of the methods in that section apply equally well to a simultaneous application to two signals. However, there are additional statistical measures that determine the interrelationship of the two signals.

If the two signals have been sampled with multiplexing, the time delay between channels must be removed by an by an interpolation of one of the channels to obtain sample values at exactly the same times before the computation methods are discussed in Section 3.2.4.5.

3.3.1.1 Joint Probability Density Function

The sample joint probability density function (pdf) of two sets of N values x_n and y_n is related to the number of value pairs (x_n, y_n) that fall within various value intervals for both x_n and y_n. Suppose the range of x_n values is divided into intervals of width w_x, centered at values X_i; $i = 1, 2, \ldots, M$; and the range of y_n values is divided into intervals of width w_y, centered at values Y_j; $j = 1, 2, \ldots, K$. Then a two-dimensional

array N_{ij} can be defined as the number of sample pairs (x_n,y_n); $n = 1, 2, \ldots, N$; that lie in each combination of value intervals w_x and w_y. The joint pdf, denoted by $p(X_i,Y_j)$ is computed by:

$$p(X_i,Y_j) = \frac{N_{ij}}{w_x w_y N}. \tag{3-132}$$

The joint pdf is normalized in magnitude so that

$$\sum_{i=1}^{M} \sum_{j=1}^{K} p(X_i,Y_j)\, w_x w_y = 1. \tag{3-133}$$

A procedure for computing a joint pdf from a set of N value pairs (x_n,y_n) can be obtained by an extension of Table 3-1.

The accuracy of the joint pdf will depend on the choices of w_x and w_y for a given number N samples. As shown in Section 3.2.1.4, good estimates for the choices of the value intervals are:

$$w_x \simeq \frac{20\, s_x}{\sqrt{N}}$$
$$w_y \simeq \frac{20\, s_y}{\sqrt{N}} \tag{3-134}$$

where s_x and s_y are the standard deviations of x_n and y_n, respectively.

3.3.1.2 Joint Cumulative Distribution Function

The joint cumulative distribution function (cdf) of a set of N value pairs (x_n,y_n) is a measure of the probability of the simultaneous conditions that x_n is less than a certain value X_k and y_n is less than a certain value Y_ℓ. The joint cdf, denoted by $P(X_k,Y_\ell)$, can be computed from the joint pdf by summing up the values of the pdf at all intervals where $X_i < X_k$ and $Y_j < Y_\ell$:

$$P(X_k, Y_\ell) = \sum_{i=0}^{k} \sum_{j=0}^{\ell} p(X_i, Y_j) \qquad (3\text{-}135)$$

3.3.1.3 Conditional Probability Density Function

The conditional pdf of a set of N value pairs (x_n, y_n) is a measure of the pdf of x_n given the condition that y_n has a certain value y or vice-versa. The pdf of x_n given a value of Y_n is denoted by $p(X_i|Y_j)$ and can be calculated from the joint pdf as follows:

$$p(X_i|Y_j) = \frac{p(X_i, Y_j)}{\sum_{i=0}^{M+1} p(X_i, Y_j)} \qquad (3\text{-}136)$$

where $\sum_{i=0}^{M+1} p(X_i, Y_j) = p(Y_j)$ is called the marginal pdf of Y_n.

The various types of pdf of two variables are related by the following:

$$p(X_i, Y_j) = p(Y_j)\ p(X_i|Y_j) = p(X_i)\ p(Y_j|X_i) \qquad (3\text{-}137)$$

If the two variables are independent then

$$p(X_i|Y_j) = p(X_i)$$

and

$$p(Y_j|X_i) = p(Y_j)$$

so that

$$p(X_i, Y_j) = p(X_i)\ p(Y_j). \qquad (3\text{-}138)$$

3.3.1.4 Covariance

The covariance (cov) of a set of N value pairs (x_n, y_n) is a measure of the relationship between the values x_n and y_n. The

cov, denoted by c_{xy}, is the average value of the cross product of the value pairs with the mean values removed:

$$c_{xy} = \frac{1}{N-1} \sum_{n=1}^{N} (x_n - \bar{x})(y_n - \bar{y}) \qquad (3\text{-}139)$$

From this formula it can be seen that the cov of a signal with itself is equal to the sample variance:

$$c_{xx} = s_x^2 \qquad (3\text{-}140)$$

The covariance is often expressed in terms of the correlation coefficient r_{xy} which is the normalized c_{xy} defined by:

$$r_{xy} = \frac{c_{xy}}{s_x \, s_y} \qquad (3\text{-}141)$$

The magnitude of the correlation coefficient is limited by:

$$-1 \leq r_{xy} \leq 1. \qquad (3\text{-}142)$$

If $r_{xy} = \pm 1$ the signals x and y are said to be perfectly correlated. If $r_{xy} = 0$ the signals x and y are said to be uncorrelated.

3.3.1.5 Regression Analysis

A regression analysis of a set of N value pairs (x_n, y_n) is a method of determining the relationship between the two values. In general, a regression of the mean value of y_n on the value of x_n is defined as the expected value of y_n given a certain value $x_n = X_i$. The regression function is denoted by $E[y_n | X_i]$ and can be computed from the conditional pdf by:

$$E[y_n | X_i] = \sum_{j=0}^{K+1} Y_j \, p(Y_j | X_i). \qquad (3\text{-}143)$$

In many applications it is desirable to approximate the regression function by an equation with a few variables. The most popular approximation is the use of a straight line and is called linear regression. Linear regression is a method of estimating the regression function by a straight line of the form:

$$y_n = ax_n + b + \varepsilon \tag{3-144}$$

where ε is a random error.

The parameters a and b are chosen so as to minimize the sum of the squared errors of the form $\varepsilon_n^2 = (y_n - ax_n - b)^2$.

For a set of N value pairs (x_n, y_n) the linear regression parameters can be computed by:

$$a = \frac{\sum_{n=1}^{N} x_n y_n - N\bar{x}\,\bar{y}}{\sum_{n=1}^{N} x_n^2 - N(\bar{x})^2} = \frac{c_{xy}}{s_x^2}$$

$$b = \bar{y} - a\bar{x}. \tag{3-145}$$

3.3.1.6 Accuracy of Joint Statistical Measures

In this Section we present methods of evaluating the accuracy of the joint statistical measures discussed in the previous Section. The discussion of confidence intervals in Section 3.2.1.8 is expanded here to apply to the various joint statistical measures.

SAMPLE JOINT PDF: The expected value of the sample joint pdf evaluated by Equation (3-132) is given by:

$$E\left[\frac{N_{ij}}{w_x w_y N}\right] \simeq p(x_i, y_j) + \frac{w_x^2}{24} p_{xx}(x_i, y_j) + \frac{w_y^2}{24} p_{yy}(x_i, y_j) \tag{3-146}$$

where $p(x_i,y_j)$ is the true joint pdf, $p_{xx}(x_i,y_j)$ is the second derivative with respect to x, and $p_{yy}(x_i,y_j)$ is the second derivative with respect to y.

The variance of the sample joint pdf is given by:

$$V\left[\frac{N_{ij}}{w_x w_y N}\right] \approx \frac{2p(x_i,y_j)}{w_x w_y N} \tag{3-147}$$

SAMPLE COVARIANCE: For a set of N value signal pairs (x_n,y_n) with variances of c_x and σ_y, respectively, the expected value of the sample covariance c_{xy} evaluated by Equation (3-139) is given by:

$$E[c_{xy}] = r_{xy}\sigma_x\sigma_y \tag{3-148}$$

where r_{xy} is the true correlation coefficient between x and y. The variance of c_{xy} is dependent on the cross covariance function defined by Equation (3-155) in Section 3.3.2. The variance is given by:

$$V[c_{xy}] = \frac{1}{N^2} \sum_{i=-n}^{N} \left[(N - |i|)\left(C_x(i\tau)\, C_y(i\tau) + C_{xy}(i\tau)\, C_{xy}(-i\tau)\right)\right] \tag{3-149}$$

LINEAR REGRESSION: For a set of N value pairs (x_n,y_n) the linear regression of y_n on x_n evaluated by Equation (3-144) has an expected value given by:

$$E[ax_n + b] = y_n \tag{3-150}$$

and a variance

$$V[ax_n + b] = \frac{1}{N-2} \sum_{n=1}^{N} \varepsilon_n^2 \qquad (3\text{-}151)$$

3.3.2 Cross Correlation Function

In this section we present a method of calculating the cross correlation function (ccf) of a set of periodically sampled value pairs (x_n, y_n). The ccf is a measure of the relationship between the values of two signals at different times, t_1 and t_2. This relationship is measured by the average of the product of the two signal values, $x(t_1)$ and $y(t_2)$. The ccf is a generalization of the acf described in Section 3.2.2 and the same considerations relative to the stationarity of the signals apply.

The time averaged ccf of a signal pair x(t) and y(t) can be estimated by the sample ccf, denoted by R_{xy}, of a set of N value pairs (x_n, y_n) sampled periodically with a time interval τ as follows:

$$R_{xy}(m\tau) = \frac{1}{N-m} \sum_{n=1}^{N-m} x_n y_{n+m} \; ; \quad m \geq 0$$

$$= \frac{1}{N+m} \sum_{n=1-m} x_n y_{n+m} \; ; \quad m \leq 0 \qquad (3\text{-}152)$$

This direct method of estimating the ccf is generally used only for small values of m, because it requires a large number of computations. When m is large a more general procedure can be used which is based on the relationship of the ccf to the cross spectral density of the two signals. This procedure is outlined in Section 3.3.3.1.

NOTE: If the value pairs (x_n, y_n) are obtained through multiplex sampling with time shift t_s such that $x_n = x(n\tau)$ and $y_n = y(n\tau + t_s)$ then the ccf is evaluated at time differences of $(m\tau + t_s)$.

The following are some general properties of the ccf:

1. For stationary signals

$$R_{xy}(m\tau) = R_{yx}(-m\tau) \tag{3-153}$$

2. For non-periodic signals

$$\lim_{m \to \infty} [R_{xy}(m\tau)] = \bar{x}\,\bar{y} \tag{3-154}$$

Figure 3-17 presents the ccf of some typical signals. The ccf of two periodic signals is also periodic. The ccf of two independent signals with zero mean is zero. For two signals with some interdependence the peak value of the ccf does not always occur at m = 0. The location of the peak value of the ccf is one indication of the time lag (or propagation time) between the two signals. Caution must be used if the time lag is a significant fraction of the record length $T = N\tau$ in which case some of the correlated parts of the two signals may not be contained in the same record length.

When the ccf is to be evaluated for signals having a non-zero mean, it is computationally desirable to remove the mean value first before using Equation (3-152). The resulting quantity is called the cross covariance function (ccvf), denoted by C_{xy}, which is defined by:

$$C_{xy}(m\tau) = \frac{1}{N-m} \sum_{n=1}^{N-m} (x_n - \bar{x})(y_{n+m} - \bar{y}) \; ; \quad m < N \tag{3-155}$$

The ccf and ccvf are related by:

$$R_{xy}(m\tau) = C_{xy}(m\tau) + \bar{x}\,\bar{y} \tag{3-156}$$

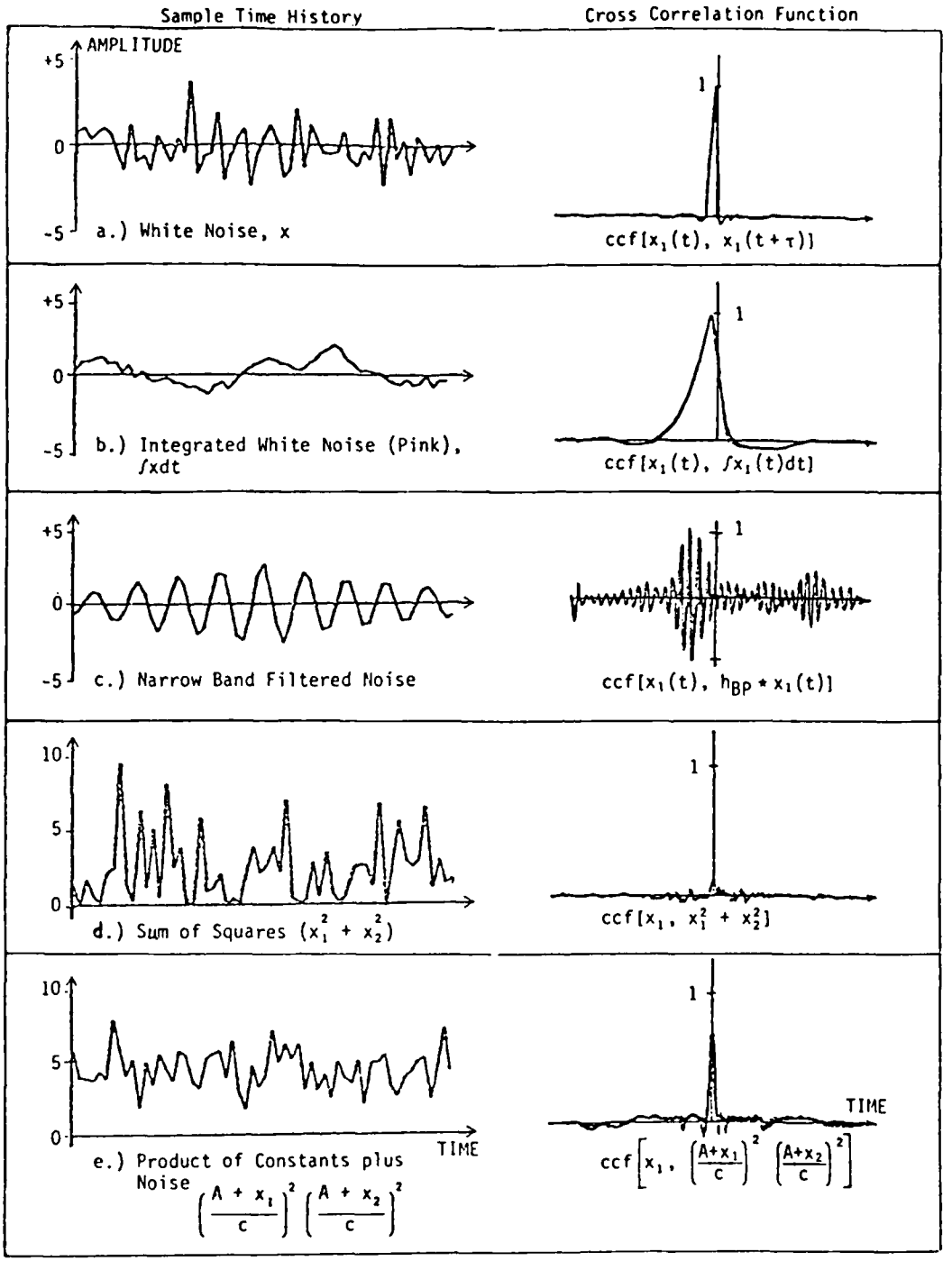


Figure 3-17. Cross Correlation Function of Some Typical Signal Pairs

The ccf is often represented by the normalized correlation function ρ which is defined by:

$$\rho_{xy}(m\tau) = \frac{C_{xy}(m\tau)}{\sqrt{C_x(0)\ C_y(0)}} \tag{3-157}$$

where C_x and C_y are the autocovariance functions of each signal defined by Equation (3-44). The magnitude of the normalized correlation function is bounded by:

$$|\rho_{xy}(m\tau)| \leq 1. \tag{3-158}$$

3.3.2.1 Statistical Errors in the CCF Estimate

The bias and variance of the ccf estimate using Equation (3-152) are given by:

$$B[R_{xy}(m\tau)] = 0 \tag{3-159}$$

$$V[R_{xy}(m\tau)] = \frac{1}{(N-m)^2} \sum_{i=-N+m}^{N-m} [(N-m-|i|)(R_x(i\tau)R_y(i\tau) + R_{xy}([m+i]\tau)R_{xy}([m-i]\tau))] \quad ; \quad m \geq 0 \tag{3-160}$$

This is the two-signal equivalent to Equation (3-50) which applies to the autocorrelation function of one signal.

In a similar manner to the discussion in Section 3.2.2.1 a biased estimate of the ccf can be defined by:

$$R^b_{xy}(m\tau) = \frac{1}{N} \sum_{n=1}^{N-m} x_n y_{n+m} \tag{3-161}$$

which has a bias error given by:

$$B[R^{b}_{xy}(m\tau)] = \left(\frac{m}{N}\right) R_x(m\tau) \qquad (3\text{-}162)$$

and a variance given by:

$$V[R^{b}_{xy}(m\tau)] = \left(1 - \frac{m}{N}\right)^2 V[R_{xy}(m\tau)] \qquad (3\text{-}163)$$

By averaging either ccf estimate over M sample records, the variance can be reduced by a factor of 1/M while the bias cannot be reduced.

3.3.3 Cross Spectral Density Function

In this section we present two methods of calculating the cross spectral density (csd) of a set of periodically sampled value pairs (x_n, y_n) obtained from continuous signals x(t) and y(t). The csd is a measure of the frequency distribution of the average product of the two signals. The csd, denoted by $G_{xy}(f)$, can be defined in terms of analog filtering as the average product of the outputs of two narrow band filters divided by the bandwidth of the filters in the limit of the bandwidth becoming small:

$$G_{xy}(f) = \lim_{\delta f \to 0} \left[\frac{x(t,f,\delta f)y(t,f,\delta f) - jx(t,f,\delta f)y\left(t + \frac{1}{4f}, f, \delta f\right)}{\delta f} \right] \qquad (3\text{-}164)$$

where $x(t,f,\delta f)$ and $y(t,f,\delta f)$ are the outputs of the bandpass filters of width δf centered at frequency f (cycles per second, or Hertz).

The csd is a complex function of the form:

$$G_{xy}(f) = C_{xy}(f) - j\, Q_{xy}(f) \qquad (3\text{-}165)$$

where $C_{xy}(f)$ is called the coincident, or in-phase, spectral density function and $Q_{xy}(f)$ is called the quadrature, or out-of-phase, spectral density function. This results from the definition in Equation (3-164) where $C_{xy}(f)$ is the average cross product of the two signals at the same time, while $Q_{xy}(f)$ is the average cross product with a 90° phase shift between the two signals. This definition of the csd reduces to that for the power spectral density (psd) in the case of a single signal since $Q_{xx}(f) = 0$.

The csd can also be represented by a magnitude $|G_{xy}(f)|$ and a phase angle $\Phi_{xy}(f)$ in the following form:

$$G_{xy}(f) = |G_{xy}(f)|\, e^{-j\Phi_{xy}(f)}$$

$$|G_{xy}(f)| = [C^2_{xy}(f) + Q^2_{xy}(f)]^{1/2}$$

where (3-166)

$$\Phi_{xy}(f) = \tan^{-1}\left[\frac{Q_{xy}(f)}{C_{xy}(f)}\right]$$

Figure 3-18 presents the csd of some typical signal pairs.

There are two approaches to calculating the csd, and these are extensions of the two methods used to calculate the psd as described in Section 3.2.3. The direct method involves the computation of the Fourier Transforms of the two signals. This method is designed to minimize computation time at the expense of introducing some statistical errors, and is better suited for analyzing stationary signals of sufficient duration to allow for averaging over a large number of sample records. The indirect method is based on the relationship of the csd to the cross correlation function (ccf). This method is designed to minimize

statistical errors in the processing of the data at the expense of some additional computation time, and is better suited for analyzing signals of limited duration.

3.3.3.1 Direct Method

The direct method of calculating the csd of N value pairs (x_n, y_n) sampled with a time interval τ is based on the principles of the Fourier Transform (FT) (see Section 3.2.3.1). If $X_T(f_k)$ and $Y_T(f_k)$ are the FT of x_n and y_n, respectively, at frequencies $f_k = k/N\tau$; $k = 0, 1, \ldots, N-1$; then the csd is given by:

$$G_{xy}(f_k) = \frac{2}{T} X_T^*(f_k)\, Y_T(f_k) \qquad (3\text{-}167)$$

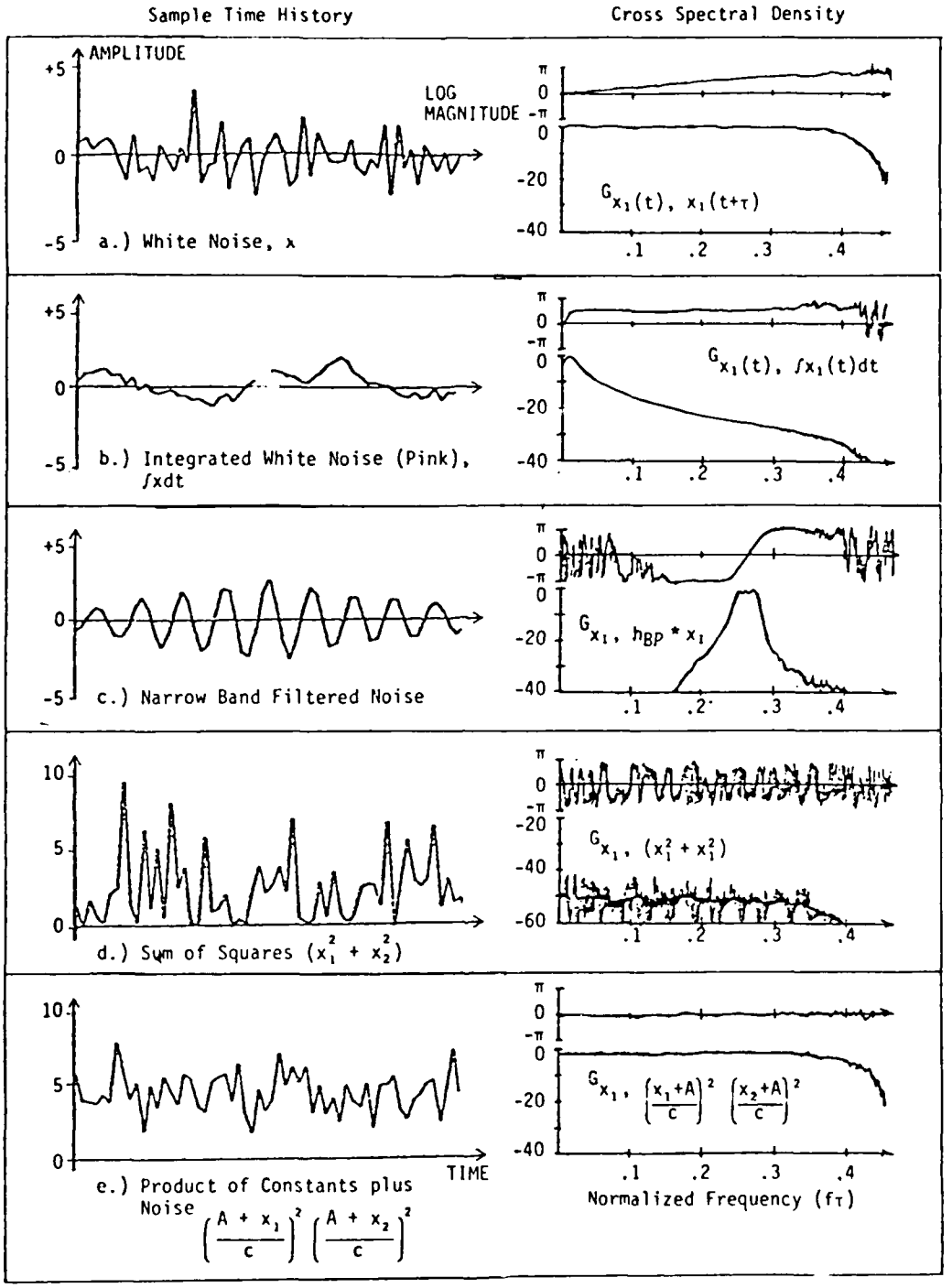


Figure 3-18. Cross Spectral Density of Some Typical Signal Pairs

where * denotes the complex conjugate and $T = N\tau$ is used as a subscript to denote a finite duration estimate of the FT.

USING THE FFT: The FFT can be used to efficiently compute the csd of the value pairs if $N = 2^p$. The sample values are stored in a complex array Z_n for input to the FFT in the format:

$$Z_n = x_n + jy_n \; ; \quad n = 1, 2, \ldots, N. \tag{3-168}$$

After the FFT operations on the input data, the FT's $X_T(f_k)$ and $Y_T(f_k)$ can be recovered from the output array Z_k by the formulas:

$$X_T(0) = \tau \, \mathrm{Re}(Z_1) + j0$$

$$Y_T(0) = \tau \, \mathrm{Im}(Z_1) + j0$$

$$X_T(f_k) = \frac{\tau}{2} \left(Z_{k+1} + Z^*_{N-k+1}\right)$$

$$Y_T(f_k) = \frac{\tau}{2j} \left(Z_{k+1} - Z^*_{N-k+1}\right)$$

$$f_k = \frac{k}{T} \; ; \quad k = 1, 2, \ldots, N-1. \tag{3-169}$$

The csd $G_{xy}(f_k)$ can then be evaluated using Equation (3-167). The psd of each signal can be evaluated using Equation (3-63).

DATA PREPARATION: The two sample records must be prepared for the FFT computation in a manner similar to that for a single record as outlined in Section 3.2.3.1. First, the mean value and slope of each sample record must be removed. Then, each sample record must be multiplied by a smooth window function such as those shown in Figure 3-14.

3.3.3.2 Indirect Method

The indirect method of computing the csd is based on its relationship to the cross correlation function (ccf). The csd $G_{xy}(f_k)$ and the ccf $R_{xy}(m\tau)$ are related by the complex Fourier series:

$$G_{xy}(f_k) = 2\tau \sum_{m=-M}^{M} R_{xy}(m\tau)\, e^{-j2\pi km/N}$$

$$R_{xy}(m\tau) = \frac{1}{2N\tau}\, G_{xy}(0) + \frac{1}{N\tau} \sum_{k=1}^{K} \mathrm{Re}\left[G_{xy}(f_k)\, e^{j2\pi km/N}\right] \qquad \text{(3-170)}$$

where N is the number of data values sampled with period τ, $f_k = k/N\tau$, and $e^{j\theta}$ is the exponential notation for the complex number defined by $e^{j\theta} = \cos\theta + j \sin\theta$.

The summation limits in Equation (3-170) are $M = \frac{N}{2} - 1$ and $K = \frac{N}{2} - 1$.

The ccf can be evaluated by one of two methods. The first is the direct summation method outlined in Section 3.3.2, Equation (3-151). This method is computationally slow because it requires approximately 1/2 MN operations. The second method is to use the FFT to compute the ccf as follows:

1. Extend both sample records to a length of 2N values by adding N zeroes to the end of both records. Do not window the data.

2. Compute the sample csd with a 2N point FFT giving a frequency interval $\Delta f = 1/2\ N\tau$.

3. Average the sample csd over ensembles of sample records.

4. Compute the (biased) ccf with a 2N point inverse FFT of the average csd giving correlation time delays from 0 to $\pm(N-1)\tau$.

5. Multiply the biased ccf by the factor $\frac{N}{N-|m|}$; $|m| \leq N-1$; to obtain the unbiased ccf.

The ccf will not, in general, go to zero at time delays $\pm(N-1)\tau$. Therefore the ccf must be multiplied by a time window, such as one shown in Figure 3-14, before using it to compute the csd. However, this is complicated because the ccf will not necessarily have its maximum value at zero time delay as with the acf of a single time series. Therefore it may not be suitable to have a window centered at zero time delay.

A time delay in the maximum value of the ccf indicates that there is a time delay between the correlated components of the two signals (due to propagation delay or reverberation in the system being measured). If this time delay is a significant fraction of the record length $T = N\tau$, then part of the correlated components of the two signals may not be contained in the same record length. This will lead to errors in the computation of the csd (for both the direct and indirect methods). This can be corrected if it is possible to shift the sampling of one of the signals in order to bring the maximum value of the ccf to zero time delay. Then the two signals can be resampled with this time delay between respective sample records and a new ccf computed. A time window centered at zero time delay can then be applied to this ccf.

The csd can then be computed from the windowed ccf. If a window of length $T = N\tau$ is used, then the ccf is compressed to one-half its original length and the csd will have a frequency interval $\Delta f = 1/N\tau$.

3.3.3.3 Time Delay Corrections

Time delays between the sampling of two channels can most easily be corrected in the frequency domain by altering the phase $\Phi_{xy}(f)$ of the csd. This applies to time delays that are fractions of a sample interval (resulting from multiplexing in the A/D conversion process) and also to time delays that are many sample intervals long (resulting from intentional time shifts to move the ccf peak value to zero delay).

A time shift of t_s seconds in one of the two signals will produce a phase shift in the csd given by:

$$\phi_s(f) = 2\pi f t_s \text{ (radians)}. \qquad (3\text{-}171)$$

The phase of the measured csd can then be corrected by subtracting the value of this function at each frequency. The magnitude of the csd does not need correcting.

3.3.3.4 Calibration and Units

If the two signals x and y have units of measure U_x and U_y, respectively, then the csd magnitude $|G_{xy}(f)|$ has units of $U_x U_y$/Hz (Hz = cycles/sec). The csd must also be calibrated both in magnitude and phase.

A convenient method of calibrating the csd computed from an FFT of a sampled signal is to use a white noise voltage signal of known bandwidth f_u and known mean square value $\overline{x^2}$ as described in Section 3.2.3.3. If the same signal is used for the two sample records x and y, then the csd is given by:

$$|G_{xy}(f)| = \frac{\overline{x^2}}{f_u}, \qquad (3\text{-}172)$$

$$\Phi_{xy}(f) = 0 \; ; \quad f < f_u$$

The computed magnitude of the csd can be calibrated to this value with units of $(\text{volts})^2/\text{Hz}$. The units of the csd of two measured signals can then be obtained by multiplying the computed csd magnitude by the sensitivity of each of the instrumentation systems in (units/volt). If the phase is not zero, but a linear function of the frequency, then a phase shift between the two sample records must be corrected as described in the previous Section.

It may also be necessary to calibrate the phase characteristics of the two instrumentation systems. This can be done by constructing an artificial environment in which the two measuring transducers are detecting the same signal. The measured csd phase in this calibration test can then be subtracted from subsequent measurements in order to obtain the correct phase.

3.3.3.5 Coherence Function

The coherence function is a measure of the frequency distribution of the relationship between two signals. The coherence function is denoted by $\gamma^2_{xy}(f)$ and is given by:

$$\gamma^2_{xy}(f) = \frac{|G_{xy}(f)|^2}{\overline{G}_x(f)\overline{G}_y(f)} \qquad (3\text{-}173)$$

where the $\overline{G}$ are ensemble averages of the power or cross spectral densities.

The magnitude of the coherence has the property that

$$\gamma^2_{xy}(f) \leq 1. \qquad (3\text{-}174)$$

When $\gamma^2_{xy}(f) = 1$ the two signals are said to be fully coherent at that frequency. When $\gamma^2_{xy}(f) = 0$ the two signals are said to be incoherent (statistically independent) at that frequency. Figure 3-19 presents the coherence function for the examples in Figure 3-18.

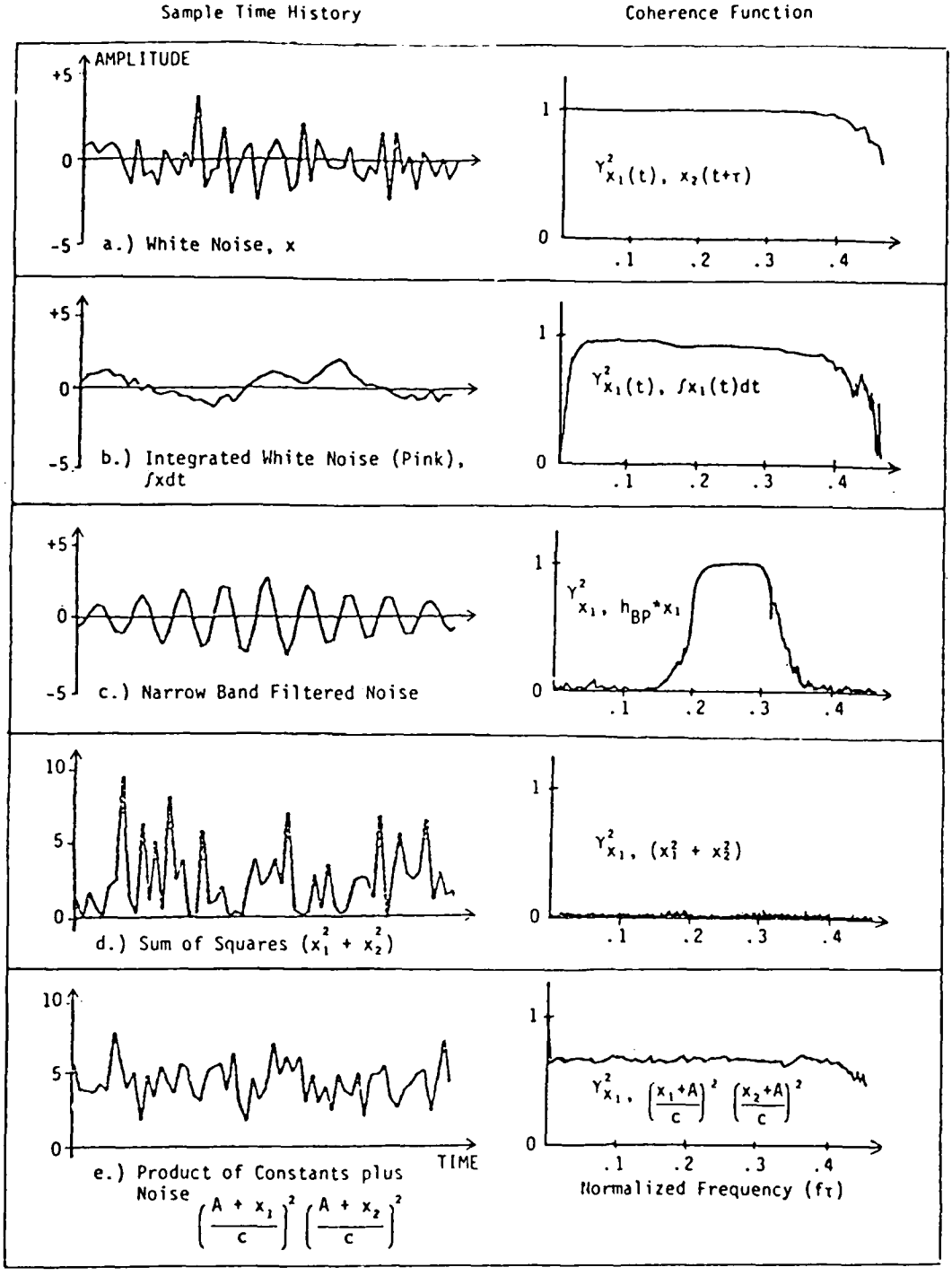


Figure 3-19. Coherence Function of Some Typical Signal Pairs

The coherence function can be used to determine the percentages of each signal which is linearly related to the other or to a common source. The magnitude of the coherence will be reduced by the presence of background noise in the measurement of the two signals or by non-linearities in the system being measured.

The spectral densities used in Equation (3-173) must be averaged over an ensemble of sample records. If the coherence is calculated using only one sample record, the value will always be unity. This can be seen from the definitions of the power and cross spectral densities which give

$$\frac{|G_{xy}(f)|^2}{G_x(f)G_y(f)} = \frac{X(f)Y^*(f)X^*(f)Y(f)}{X^*(f)X(f)Y^*(f)Y(f)} = 1. \qquad (3\text{-}175)$$

3.3.3.6 Frequency Response Functions

In many physical systems it is desirable to measure a frequency response function defined by:

$$H_{y|x}(f) = \frac{G_{xy}(f)}{G_x(f)} \qquad (3\text{-}176)$$

The magnitude of this frequency response function is given by:

$$|H_{y|x}(f)|^2 = \frac{G_y(f)}{G_x(f)} \qquad (3\text{-}177)$$

If x is the input to a linear system and y is the output, then the frequency response function is a characteristic of the system alone and is independent of the spectrum of the input. More detailed information on the types and uses of frequency response functions is included in Chapters 5-7 of this Compendium.

In order to average the frequency response function over an ensemble of sample records it is necessary to first average G_{xy} and G_x separately and then compute the ratio. Therefore,

$$\overline{H_{y|x}}(f) = \frac{\overline{G_{xy}}(f)}{\overline{G_x}(f)} \tag{3-178}$$

3.3.3.7 Statistical Errors in the CSD Estimate

The statistical errors in the estimation of the csd can be evaluated in a manner similar to that for the psd (Section 3.2.3.4). The bias error for the real part, imaginary part, and magnitude of the csd are given by [3-4]:

$$B[C_{xy}(f)] \simeq \frac{f_b^2 \, C''_{xy}(f)}{24} \tag{3-179}$$

$$B[Q_{xy}(f)] \simeq \frac{f_b^2 \, Q''_{xy}(f)}{24} \tag{3-180}$$

$$B[|G_{xy}(f)|] \simeq \frac{f_b^2 \, |G_{xy}(f)|''}{24} \tag{3-181}$$

where f_b is the effective bandwidth of the FFT window used.

The variances of these quantities are dependent on the amount of averaging used in their computation. If the csd is averaged over an ensemble of M sample records and K adjacent frequency points, the variances are given by [3-4]:

$$V[\overline{C}_{xy}(f)] \simeq \frac{G_x(f)G_y(f)}{2MK} \tag{3-182}$$

$$V[\overline{Q}_{xy}(f)] \simeq \frac{G_x(f)G_y(f)}{2MK} \tag{3-183}$$

$$V[|G_{xy}(f)|] \simeq \frac{|G_{xy}(f)|^2}{2MK}\left(1 + \frac{1}{\gamma_{xy}^2}\right) \tag{3-184}$$

The variances of the averaged coherence can be derived from those of the psd and csd as follows [3-4]:

$$V[\overline{\gamma_{xy}^2}(f)] \simeq \frac{2\gamma_{xy}^2(1 - \gamma_{xy}^2)^2}{MK} \tag{3-185}$$

3.3.4 Cepstral Analysis

The cepstrum of a frequency response function is defined as the inverse FT of the log of frequency spectrum. Although the cepstrum is a function in the time domain, the parameters of the cepstrum have historically been given names which are derived from clever reversals of syllables in the comparable parameters in the frequency spectrum [3-17]. (For example, "time" is replaced by "quefrency"). This has its origin in the use of the cepstrum for echo detection.

The use of the cepstrum can be seen by considering the output spectrum $G_y(f)$ of a system with a transfer function $H(f)$ and an input spectrum $G_x(f)$ so that:

$$G_y(f) = |H(f)|^2 G_x \tag{3-186}$$

Then the log of the output is given by

$$K_y(f) = \log G_y(f) = \log |H(f)|^2 + \log G_x \tag{3-187}$$

By taking the inverse Fourier Transform $(FT)^{-1}$ of Equation (3-187), the cepstrum $k_y(q)$ is achieved:

$$k_y(q) = FT^{-1} \{\log |H(f)|^2\} + FT^{-1} \{\log G_x\} \qquad (3\text{-}188)$$

This is called the "real" cepstrum, because it is the FT^{-1} of a real, even function.

If $\log G_x$ is a smooth function of time, its FT^{-1} will be concentrated at small values of q ("low time"). If $\log |H(f)|^2$ has some irregularities or ripples due to echoes or reverberation in the system, its FT^{-1} will have the associated contribution at larger values of q ("high time"). The contributions from this echo or reverberation can then be separated out by "low time" windowing the cepstrum and computing the log spectrum of the "dereverberated" signal through the FT of the windowed cepstrum.

The "complex" cepstrum is defined as the FT^{-1} of the log of the complex transfer function $H(f) = |H(f)|e^{j\phi}$ where ϕ is the phase angle of the H(f). Then

$$K_H(f) = \log |H(f)| + j\phi \qquad (3\text{-}189)$$

and the complex cepstrum is found from

$$k_H(q) = FT^{-1} \{K_H(f)\} \qquad (3\text{-}190)$$

This can be computed with a complex inverse FFT routine similar to that used for 2-channel signal processing where the complex array $K_H(f)$ is used as the input and the complex array $k_H(q)$ is the output. One difficulty that arises in the computation of Equation (3-190) is in obtaining the correct value of ϕ. The value of ϕ computed by the trigonometric relationship:

$$\phi(f) = \tan^{-1} \frac{\mathrm{Im}[H(f)]}{\mathrm{Re}[H(f)]} \qquad (3\text{-}191)$$

only gives the principal value in the range $-\pi < \phi < \pi$. Therefore it is necessary to "unwrap" the computed phase to obtain the true value.

More detailed information on these procedures can be found in References [3-3, 3-10].

3.4 MULTIPLE TIME SERIES

In this section we present various methods of analyzing the interrelationships of multiple signals sampled over the same time period. The methods described in the previous two sections for single and dual time series apply equally well to individual signal or pairs of signals in this section. In fact, this section is basically a generalization of the dual series analysis with the addition of a few special methods for handling the multichannel data efficiently.

3.4.1 Multiple Statistical Measures

This section is an extension of Section 3.3.1 which is limited to two channel analysis. In this section it is necessary to distinguish between a series of values obtained from a single signal (called a sequence) and a set of values obtained from a single realization of a set of signals (called a vector). If M signals are each sampled N times over the same time period, then the data is represented by a sequence of N vectors of the form $(x_1, x_2, \ldots x_M)_n$. In matrix notation this sequence is denoted by:

$$\underline{X} = \begin{bmatrix} x_{11} & x_{12} \cdots x_{1M} \\ x_{21} & x_{22} \cdots x_{2M} \\ \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot \quad\quad \cdot \\ x_{N1} & x_{N2} \cdots x_{NM} \end{bmatrix} \qquad (3\text{-}192)$$

where x_{ij} is the ith sample of the jth signal.

3 4.1.1 Joint Probability Density Function

The joint pdf of a sequence of N vectors $\underline{x}_n$ is related to the number of vector samples $N_{12...M}$ whose values fall within a set of value intervals designated by the vector of center values $\underline{X}$ and interval widths $\underline{w}$:

$$p(\underline{X}) = \frac{N_{12...M}}{w_1 w_2 ... w_M N} \tag{3-193}$$

A procedure for computing this pdf for a sequence of N vectors can be obtained by an extension of Table 3-1. If all of the signals are mutually independent the joint pdf is the product of the individual pdf's:

$$p(\underline{X}) = p(\underline{X}_1)\ p(\underline{X}_2)\ \cdots\ p(\underline{X}_M) \tag{3-194}$$

Except in a few special cases, the matrix $N_{12...M}$ is difficult to represent because of its multidimension. More frequently the joint pdf is represented by a matrix of the joint pdf's of all possible value pairs (called a bivariate matrix). The moments of this pdf matrix are also bivariate matrices, the most common being the covariance matrix $\underline{c}_{xx}$ defined by:

$$\underline{c}_{xx} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1M} \\ c_{21} & c_{22} & \cdots & c_{2M} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ c_{M1} & c_{M2} & \cdots & c_{MM} \end{bmatrix} \tag{3-195}$$

where each term $c_{ij} = c_{x_i x_j}$ is defined by Equation (3-139).

3.4.1.2 Joint Cumulative Distribution Function

The joint cdf of a sequence of N vectors can be obtained by a multiple summation of the joint pdf as an extension of Equation (3-135) for the two signal case.

3.4.1.3 Conditional Probability Density Function

The conditional pdf of a sequence of N vectors is a measure of the joint pdf of a subset of the vector values given that the remaining values are fixed is calculated by:

$$p(X_1, \ldots X_q | X_{q+1}, \ldots X_M) = \frac{p(\underline{X})}{p(X_{q+1}), \ldots p(X_M)} \qquad (3\text{-}196)$$

where $p(X_i)$ is the marginal pdf of x_i given by

$$p(X_i) = \sum_{X_i=-\infty}^{\infty} p(\underline{X}) \qquad (3\text{-}197)$$

If all of the signals are mutually independent the conditional pdf is independent of the values of the fixed signals:

$$p(X_1, \ldots X_q | X_{q+1}, \ldots X_M) = p(X_1, \ldots X_q) \qquad (3\text{-}198)$$

3.4.1.4 Multiple Regression Analysis

The method of linear regression discussed in Section 3.3.1 for two signals is extended here for multiple signals. Multiple regression is used to obtain an estimate for the value of one signal (usually referred to as the "output" signal) based on the values of the other signals (usually referred to as the "input" signals). If N samples are obtained of M+1 signals (x_1, x_2, ... x_{M+1}), then a linear regression of the signal x_{M+1} on the values of the other signals is of the form:

$$x_{M+1} = a_o + a_1x_1 + a_2x_2 + \ldots + a_Mx_M + \varepsilon \qquad (3\text{-}199)$$

where the a_i are the coefficients to be determined and ε is the error.

For convenience an "output" signal y is defined by:

$$y = x_{M+1} - a_o = a_1x_1 + a_2x_2 + \ldots a_Mx_M + \varepsilon \qquad (3\text{-}200)$$

Evaluating this equation for the N samples gives the matrix form

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ a_M \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_N \end{bmatrix} \qquad (3\text{-}201)$$

or

$$\underline{Y} = \underline{X}\,\underline{A} + \underline{\varepsilon} \tag{3-202}$$

The values of the parameters $\underline{A}$ which minimize the sum of the squared errors ε_i^2 can be found by [3-4]:

$$\underline{A} = (\underline{X}^T\underline{X})^{-1}\,\underline{X}^T\underline{Y} = \underline{c}_{xx}^{-1}\,\underline{c}_{xy} \tag{3-203}$$

where X^T is the transpose of X (i.e., the subscripts are reversed). The matrix $\underline{c}_{xx}$ is the covariance matrix for the M "input" signals in the form of Equation (3-195) and $\underline{c}_{xy}$ is the covariance vector between the output and the input signals:

$$\underline{c}_{xy} = \begin{bmatrix} c_{1(M+1)} \\ c_{2(M+1)} \\ \cdot \\ \cdot \\ \cdot \\ c_{M(M+1)} \end{bmatrix} = \begin{bmatrix} c_{x_1y} \\ c_{x_2y} \\ \cdot \\ \cdot \\ \cdot \\ c_{x_My} \end{bmatrix} \tag{3-204}$$

The value of a_o is then found by:

$$a_o = \bar{x}_{M+1} - \sum_{i=1}^{M} a_1\bar{x}_i \tag{3-205}$$

3.4.1.5 Multiple Correlation Coefficient

The multiple correlation coefficient is a measure of the linear relationship between an output signal y and a set of input signals x_i; i = 1, 2, ... M. This multiple correlation coefficient, denoted by $r_{y|x}$, is computed by:

$$r^2_{y|x} = \frac{\underline{A}^T \underline{c}_{xy}}{c_{yy}} = \frac{\underline{c}^T_{xy} \, \underline{c}^{-1}_{xx} \, \underline{c}_{xy}}{c_{yy}} \qquad (3\text{-}206)$$

where $\underline{A}$ is defined by Equation (3-203). The value of $r^2_{y|x}$ can be interpreted as being the mean square value of the predicted output signal using linear regression divided by the total mean square value of the output signal. Therefore, its value is bounded by $0 \leq r^2_{y|x} \leq 1$. A value of $r^2_{y|x} = 1$ implies that y is perfectly correlated with a linear combination of the x_i, while a value of $r^2_{y|x} = 0$ implies that y is completely uncorrelated with all of the x_i. Other special cases of multiple correlation coefficients are given in Reference 3-6.

3.4.2 Multiple Correlation Relationships

This section is an extension of Section 3.3.2 on the cross correlation function for two-channel analysis. The correlation relationships between multiple signals are often expressed in terms of a matrix of cross correlation functions:

$$\underline{R}_{xx}(m\tau) = [R_{ij}(m\tau)] = \begin{bmatrix} R_1(m\tau) & R_{12}(m\tau) & \cdots & R_{1M}(m\tau) \\ R_{21}(m\tau) & R_2(m\tau) & \cdots & R_{2M}(m\tau) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ R_{M1}(m\tau) & R_{M2}(m\tau) & \cdots & R_{MM}(m\tau) \end{bmatrix} \quad (3\text{-}207)$$

Based on the characteristics of the ccf (see Section 3.3.2) the elements in $\underline{R}_{xx}(m\tau)$ are related by:

$$R_{ij}(-m\tau) = R_{ji}(m\tau) \quad (3\text{-}208)$$

In a similar manner the matrix of cross-covariance functions $\underline{C}(m\tau)$ relates the signals with the mean value of each removed. These types of matrices are called bivariate matrices since they contain parameters evaluated for all possible signal pairs.

Because the ranges of values for $R_{ij}(m\tau)$ and $C_{ij}(m\tau)$ can vary for each signal pair, it is difficult to compare relative degrees of correlation. Therefore, it is more convenient to use the bivariate matrix of the normalized correlation function, $\rho_{ij}(m\tau)$, defined by Equation (3-157), since these values are bounded by ± 1.

3.4.2.1 Multiple Correlation Function

In the case where a single output signal y is to be related to a set of input signals x_i; i = 1, 2, ... M; a multiple correlation function $\rho_{y|x}(m\tau)$ can be defined in a manner similar to that for $r_{y|x}$ in Section 3.4.1.5. For this case,

$$\rho^2_{y|x}(m\tau) = \frac{\underline{C}^T_{xy}(m\tau)\ [\underline{C}^{-1}_{xx}(0)]^T\ \underline{C}_{xy}(m\tau)}{C_{yy}(0)} \quad (3\text{-}209)$$

The value of $\rho^2_{y|x}(m\tau)$ is then a measure of the linear relationship of y to the x_i for various time delays, $m\tau$.

3.4.2.2 Partial Correlation Function

The partial correlation function is a measure of the relationship between two signals with the effects of other signals removed. It is related to the partial covariance function which is defined by, for example:

$$C_{ij|1}(m\tau) = C_{ij}(m\tau) - \frac{C_{1i}(m\tau)\ C_{1j}(m\tau)}{C_{11}(m\tau)} \qquad (3\text{-}210)$$

$C_{ij|1}(m\tau)$ is the cross covariance between x_i and x_j with their linear dependence on x_1 removed [3-4]. The last term in Equation (3-210) has the effect of removing that part of $C_{ij}(m\tau)$ which is due to the linear relationships of x_i and x_j to x_1.

The partial correlation function $\rho_{ij|1}(m\tau)$ is then defined by:

$$\rho^2_{ij|1}(m\tau) = \frac{C^2_{ij|1}(m\tau)}{C_{ii|1}(0)\ C_{jj|1}(0)}$$

$$= \frac{\left(\rho_{ij}(m\tau) - \rho_{1i}(m\tau)\rho_{1j}(m\tau)\right)}{\left(1 - \rho^2_{1i}(m\tau)\right)\left(1 - \rho^2_{1j}(m\tau)\right)} \qquad (3\text{-}211)$$

The partial correlation functions can be used to obtain an efficient evaluation of the multiple correlation function $\rho_{y|x}$ as follows. (Note: The dependence on $(m\tau)$ is omitted for convenience).

1. Evaluate the normalized correlation functions $\rho_{x_i y}$ between the output y and all inputs x_i; i = 1, 2, ... M.

2. Evaluate the normalized correlation functions $\rho_{x_1 x_i}$; i = 2, ... M.

3. Evaluate the partial correlation functions $\rho_{x_i y|x_1}$; i = 2, ... M by Equation (3-211).

4. Evaluate the partial correlation functions $\rho_{x_2 x_i|x_1}$; i = 3, ... M by Equation (3-211).

5. Evaluate the partial correlation functions $\rho_{x_i y|x_1 x_2}$; i = 3, ... M by

$$\rho^2_{x_i y|x_1 x_2} = \frac{\left(\rho_{x_i y|x_1} - \rho_{x_2 x_i|x_1}\, \rho_{x_2 y|x_1}\right)^2}{\left(1 - \rho^2_{x_2 x_i|x_1}\right)\left(1 - \rho^2_{x_2 y|x_1}\right)} \qquad (3\text{-}212)$$

6. Evaluate the partial correlation functions $\rho_{x_3 x_i|x_1 x_2}$; i = 4, ... M by

$$\rho^2_{x_3 x_i|x_1 x_2} = \frac{\left(\rho_{x_3 x_i|x_1} - \rho_{x_2 x_3|x_1}\, \rho_{x_2 x_i|x_1}\right)^2}{\left(1 - \rho_{x_2 x_3|x_1}\right)\left(1 - \rho_{x_2 x_i|x_1}\right)} \qquad (3\text{-}213)$$

7. Continue the process until the value of $\rho_{x_M y|x_1 x_2 \ldots x_{M-1}}$ has been evaluated.

8. Then the value of $\rho_{y|x}$ can be evaluated by

$$\rho^2_{y|x} = 1 - (1 - \rho^2_{x_1 y}) \prod_{i=2}^{M} (1 - \rho^2_{x_i y | x_1 \ldots x_{i-1}}) \qquad (3\text{-}214)$$

3.4.3. Multiple Spectral Analysis

This section is an extension of Section 3.3.3 on the cross spectral density function for two channel analysis. The spectral relationships between multiple signals are often expressed in terms of a matrix of cross spectral density functions:

$$\underline{G}_{xx}(f) = [Gij(f)] = \begin{bmatrix} G_{11}(f) & G_{12}(f) & \ldots & G_{1M}(f) \\ G_{21}(f) & G_{22}(f) & \ldots & G_{2M}(f) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ G_{M1}(f) & G_{M2}(f) & \ldots & G_{MM}(f) \end{bmatrix} \qquad (3\text{-}215)$$

The parameters on the diagonal of this matrix are real quantities while the remaining parameters are generally complex. Based on the characteristics of the csd (see Section 3.3.3), the elements in $\underline{G}_{xx}(f)$ are related by:

$$G_{ij}(f) = G^*_{ji}(f) \qquad (3\text{-}216)$$

where * denotes the complex conjugate.

In order to describe the relationships between the spectral content of the various signal pairs it is convenient to use a matrix of the coherence functions $[\gamma^2_{ij}(f)]$ defined in Section 3.3.3.5. Each value in this matrix measures the strength of the linear relationship between two signals at a particular fre-

quency. However, if there is a linear relationship between two signals, the coherence matrix cannot determine whether the two signals are directly related or whether they are each related to a common third signal. In order to determine this, use must be made of the multiple and partial coherence.

3.4.3.1 Multiple Coherence Function

In the case where the spectrum of a single output signal y is to be related to those of a set of input signals x_i; i = 1, 2, ... M; a multiple coherence function $\gamma^2_{y|x}(f)$ can be defined by:

$$\gamma^2_{y|x}(f) = \frac{\underline{G}^T_{xy}(f) \; [\underline{G}^{-1}_{xx}(f)]^T \; \underline{G}_{xy}(f)}{G_{yy}(f)} \qquad (3\text{-}217)$$

The value of $\gamma^2_{y|x}(f)$ is then a measure of the linear relationship of y to the x_i for a particular frequency f.

3.4.3.2 Partial Coherence Function

The partial coherence function is a measure of the relationship between two signals with the effects of other signals removed. It is related to the residual (or partial) spectral density function which is defined by, for example:

$$G_{ij|1}(f) = G_{ij}(f) - \frac{G^*_{1i}(f) \; G_{1j}(f)}{G_{11}(f)} \qquad (3\text{-}218)$$

$G_{ij|1}(f)$ is the residual cross spectral density between x_i and x_j with their linear dependence on x_1 removed [3-19]. The last term in Equation (3-218) has the effect of removing that part of $G_{ij}(f)$ which is due to the linear relationship of x_i and x_j

to x_1.

The partial coherence function $\gamma^2_{ij|1}(f)$ is then defined by [3-20]:

$$\gamma^2_{ij|1}(f) = \frac{|G_{ij|1}(f)|^2}{G_{ii|1}(f)\ G_{jj|1}(f)}$$

$$= \frac{\gamma^2_{ij}(f) - 2\ \mathrm{Re}\left(\dfrac{G_{ij}(f)G_{1i}(f)G^*_{1j}(f)}{G_{ii}G_{jj}G_{11}}\right) + \gamma^2_{1i}(f)\gamma^2_{1j}(f)}{\left(1 - \gamma^2_{1i}(f)\right)\left(1 - \gamma^2_{1j}(f)\right)} \qquad (3\text{-}219)$$

The residual spectral density functions can be used to obtain an efficient evaluation of the multiple coherence function $\gamma^2_{y|x}$ as follows. (Note: The dependence on (f) is omitted for convenience).

1. Evaluate the cross spectral densities $G_{x_i y}$ between the output y and all inputs x_i; i = 1, 2, ... M; and also G_{yy}.

2. Evaluate the cross spectral densities $G_{x_1 x_i}$; i = 1, 2, ... M; and also $\gamma^2_{x_1 y}$.

3. Evaluate the residual spectral densities $G_{x_i y|x_1}$; i = 2, ... M; and also $G_{yy|x_1}$.

4. Evaluate the cross spectral densities $G_{x_2 x_i|x_1 j}$; i = 2, ... M; and also $\gamma^2_{x_2 y|x_1}$.

5. Evaluate the residual spectral densities $G_{x_i y|x_1 x_2}$; i = 3, ... M; and also $G_{yy|x_1 x_2}$ by:

$$G_{x_i y|x_1 x_2} = G_{x_i y|x_1} - \frac{G^*_{x_2 x_i|x_1} G_{x_2 y|x_1}}{G_{x_2 x_2|x_1}} \quad (3\text{-}220)$$

6. Evaluate the cross spectral densities $G_{x_3 x_i|x_1 x_2}$; $i = 3, \ldots M$; and also $\gamma^2_{x_3 y|x_1 x_2}$ by:

$$\gamma^2_{x_3 y|x_1 x_2} = \frac{|G_{x_3 y|x_1 x_2}|^2}{G_{x_3 x_3|x_1 x_2} G_{yy|x_1 x_2}} \quad (3\text{-}221)$$

7. Continue this process until the value of $\gamma^2_{x_M y|x_1 x_2 \ldots x_{M-1}}$ has been evaluated.

8. Then the value of $\gamma^2_{y|x}$ can be evaluated by:

$$\gamma^2_{y|x} = 1 - (1 - \gamma^2_{x_1 y}) \prod_{i=2}^{M} (1 - \gamma^2_{x_i y|x_1 \ldots x_{i-1}}) \quad (3\text{-}222)$$

3.4.3.3 Multiple Frequency Response Function

The frequency response function defined in Section 3.3.3.6 can be generalized to multiple signals by defining a separate transfer function between each input signal x_i; $i = 1, 2, \ldots M$; and an output signal y as follows [3-20]:

$$H_1(f) = \frac{G_{1y|23 \ldots M}(f)}{G_{11|23 \ldots M}(f)}$$

$$H_2(f) = \frac{G_{2y|13 \ldots M}(f)}{G_{22|13 \ldots M}(f)}$$

$$H_M(f) = \frac{G_{My|12 \ldots M-1}(f)}{G_{MM|12 \ldots M-1}(f)} \qquad (3\text{-}223)$$

These transfer functions identify the unique frequency response characteristics of each input signal and are not necessarily equal to the transfer functions which would be obtained by having only one input signal acting at a time.

3.4.3.4 Statistical Errors in Multiple Time Series Analysis

The statistical errors in the analysis of multiple signals are very similar to those for two channels because most of the analysis is done with bivariate matrices. Therefore, the formulas presented in Section 3.3 for the various statistical errors are also applicable here with one exception. For those statistical measures which are conditioned to remove the dependence of other signals, the number of degrees of freedom are reduced accordingly. This can be accounted for in the formulas for statistical errors by reducing the effective number of ensemble averages used in the computation by the number of conditions placed on the computation. For example, the computation of $\gamma^2_{x_3y|x_1x_2}$ with 20 ensemble averages would have an effective number of averages equal to 18.

LIST OF REFERENCES

3-1. Stockham, T.G., "A/D and D/A Converters: Their Effect on Digital Audio Fidelity," Proc. of 41st Conv., Audio Eng. Soc., October 1971, New York, NY.

3-2. Widrow, B., "Statistical Analysis of Amplitude-Quantized Sampled-Data Systems," AIEE Trans. Appl. Indust., Vol. 81, January 1961, p. 555.

3-3. Oppenheim, A.V., and Schaefer, R.W., Digital Signal Processing, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1975.

3-4. Jenkins, G.M., and Watts D.G., Spectral Analysis and Its Applications, Holden-Day, Inc., San Francisco, CA, 1968.

3-5. Bendat, J.S., and Piersol, A.G., Random Data: Analysis and Measurement Procedures, John Wiley and Sons, Inc., New York, NY, 1965.

3-6. Merkle, R.G., "Statistical Measures, Probability Densities, and Mathematical Models for Stochastic Measurements," *AFFDL-TR-76-83, October 1976.*

3-7. Parzen, E., Modern Probability and Its Applications, John Wiley and Sons, Inc., New York, NY, 1960.

3-8. Cooley, J.W., and Tukey, J.W., "An Algorithm for the Machine Calculation of Complex Fourier Series," Math. Computation, Vol. 15, 1965, p. 297.

3-9. Welch, P.D., "The Use of the FFT for Estimation of Power Spectra: A Method Based on Averaging over Short Modified Periodograms," IEEE Trans. Audio Electroacoust., Vol. AU-15, No. 2, 1967, p. 70.

3-10. Rabiner, L.R., and Gold, B., Theory and Application of Digital Signal Processing, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1975.

3-11. Rabiner, L.R., "Techniques for Designing Finite Duration Impulse Response Digital Filters," IEEE Trans. Commun. Technol., Vol. COM-19, 1975, p. 188.

3-12. Rabiner, L.R., Gold, B., and McGonegal, C.A., "An Approach to the Approximation Problem for Nonrecursive Digital Filters," IEEE Trans. Audio Electroacoust., Vol. AU-18, 1970, p. 83.

3-13. Burrus, C.S., and Parks, T.W., "Time Domain Design of Recursive Digital Filters," IEEE Trans. Audio Electroacoust., Vol. AU-18, 1970, p. 141.

3-14. Steiglitz, K., "Computer-Aided Design of Recursive Digital Filters," IEEE Trans. Audio Electroacoust., Vol. AU-18, 1970, p. 123.

3-15. Rabiner, L.R., and Steiglitz, K., "The Design of Wide-Band Recursive and Nonrecursive Ditital Differentiators," IEEE Trans. Audio Electroacoust., Vol. AU-18, 1970, p. 204.

3-16. Crochiere, R.E., and Rabiner, L.R., "Optimum FIR Digital Filter Implementations for Decimation, Interpolation and Narrow-Band Filtering," IEEE Trans. Acoust. Speech Signal Proc., Vol. ASSP-23, No. 5, 1975, p. 444.

3-17. Bogert, B.P., Healy, M.J.R., and Tukey, J.W., "The Quefrency Alanysis of Time Series for Echoes: Cepstrum, Pseudo-Autocovariance, Cross-Cepstrum, and Saphe Cracking," Proc. Symp. Time Series Analysis, M. Rosenblatt, Ed., John Wiley and Sons, Inc., New York, NY, 1963.

3-18. Harris, F.J., "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," Proc. IEEE, Vol. 66, No. 1, 1978, p. 51.

3-19. Dodds, C.J., and Robson, J.D., "Partial Coherence in Multivariate Random Processes," J. Sound Vib., Vol. 42, No. 2, 1975, p. 243-249.

3-20. Bendat, J.S., "Solutions for the Multiple Input/Output Problem," J. Sound Vib., Vol. 44, No. 3, 1976, p. 311-325.

CHAPTER 4. TEST PROGRAM DESIGN

By

R. E. POWELL
J. A. MOORE
J. E. MANNING

CHAPTER 4. TEST PROGRAM DESIGN

TABLE OF CONTENTS

		Page
	TABLE OF CONTENTS	i
	LIST OF TABLES	iii
	LIST OF SYMBOLS	iv
4.1	OVERVIEW OF TEST PROGRAM DESIGN	4-1
4.2	PROBLEM DESCRIPTION	4-2
	4.2.1 Identification of the Problem	4-2
	4.2.2 History of the Problem	4-2
	4.2.3 Extent of the Problem	4-2
	4.2.4 Impact of the Problem	4-3
	4.2.5 Suggested Scope of the Test Program	4-3
4.3	PROGRAM OBJECTIVE	4-3
	4.3.1 Problem Review and Statement	4-4
	4.3.2 Primary Program Objective	4-4
	4.3.3 Secondary Objectives	4-5
4.4	DATA REQUIREMENTS	4-5
	4.4.1 Test Identification	4-7
	4.4.1.1 Tests to Quantify Vibration or Acoustic Levels	4-7
	4.4.1.2 Tests to Acquire Diagnostic Information on Vibration Sources and Transmission Paths	4-8
	4.4.1.3 Tests to Verify a Design Which Represents a Possible Problem Solution	4-8

4.4.2 Measured Data..... 4-8

4.4.3 Analytical Models..... 4-9

4.5 DEVELOPMENT OF TEST PLANS..... 4-10

4.5.1 Test Objectives..... 4-10

4.5.2 Test Configuration..... 4-10

4.5.3. Measurement Types and Sensor Locations..... 4-12

4.5.4 Instrumentation Requirements..... 4-13

4.5.4.1 Moderate Accuracy Requirements..... 4-14

4.5.4.2 High Accuracy Requirements..... 4-15

4.5.4.3 Extreme Accuracy Requirements..... 4-17

4.5.5 Data Analysis Requirements..... 4-18

4.5.5.1 Number of Channels to Acquire Simultaneously..... 4-19

4.5.5.2 Analog Filtering Prior to Sampling... 4-20

4.5.5.3 Sampling Rate..... 4-20

4.5.5.4 Continuous Data Lengths..... 4-20

4.5.5.5 Total Data Lengths..... 4-21

4.5.5.6 Digital Filter Types..... 4-21

4.5.5.7 Fast Fourier Transform (FFT) Lengths. 4-21

4.5.5.8 Window Types..... 4-22

4.5.5.9 Number of Averages..... 4-22

4.5.6 Test Report Requirements..... 4-23

4.6 DOCUMENTATION OF RESULTS..... 4-23

TABLES..... 4-25

REFERENCES..... 4-35

CHAPTER 4. LIST OF TABLES

		Page
4.1	Primary Task Areas.....	4-25
4.2	Problem Description.....	4-27
4.3	Program Objective.....	4-29
4.4	Data Requirements.....	4-30
4.5	Test Plan Development.....	4-32
4.6	Documentation of Results.....	4-34

CHAPTER 4. LIST OF SYMBOLS

A/D	Analog to Digital
DC	Direct Current
FFT	Fast Fourier Transform

CHAPTER 4. TEST PROGRAM DESIGN

This chapter presents the procedures to be followed in designing a test program for collecting vibration and acoustic data. Its purpose is to link earlier chapters on measurement and data analysis technologies with subsequent chapters in which specific methods for applying measured data to vibration and acoustic problems are presented.

4.1 OVERVIEW OF TEST PROGRAM DESIGN

A good test program is an important ingredient in the overall success of a project because it facilitates the coordination and cooperation of all personnel involved. It requires interaction between those specifying the data to be collected and those who will actually carry out the data collection and analysis.

The primary tasks for test program design are listed in Table 4.1. These include problem description, program objectives, data requirements, detailed test plans, and documentation of results. Many test programs fail because attention is concentrated on the detailed test plan with too little attention being given to the other tasks of the design. Program objectives and data requirements may be poorly defined or the data collected may be poorly documented. In either case an expensive and time consuming test may provide little useful information even though the test itself is properly carried out.

In the following Sections each task to be carried out in designing a test program will be discussed. Tables 4.1 through 4.6 present summary material in outline form. These tables will provide a convenient guide to follow in preparing a good test plan.

4.2 PROBLEM DESCRIPTION

Most test programs are oriented toward the solution of a perceived problem. The first step in test program design is to present the problem as clearly as possible in a problem description. Typically, this is done by the organizational unit actually experiencing the problem. It is important that this group convey the necessary information to those who will plan and carry out the test program. In the following paragraphs the basic elements of a problem description are presented. Table 4.2 is included as a summary of the problem description format.

4.2.1 Identification of the Problem

The problem description should concisely identify the problem in the context of the affected system(s). Evidence of an actual or potential problem should be clearly presented, and any special terminology which is used to describe the problem should be defined. Excessive detail in describing the structure or system should be avoided at this point in the program description.

4.2.2 History of the Problem

A history of the problem should be included, giving a chronology of events related to the problem. Relevant test reports and other documentation should be referenced in this section. A discussion of any previous testing programs or remedial actions taken can be particularly valuable in describing the problem.

4.2.3 Extent of the Problem

Additional background information describing the details of the problem and its relation to various structures and functions in the system may be included in this section. Hypotheses as to the cause of the problem or other information which may be

helpful in solving the problem should also be included. For example, if the test program is expected to produce design changes to a component, then design constraints caused by related systems and functions should be described.

4.2.4 Impact of the Problem

The problem description must also provide the motivation for the test program by describing the impact of the problem on affected systems. Approximate cost estimates (which need not be specific) will assist the test designers in judging the importance and urgency of solving the problem quickly. Cost savings and improved performance or safety which would result from a successful solution to the problem should be mentioned here.

4.2.5 Suggested Scope of the Test Program

The problem description should conclude with a brief summary of the proposed scope of the testing program. Such a discussion may have to be rather general at the beginning of the project, since the subsequent program plan will be providing specific information. Some of the areas which could be mentioned are: types of measurements, testing environments, data analysis methods, and analytical or statistical models to be used.

4.3 PROGRAM OBJECTIVE

The program objective is a discussion of the goals of the test program. It is written after a review of the problem description and serves as a bridge between the group actually experiencing the problem and the group responsible for designing and implementing the test plan. In addition to reviewing the problem description, the program objective must specify the primary goal of the program including criteria that will be used

to evaluate the solution. The three areas of the program objective are outlined in Table 4.3 and described below.

4.3.1 Problem Review and Statement

The first step in forming the program objective is a review of the problem description. If there are any points which require further clarification or additional information, these difficulties should be resolved by communication with the author of the problem description. The key points of the problem description should be summarized into a one or two paragraph problem statement. This statement may be used as a basis for the background section of subsequent test plans and reports.

An important element of the problem review and statement is the identification of any constraints which have been placed on the test program. Examples of such constraints would be: time constraints such as limited availability of the structure for testing, design constraints such as limited weight addition for damping treatment, or budget constraints which limit the extent of measurements and data analysis.

4.3.2 Primary Program Objective

The second part of the program objective is the actual statement of the primary program objective. After reviewing the problem and discussing possible solutions, a consensus must be achieved on the nature of an acceptable solution to the problem. It is desirable that the nature of the solution be described here (as the primary program objective) so that the preparation and execution of the program plan may be guided by a well-defined goal.

Realistically, the solution of the problem cannot be fully specified before testing. What is needed is a description of the character of the desired solution, including the results which are to be achieved and the quantitative measure which will be used to evaluate the results. A particularly concise example

of a primary program objective was contained in the "Preliminary Test Plan for F-4G UHF Blade Antenna Tests" (1978) by J. Pearson:

> "Primary Objective: The primary objective of this test is to determine the in-flight environment being experienced by the UHF blade antenna as installed on the nosewheel door, to determine the cause of the failures, and to develop a modification for present antennas and a new vibration specification for new antennas which will ensure 1000 hours of useful life".

4.3.3 Secondary Objectives

In stating the program objectives it is important to distinguish between the primary objective and other secondary objectives. The statement of secondary objectives is an opportunity to describe objectives that are desirable but not critically important to the success of the program. The description should include a discussion of the relative importance of each objective. With this description the test engineer can design a flexible test plan without losing sight of the overall program objectives.

4.4 DATA REQUIREMENTS

Chapters 5, 6, 7, and 8 of this report present a number of methods for applying measured data to vibration and acoustic problems. A key step in developing a test program design is to select those methods that can be used to reach a problem solution. Since the different methods have different data requirements, it is important to select the methods that will be used before proceeding with the detailed test design.

In most cases a satisfactory problem solution will require the use of several different methods. Often more than one type

of test will be needed to obtain the necessary data. For example, it is often necessary to conduct both flight and ground tests in order to reach a satisfactory problem solution.

The data requirements section of the test program plan presents a general description of the tests to be performed, the data that are to be collected, and the analytical models that are to be used. It includes a discussion of the motivation in performing the different tests and how they relate to each other in the context of the overall test program. It involves the identification of methods and tests which will be used to meet the primary program objective. Appropriate methods must be fit together into a coherent plan for solving the problem.

Again taking as an example the antenna study, with a program objective quoted in Section 4.3.2, a data requirements section of the program plan could read as follows:

> Data Requirements: A flight test of an F-4G with an instrumented antenna will be made. Antenna vibration levels and dominant frequencies are to be determined from the flight data by spectrum analysis. A laboratory shake test will then be performed on an individual antenna so that strain levels may be determined under simulated in-flight vibration. Fatigue curves of antenna materials will be used to predict life at the maximum alternating stress amplitudes found in the shake test, and to calculate the stress reduction necessary to at least triple this life. As part of the laboratory testing, the frequencies, damping, and mode shapes of antenna resonances which contribute to the measured in-flight vibration will be determined by modal analysis. The results will be used to design a modification to the antenna which will achieve the desired stress reduction and life improvement. The modified antenna will be tested on a shake-table to verify the fix and, if necessary, an additional flight test will be scheduled to evaluate the design.

The topics to be considered in preparing the data requirements are listed in Table 4.4 and discussed in more detail below.

4.4.1 Test Identification

The first step in preparing the data requirements is to identify the tests that are required. Three types of tests should be considered: (1) tests to quantify vibration and/or acoustic levels, (2) tests to provide diagnostic information on vibration sources and transmission paths, and (3) tests to verify a design which represents a possible problem solution.

4.4.1.1 Tests to Quantify Vibration or Acoustic Levels

Tests to provide data on vibration and acoustic levels are used to develop a quantitative characterization of the problem. Although the problem description contains information of this nature, more information will be required in many cases. For example, a previous test may have reported the vibration amplitude at a critical location but neglected the frequency content. The data may be available for re-analysis or another test may be necessary.

Quantifying the problem is important in providing a benchmark to judge how severe it is and the extent to which it must be improved. The success of any solutions developed will be evaluated by comparison with pre-solution results. This important comparison is best made with data of known quality.

Consider the antenna example above. As presented the problem description may indicate observed in-service failure rates. Program cost or scheduling constraints may not allow the repeated in-service monitoring necessary to document the failure rates of candidate solutions. A single in-service survey of the noise or vibration levels that the component is exposed to quantifies the problem so that laboratory testing of the component can provide a comparable environment for studying the failure or for judging the performance of solution designs.

4.4.1.2 Tests to Acquire Diagnostic Information on Vibration Sources and Transmission Paths

The second type of tests to be considered involve diagnostic investigations of sources and transmission paths. The purpose of these tests is to build an understanding of the problem so that the appropriate models describing the system's dynamic behavior can be chosen. The methods used in diagnostic tests include those of estimation of system characteristics, identification of transmission paths, and identification of sources. In the antenna example, diagnostic tests for frequency response functions will be needed as part of the modal analysis method selected for describing the problem on this small part.

4.4.1.3 Tests to Verify a Design Which Represents a Possible Problem Solution

The third type of test is performed to verify the performance of a candidate solution. Although the use of measured vibration and acoustic data greatly improves the chances for success in solving a particular problem, a final test of a prototype design is often necessary. Ideally, such a test is carried out under a service environment. Often, however, a test under simulated conditions can provide a reasonable validation of the design.

4.4.2 Measured Data

The next step after identifying the tests that are to be performed is to identify the vibration and acoustic data that are to be measured. A detailed description of measurement locations and instrumentation is not needed at this point. However, it is important to set out the variables that are to be measured and to identify the methods that will be used in applying the measured data to the problem.

4.4.3 Analytical Models

The development of analytical models of dynamic behavior are important in interpreting the data. A model provides a basis for identifying key system parameters that affect the severity of the problem. A model also serves as a design tool for developing solution approaches. The development of a system model must take into consideration potential solution approaches in order that it may demonstrate the performance improvement associated with the solution.

Testing is often required to evaluate model parameters. In addition, laboratory tests on component subsystems or on idealized structures representative of the actual system are often important in refining the model. The measurements can focus on individual parameters in the model, thereby affording a simpler interpretation of results in comparison with measurements on the complete system. The structures tested can be subsections of the actual system or fabricated test modules which exhibit similar dynamic behavior as the actual system.

System damping levels are a typical area where parameter values are most appropriately defined based on measurements. Analytical models of damping exist for particular material configurations (e.g. constrained layer damping treatments), but most often measured values on actual structures or empirical estimates from data for comparable structures are the source for model damping values.

Verification of the model is important in establishing its credibility as the basis for designing problem solutions. This generally involves comparisons with measured data from the actual system under in-service operating conditions. An alternative to comparisons with in-service measurements with actual sources is comparisons with data from source simulation measurements.

It is sometimes sufficient to establish the accuracy of the model for the subsection of the total system in which the solution will be implemented. The reduced scope of the

comparison is appropriate for demonstrating the model as a basis for design of the solution and for evaluating the performance improvement. Engineering judgment is required in deciding how to limit the size of the subsection and the scope of the modeling and comparisons that are carried out.

Chapter 8 on system response prediction describes the methods which are most closely associated with this task area. In the antenna example, the method chosen for describing the system's dynamic behavior was modal analysis.

4.5 DEVELOPMENT OF TEST PLANS

Sections 4.4.1 through 4.4.3 identify the tests to be carried out. The program plan should include a detailed test plan for each test that completely describes the test conditions, measurements, and data analysis. A suggested test plan outline is given in Table 4.5.

4.5.1 Test Objective

The test objectives section of the measurement test plan provides the motivation for the test. It includes a description of the measurements which will be performed, the analysis methods which will be applied to the measured data, and the relation of the test results to the program objective. The objectives section should be brief and introductory as later sections of the test plan will describe the details of the test procedures.

4.5.2 Test Configuration

The first detailed section of the test plan describes the object to be tested and the environment for the test. The physical configuration of the test object should be described, including any variations in configuration to be tested. These configuration variations to be tested are often useful for

transmission path identification methods and for testing the effect of potential solution approaches. Backup material for describing the configuration could include drawings or photographs of the test object.

Mounting configurations for ground-based tests should be specified. Free boundary conditions can be simulated by hanging the object from bungee cords or by supporting it on air bags or soft urethane foam pads. Fixtures or jigs for vibration tests must be carefully designed, considering their effect on the object and the measurements. For system identification tests involving shaker excitation, the shaker attachment arrangement should be specified to minimize the effects of rotational inertia and mass below the force gauge. For damping measurements on lightly-damped structures, the shaker damping can be a significant addition to the system damping, so a provision for shaker removal or the use of impact excitation should be considered.

The range of operating conditions to be tested must also be described. For in-service tests, this description would include the maneuvers to be performed and the timing of the test to ensure adequate data lengths for each maneuver. The choice of operating conditions is extremely important for source identification methods. Some sources will be stronger than others during certain maneuvers, which makes them easier to identify in analyzing the data. For ground-based or laboratory tests the operating conditions are primarily the form of excitation to be used (natural, random, impact, etc). The total amount of data to be recorded and analyzed can increase greatly if many operating conditions and configuration variations are called for, so that the cost of the test must be considered in setting these options.

In complicated testing situations involving many people and large amounts of equipment, the configuration and operating conditions of the test site itself may need to be specified. Availability of personnel and equipment often requires careful

scheduling. Certain types of test may also require special safety equipment. Any special test site requirements such as these should be written into this section of the test plan.

4.5.3 Measurement Types and Sensor Locations

The test parameters specified in this section of the test plan concern the selection of the particular quantities to be measured, the number of sensors required, and the sensor locations on the test object. The choice of measurement quantity (acceleration, strain, pressure, etc) depends on the purpose of the test. System response measurements are always required, while source measurements such as force or aerodynamic pressure are necessary for many system identification methods and they may be useful in the identification of sources and transmission paths.

The choice of measurement quantity may also be used to ease the dynamic range requirements for the instrumentation system. A displacement measurement will often require a very high dynamic range in order to resolve high frequency vibration. In this case, an acceleration measurement instead of displacement would emphasize the higher frequencies. The goal is always to have the measured spectrum be as flat as possible over the frequency range of interest. Previously measured or predicted levels or spectra can be valuable in choosing the measurement quantity, as well as in choosing appropriate transducers and signal conditioners as described in the next section.

The choice of transducer locations is also guided heavily by previous experience. Conflicting with the desire to measure and record everything about the test structure are a number of practical limitations. Mounting difficulties are encountered when surfaces are curved or fragile. Environments of temperature and pressure extremes, nuclear radiation, or corrosive fluids may exist. Sensor output signals must be routed by cable or else telemetry will be required. Interference with normal and safe

functioning of the test object must be avoided. The number of channels available for transmitting or recording test data is always limited by the availability of these systems. Because of these limitations, the choice of sensor locations which will satisfy the test objective is usually a compromise.

In tests which will use artificial excitation, the locations for the exciter device(s) must also be specified. When the test object is large and massive compared to the vibration exciter, the attachment should be made at a location of low mechanical impedance (vibration antinodes) in order to maximize the power that can be delivered to the structure. If the test object is thin or light compared with the moving parts of the exciter, the attachment is better located in an area of higher impedance on the structure in order to avoid mass loading or rotational inertia effects.

All transducer and exciter locations should be carefully marked and identified by some code (letters and/or numbers) on drawings of the test structure. Key dimensions should be noted to aid in locating the proper points on the structure, especially if the drawings are not to scale. A table of location codes, coordinates, and a brief description of each location should also be included in the test plan.

4.5.4 Instrumentation Requirements

The instrumentation system for the test includes the measurement transducers, signal conditioners, and recording system, as well as the playback system to be used for data analysis. In this section of the test plan, the instrumentation system must be fully specified. The design techniques for instrumentation systems and their components are discussed in detail in Chapter 2 of this report. Section 2.4.2 on System Design includes the procedures for designing a system to meet the data accuracy requirements.

The data accuracy requirements of the methods described in this report for applying measured data to vibration and acoustic problems (listed in Table 4.4) can generally be placed in one of three classes: moderate, high, and extreme accuracy.

4.5.4.1 Moderate Accuracy Requirements

Methods using single time series analysis methods without combining measured results do not require as much accuracy as do those requiring the phase between dual time series or methods which use subtraction or division operations on measured data. These methods obtain satisfactory results with less stringent instrumentation requirements. The methods in this "moderate" instrumentation accuracy class include those using vibration or acoustic levels, power spectra, or correlation with single channel analysis. Examples are:

- Damping measurement by decay rate
- Transmission path determination by correlation
- Response level measurements for comparison with predictions, fatigue prediction, or vibration qualification tests.

The number of channels required for a test using these methods can be just as high as any test. The difference is that phase differences (small time delays) between the channels are not important. A number of small tape recorders could be used to record the necessary channels instead of one large, expensive one. In terms of frequency band requirements, it is sufficient to specify the amplitude response over the frequency band of interest. Instrumentation bandwidths are less of a problem when phase linearity is not important.

The dynamic range requirements are usually less than other methods, ranging from 30 dB to 40 dB, depending on the importance of the spectrum shape at lower levels. The upper end of the dynamic range is determined by the linearity of the

instrumentation chain. For the methods in this class, the small non-linearities that can be encountered from some transducers (1-5%) generally produce amplitude errors of the same order in the results. FM tape recorder overmodulation must still be avoided, as that produces far worse degradation. Tape transport flutter is unimportant unless discrete tones and sidebands are to be analyzed. The lower end of the instrumentation system dynamic range is set by noise and interference. These effects are less important for this class of methods, since the results are usually dominated by the largest components of the signal.

4.5.4.2 High Accuracy Requirements

The next class of methods for applying measured data to vibration and acoustic problems requires higher data accuracy to achieve satisfactory results. This class includes the methods which use dual time series analysis where interchannel phase is important. Also in this class are techniques which use further processing on the single time series results from the previous class. Some examples are:

- Frequency response estimation using single input excitation
- Cepstral analysis for transmission path identification
- Source identification methods using acoustic intensity, cross-correlation, or cross-spectrum methods.

The number of channels for these methods is usually two or more, with the phase between channels an important quantity. In a test with a large number of channels, the channels used as a pair for these methods must be recorded on the same tape recorder. If the recorder has multiple record heads, better results will be achieved if the channel pair is recorded on the same head.

Frequency bandwidths must be specified in terms of both amplitude and phase response for this class of methods. Sensor

and instrumentation phase responses often deviate from the ideal linear shape at frequency range extremes. Non-linear phase responses can be tolerated in most of the analysis methods as long as the phase differences between the channel pairs are small over the frequency range of interest (5 degrees is a typical spec). Exceptions to this rule are methods which require peak amplitudes versus time, such as shock responses or certain fatigue methods. Phase response must be linear over a very wide frequency range for good results with peak response methods.

Dynamic range requirements vary with different tests, but they are generally in the 40-60 dB range for the complete instrumentation system. Linearity is much more important for this "high accuracy" class of methods because of its effect on the coherence between two signals. If the structure being tested vibrates beyond the linear range of a response transducer, the coherence between that transducer and all the other channels becomes degraded. This effect causes errors in frequency response estimates and the coherence function.

For system identification techniques using impact excitation, the linearity and noise performance of the input signal channel are especially important. Amplifier overload may be difficult to detect after filtering, so particular attention must be paid to the sensitivity of the impact sensor. Noise signals in the input channel produce a biased estimate of the standard frequency response function (regardless of the type of excitation). Noise signals in the response channels simply increase the random error of the frequency response function estimates. The dynamic range of the input channels is therefore more important than that of the response channels for this class of methods.

Interference signals cause differing effects depending on whether they are correlated with the input signal. A correlated interference pickup (such as accelerometer base strain) will bias the frequency response function estimate. Uncorrelated pickup appears as noise to the analysis.

4.5.4.3 Extreme Accuracy Requirements

The most demanding class of methods includes those requiring simultaneous multiple channel analysis. In these methods, extreme data accuracy is needed. The phase response between many channels is critical. In addition, the measured results are often used in complex calculations involving subtraction, divison, or matrix operations. Small errors in the measurements can become magnified by the processing operations. The methods in this class which are covered in other report chapters are:

- Source identification by partial coherence
- Multiple input frequency response estimation

The number of channels for these methods is always three or more, with the phase between all channels an important quantity. In designing a recording system for such measurements, tape speed fluctuations and interchannel phase variations must be minimized. Digital recording systems are often necessary in order to meet phase response and dynamic range requirements.

Frequency bandwidth specifications are the same for this "extreme accuracy" class as for the high accuracy class, that is, both amplitude and phase response must be specified. The channel-to-channel variations that can be tolerated are somewhat smaller, with more precision needed as the number of sources increases.

Dynamic range requirements also depend on the number of sources or inputs. As the contribution from each source is removed from a measurement by the partial coherence analysis, the effective dynamic range of the remainder is reduced (the noise level remains constant while the signal level is reduced). For example, in the case of two uncorrelated sources contributing equally to a response measurement, the dynamic range is reduced by 3 dB when the contribution of one source is subtracted. If

the contribution from one source is 20 dB higher than that from the second, almost 20 dB of dynamic range is lost when the first is subtracted.

Instrumentation system linearity is also important for this "extreme accuracy" class of methods because of its effect on coherence functions. Errors caused by non-linearities in the test object or in the instrument chain result in portions of the measured signals that are not coherent with any source. Most multichannel analysis methods assume that the vibrational or acoustical propagation from inputs to outputs is linear, so they do not account for the incoherent signals caused by non-linearities.

The effect of noise on these analysis methods is much the same as the effect on dual-channel methods discussed above. An additional consideration in multi-input situations is that the random error of frequency response estimates increases in the presence of output noise or of additional uncorrelated inputs (see Sections 5.2.2 and 5.3.1.1). Contributions to a response signal from uncorrelated sources appear as noise when the frequency response function from one input is being calculated. Because of this effect, longer data records must be acquired so that the analysis can achieve the same accuracy as in the single-input case.

Interference signals cause similar effects for tnis class of analysis methods as discussed above for the "high accuracy" class. A correlated interference pickup (such as accelerometer base strain) will bias the results because it is indistinguish-able from the desired signal. Uncorrelated pickup appears as noise to the analysis.

4.5.5 Data Analysis Requirements

The data analysis procedures needed for the application of the chosen methods to the measured test data are specified in this section of the test plan. Types of basic and higher order

analyses as well as the detailed analysis parameters should be specified. Signal analyzers and computer software packages which are to be used should also be specified. Data presentation formats must also be settled on.

Typical analysis parameters to be specified include:

- Number of channels to acquire simultaneously
- Analog filtering prior to sampling
- Sampling rate
- Continuous data lengths
- Total data lengths
- Digital filter types
- FFT lengths
- Window types
- Number of averages

Each method for applying measured data to acoustic and vibration problems has its particular analysis steps, which are described in later report chapters. There are a number of basic analysis parameters, such as those listed above, that are common to many of the methods. Several of these parameters may also affect the test operations and the instrumentation requirements.

4.5.5.1 Number of Channels to Acquire Simultaneously

The number of channels which may be acquired simultaneously for analysis obviously depends on the type of hardware available to the analysts. Similar restrictions probably exist in the choice of A/D converter resolution and on the maximum sampling rate of the data acquisition system. Analysis methods that require the phase between channels must have at least two channel capability. Multichannel analysis methods can proceed more quickly if all of the recorded test channels can be digitized simultaneously, thereby eliminating the synchronization processing steps that would be required otherwise.

4.5.5.2 Analog Filtering Prior to Sampling

Anti-alias filtering of the playback signals prior to sampling is always necessary. Methods which require good linearity and wide dynamic range may require more attention to aliasing. A quick-look spectrum analysis of the playback signals over a wide frequency band will usually indicate if the filters are doing an adequate job. Spectrum components above half the sampling rate must be attenuated sufficiently to meet the dynamic range requirement of the analysis method. At the same time, the frequency bandwidth requirements must not be ignored, so the filters must have good amplitude and phase characteristics in the frequency band of interest.

4.5.5.3 Sampling Rate

The sampling rate for the analysis is often the most important parameter choice. In addition to the aliasing and frequency response criteria, some analysis methods require over-sampling for extra time resolution (especially correlation time delay methods) or require power-of-two sampling rates so that frequency bands are in round numbers after FFT spectrum estimation. Sampling rates too low sacrifice frequency bandwidth and time resolution, while overly high sampling rates increase the quantity of data to be handled and processed. A high sampling rate may cause the actual time length of continuous data to be reduced (due to finite storage) and thereby reduce the frequency resolution available for analysis.

4.5.5.4 Continuous Data Lengths

The time length of continuous data records limits the frequency resolution of all analysis methods, because of the fundamental time/frequency uncertainty. Data record lengths must be long enough to ensure adequate frequency resolution for the chosen analysis methods. The bias errors produced in different spectrum analysis estimates by inadequate frequency resolution

are discussed in Sections 3.2.3.4 and 3.3.3.7. Time domain methods such as correlation also require that record lengths must be much longer than the delay times of interest. Correlation bias errors are discussed in section 3.3.2.1.

4.5.5.5 Total Data Lengths

The total length of data needed for analysis is the product of the length of continuous records and the number of such records that must be averaged. If the total length of recorded data cannot be continuous because of test conditions (short events) or data acquisition limitations, then some means of identifying the continuous segments of data must be provided. If the total length of data is continuous, then overlap averaging may be used to increase the statistical accuracy of the results.

4.5.5.6 Digital Filter Types

After the data has been digitally sampled, further processing by the application of digital filters may be desirable. These operations may involve DC and trend removal; low-pass, high-pass, or bandpass filtering; sampling rate conversion; differentiation; or integration. This processing is performed either to increase the accuracy of further processing steps, or to "shape" the data for the purpose of display and later use.

4.5.5.7 Fast Fourier Transform (FFT) Lengths

The choice of FFT lengths, when applicable, is governed by a tradeoff of frequency resolution (long FFT's) and statistical accuracy (short FFT's and more averaging). Maximum transform lengths are limited by the continuous data length discussed above. When the data length is very long and the number of averages has been set for accuracy requirements, processing time or storage limitations become the limiting factor on transform

lengths. Other considerations, such as zero padding, are required for certain analysis methods.

4.5.5.8 Window Types

Windowing of data records before transform processing is necessary when the actual data extends beyond the record (FFT) length. The finite length of a data record causes "leakage" of spectral energy from frequency ranges of high energy to ranges with low energy. Frequency response estimates are particularly vulnerable to leakage effects around system resonances, where the input power spectral densities may be quite small and thus subject to leakage noise from nearby spectral peaks. This effect can be controlled (at the expense of some frequency resolution) by the application of smoothly tapering windows to the data record (see section 3.2.3.1). The choice of window types depends primarily on the signal being analyzed - Hanning or Hamming windows are usually used for random data, while windows with narrower time widths and correspondingly wider frequency bandwidths produce better results for periodic signals with well-separated peaks. Transient data captured completely within a record does not require windowing in theory, but in practice exponential or trapezoidal windows are occasionaly useful to reduce noise at the tail end of the transient. When exponential windows are used to force a transient decay to zero, the damping estimates will be biased high. This is a sign of insufficient record length (or, equivalently, insufficient frequency resolution) and must be corrected by using lower sampling rates or longer FFT lengths.

4.5.5.9 Number of Averages

The amount of averaging which is needed depends on the signal type and on the accuracy requirement for the analysis method. Random data requires averaging in order to reduce statistical errors (see section 3.2.3.4), while periodic or

transient data is averaged to reduce noise. If appropriate triggering can be arranged, periodic or transient data may be averaged in the time domain, but there is always some loss of high frequencies in time averaging. Overlap averaging, where portions of the previous data record are used in the current average, can increase the statistical accuracy if the total data length is fixed, but the averages are not independent. This means that the effective number of independent averages (used in the error formulas) is not as great as the number of overlapped averages.

4.5.6 Test Report Requirements

The requirements for documenting the conduct of the test should be specified in this section of the test plan. For simple tests, the test plan itself and the plotted test results may be sufficient. However, careful consideration of the value of test data (and the cost to repeat the test) leads to more effort in creating detailed records of the test for the future use of the results. Some items which might be included in this section are:

- Variations from original test plan specifications
- Types of intermediate results to include
- Presentation format for test results
- Long-term data storage

4.6 DOCUMENTATION OF RESULTS

An outline for documenting the results of the program is shown in Table 4.6. The use of material that has previously been prepared during the course of the program is emphasized to diminish the need for an extensive report preparation effort.

The problem description is taken from the material prepared according to the guidelines in Section 4.2. The statement of the

primary program objective, Section 4.3, is also readily adapted for the final documentation. The discussion of tests and data requirements in Section 4.4 can be condensed to describe the tests that were actually performed. The analytical models are then presented, accounting for modifications developed during the course of the program. Details of the development of analytical models and the detailed test plans and reports are presented in the appendices. Where material of unusual interest was generated as part of the program it would be included within the text of the final document.

The remaining sections involve a presentation and evaluation of results and the conclusions and recommendations. These discussions have as a primary focus the program objective of Section 4.3 and should be concisely formulated and responsive to the original statement of the problem.

Table 4.1 Primary Task Areas of Test Program Design

I. Problem Description: Presentation of the Problem to the Test Program Group.

1. Identification of the Problem
2. History of the Problem
3. Extent of the Problem
4. Impact of the Problem
5. Suggested scope of the test program

II. Program Objective: Formulation of the Goals and Approaches of the Test Program to Achieve an Efficient Solution of the Problem.

1. Problem review and statement
2. Primary program objective
3. Secondary objectives

III. Data Requirements: Development of an Overall Plan for Achieving the Program Objective

1. Test identification
2. Measured data
3. Analytical models

Table 4.1 Primary Task Areas of Test Program Design (continued)

IV. Development of Test Plans: Preparation of Detailed Plans Describing Each Required Test.

1. Test objective
2. Test configuration
3. Measurement types and sensor locations
4. Instrumentation requirements
5. Data Analysis requirements
6. Test report requirements

V. Documentation of Results: Preparation of the Test Program Description and Conclusions.

1. Problem description repeated from I
2. Program objectives repeated from II
3. Description of tests, measurements and analytical models
4. Presentation and evaluation of results
5. Conclusions and recommendations

Table 4.2 Problem Description

I. Identification of the Problem

1. Statement of the problem, its general nature and character.
2. Definitions of any special terminology employed.
3. Brief description of the system having the problem.
4. Evidence indicating the existence of present or potential problems.

II. History of the Problem

1. Chronology of events pertaining to the problem.
2. Remedial Measures taken to alleviate or bypass the problem.
3. Improvements resulting from remedial measures.
4. New problems introduced by remedial measures.

III. Extent of the Problem

1. Additional description of the system and the problem.
2. Hypotheses of the causes of the problem.
3. Possible solution methods.
4. Constraints on problem solutions.

Table 4.2 Suggested Outline for Problem Description (continued)

IV. Impact of the Problem

1. Lost capabilities and functions of the system.
2. Diminished performance and efficiency of the system.
3. Cost of down time for maintenance or repair.
5. Extended demands or safety problems for personnel.

V. Suggested Scope of the Test Program

1. Types of measurements: strain, acceleration, pressure, etc.
2. Locations where sensing devices should be placed.
3. Test conditions when measurements should be recorded.
4. Types of data analysis desired.
5. Analytical or statistical models to be used.

Table 4.3 Program Objective

I. Problem Review and Statement

1. Review problem description.
2. Resolve unclear areas.
3. Gather any additional information if necessary.
4. Identify any imposed solution constraints.
5. Summarize the key points into a one or two paragraph problem description.

II. Primary Program Objective

1. Hypothesize and discuss possible causes of the problem.
2. Evaluate potential solution approaches and criteria.
3. Describe the solution goal and the criteria for its evaluation in a statement of the primary program objective.

III. Secondary Objectives

1. Describe objectives of the test program which are desirable but of lower importance than the primary program objective.
2. Discuss the relative importance of these secondary objectives.

Table 4.4. Data Requirements

I. Test Identification

1. Tests for quantifying noise/vibration environment
2. Tests for diagnosing sources and transmission paths
3. Tests to check or verify potential problem solutions

II. Measured Data

1. Identify data needed by each test
2. Identify data to be measured by each test
3. Identify methods to be used for applying measured data

Table 4.4 Data Requirements (continued)

III. Analytical Models

1. Models to be developed
   a) system characteristics
   b) sources
   c) transmission paths
   d) response prediction

2. Model parameters to be obtained from test data

3. Verification of the Model

4. Demonstration of possible solution designs

Table 4.5 Test Plan Development

I. Test Objective:

1. Description of measurement to be performed
2. Data to be obtained from measurement
3. Analysis methods to be applied to data
4. Relation of test results to the program objective

II. Test Configuration

1. Test object and physical configuration
2. Variations in test object configuration to be evaluated
3. Range of operating conditions
4. Test site configuration and operation

III. Measurement Types and Sensor Locations

1. Quantities to be measured
2. Measurement transducer locations
3. Source simulation (excitation) device locations

IV. Instrumentation Requirements

1. Measurement transducers
2. Signal conditioning
3. Recording and playback system

V. Data Analysis Requirements

1. Data analysis instrumentation
2. Data analysis parameters
3. Types of analysis
4. Analyzer hardware and/or software packages
5. Data presentation parameters

VI. Documentation Requirements

1. Variations from original test plan specifications
2. Types of intermediate results to include
3. Presentation format for test results
4. Long-term data storage

Table 4.6 Documentation of Results

I. Problem description

II. Program objectives

III. Description of tests, measurements and analytical models

IV. Presentation and evaluation of results

V. Conclusions and recommendations

Appendices:

Analytical Models
derivations
refinements

Measurement Results
test plans
test reports

CHAPTER 4. REFERENCES

Pearson, J. "F-4G UHF Blade Antenna Tests", Preliminary Test Plan, Air Force Wright Aeronautical Laboratories, Wright-Patterson, Ohio (1978)

Bolds, P. "Vibration and Aeroelastic Facilities", Air Force Wright Aeronautical Laboratories, TR-82-3054

CHAPTER 5. ESTIMATION OF SYSTEM CHARACTERISTICS

By

D. H. KEEFE
R. G. DEJONG

CHAPTER 5. ESTIMATION OF SYSTEM CHARACTERISTICS

TABLE OF CONTENTS

		Page
	LIST OF FIGURES	5-iii
	LIST OF SYMBOLS	5-iv
5.1	THE FREQUENCY RESPONSE FUNCTION	5-1
	5.1.1 Basic Definitions	5-1
	5.1.2 Relationship to the Impulsive Response Function	5-7
	5.1.3 Effect of the Transducers on the Frequency Response Function	5-8
	5.1.4 Effect of Noise on the Frequency Response Function	5-13
5.2	DIRECT ESTIMATION OF FREQUENCY RESPONSE FUNCTIONS	5-20
	5.2.1 Single Excitation Point	5-20
	5.2.1.1 Single Response Point	5-20
	5.2.1.2 Multiple Response Points	5-27
	5.2.2 Multiple, Incoherent Excitations	5-30
	5.2.3 Multiple, Coherent Excitations	5-36
5.3	INDIRECT ESTIMATE OF FREQUENCY RESPONSE FUNCTION	5-39
	5.3.1 Decomposition of Multiple, Partially Coherent Excitations	5-39
	5.3.1.1 Errors	5-56
	5.3.1.2 Example	5-57

5.4 ESTIMATION OF DAMPING..... 5-60

5.4.1 Measures of Damping..... 5-60

5.4.1.1 Introduction..... 5-60
5.4.1.2 Experimental Methods..... 5-61
5.4.1.3 Example..... 5-65

5.4.2 Energy Decay Method..... 5-65

5.4.3 Autocorrelation and Random Decrement Methods.. 5-68

5.4.4 Spectral Bandwidth Method..... 5-70

5.4.5 Integrated Impulse Method..... 5-72

REFERENCES..... 5-74

LIST OF FIGURES

Figure		Page
5-1.	Frequency Response Function.....	5-2
5-2.	Impulse Decomposition of System Excitation and Response.....	5-9
5-3.	Single Input, Single Output System with Transducers.....	5-10
5-4.	Single Input, Single Output System with Noisy Transducers.....	5-12
5-5.	Experimental Setup for Drivepoint Measurement.....	5-26
5-6.	Measured Drive Point Accelerance of Flat Plate....	5-28
5-7.	Single Input, Multiple Output System.....	5-29
5-8.	Two Input, Single Output System.....	5-32
5-9.	Two Coherent Input System.....	5-38
5-10.	General Two Input System.....	5-42
5-11.	Equivalent Two Input System.....	5-43
5-12.	General Multiple Input/Single Output Model.....	5-48
5-13.	Multiple Input/Single Output System for Equivalent Incoherent Inputs.....	5-49
5-14.	Reduction of System Input to Equivalent Input.....	5-50
5-15.	Computation of Conditional Spectral Density Functions.....	5-53
5-16.	Measured Multiple Coherence Function.....	5-58
5-17.	Measured Coherent Output Power.....	5-59
5-18.	Energy Level Decay and Reverberation Time.....	5-63
5-19.	Measured Transfer Function of Beam.....	5-66

LIST OF SYMBOLS

A	Acceleration
B(G)	Signal processing bias error in estimate of spectral density function
C(f)	Coincident component of transfer
D(t)	Randomec time signature
E(t)	Energy in mechanical system
F	Force
f	Frequency, cycles/sec
G(f)	Spectral density function
H(f)	Transfer function
h	Impulse response function
j	$\sqrt{-1}$ complex number
K	Number of frequency averages
L(f)	Transfer function between coherent portions of signals
M	Number of curseuble averages
m(t), M(f)	Noise time history, spectrum at the input
n(t), N(f)	Noise time history, spectrum at the output
P	Acoustic pressure
Q(f)	Quadrature component of transfer function
q	Input signal index in multiple input system
R(τ)	Correlation function
S	Surface area
T	Time period
T(f)	Transducer frequency response function
t	Time
U	Volume velocity
U(f)	Generalized input variable in the absence of noise
V	Velocity
V(f)	Generalized output variable in the absence noise

$V(G(f))$	Variance in estimate of spectral density function
X	Displacement variable
$x(t)$, $X(f)$	Generalized input variable and its transform
Y	Mobility
$y(t)$, $Y(f)$	Generalized output variable and its transform
Z	Impedance
z	Generalized signal variable

α	Phase factors, radians
β	Phase factors, radians
$\gamma^2(f)$	Coherence function
ε	Normalized bias error in estmate of spectral density function
ϕ	Phase, radian
η	Damping loss factor
Σ	Summation symbol
τ	Time delay, sec
ω	Angular frequency, radians/sec

CHAPTER 5. ESTIMATION OF SYSTEM CHARACTERISTICS

In this chapter we present various methods of using measured data to estimate the characteristics of dynamic systems. The most commonly used descriptor of system characteristics is the frequency response function. The majority of this chapter deals with the methods of estimating this function for different types of system excitations.

An alternative descriptor is the impulse response function which is the time domain equivalent to the frequency response function. Section 5.1 presents the relationship between these two functions.

Another descriptor of system characteristics closely related to the frequency response function is the level of damping or energy dissipation. Section 5.4 presents various methods of measuring damping of a dynamic system.

5.1 THE FREQUENCY RESPONSE FUNCTION

5.1.1 Basic Definitions

In general, the frequency response function of a dynamic system can be defined as the level of response of the system due to a unit level of excitation at any given frequency. The parameters which measure the levels of excitation and response are dependent on the system being considered. Some examples of this will be given later in this section.

In order to better define the frequency response function it is convenient to consider the excitation as an input to the dynamic system, and the response as an output, as shown in Figure 5-1. If the input X(f) is a sinusoidal signal with frequency f and the output at the same frequency is Y(f), then the system transfer function H(f) is given by

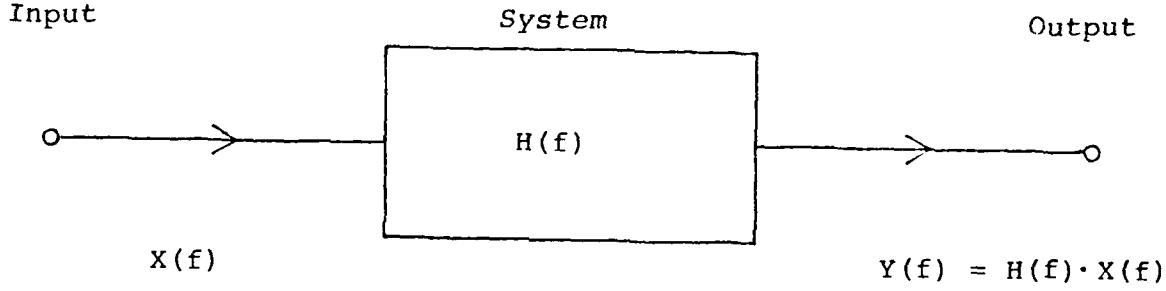


Figure 5-1. Frequency Response Function

$$H(f) = Y(f)/X(f) \tag{5-1}$$

This transfer function is, in general, a function of frequency and is equivalent to the frequency response function of the system being modeled.

To be complete, the frequency response function must represent both the relative magnitude and the relative phase of the system response. Therefore H(f) is defined as a complex number of the form

$$H(f) = |H(f)|\ e^{j\phi(f)} \tag{5-2}$$

where $|H(f)|$ is the magnitude and $\phi(f)$ is the phase of the frequency response function. If the time history of the input signal is given by $x(t) = |X(f)|\ \sin[2\pi ft + \alpha(f)]$ and that of the output signal is given by $y(t) = |Y(f)|\ \sin[2\pi ft + \beta(f)]$ then

$$|H(f)| = |Y(f)|/|X(f)| \tag{5-3}$$

and

$$\phi(f) = \beta(f) - \alpha(f) \tag{5-4}$$

An alternative representation of the frequency response function uses the real and imaginary parts of the complex number as follows:

$$H(f) = C(f) + j\ Q(f) \tag{5-5}$$

where

$$C(f) = |H(f)|\ \cos\ \phi(f) \tag{5-6a}$$

$$Q(f) = |H(f)|\ \sin\ \phi(f) \tag{5-6b}$$

The real and imaginary parts of H(f) are C(f) and Q(f), respectively. Physically, the real part of the frequency response

function is a measure of that part of the output signal which is in phase with the input signal (the coincident component). The imaginary part is a measure of that part of the output signal which is 90° out of phase with the input signal (the quadrature component).

Three assumptions are generally made about a system input/output model in order to simplify the analysis of the frequency response function. The first assumption is that the system is linear. In a linear system a proportional change in the level of the input produces the same proportional change in the level of the output. This means that the value of the frequency response function is independent of the absolute levels of the input and output.

Another characteristic of a linear system is that an input of a single frequency will produce an output at only the same frequency.

The second assumption is that the system is passive. A passive system has no internal sources of power.

The third assumption is that the system is time-invariant, that is, its characteristics do not change with time. In a time-invariant system an estimate of the frequency response function over a finite time period can be considered a stationary random variable (see Section 3.3.3.6).

Although the definition of the frequency response function in Equation (5-1) was stated for signals of a single frequency, it can also be applied to signals of arbitrary time history. This can be done by defining X(f) and Y(f) as the components of x(t) and y(t), respectively, with frequency f. The functions X(f) and Y(f) are the Fourier Transforms (FT) of x(t) and y(t), respectively, defined by

$$X(f) = \int_{-\infty}^{\infty} x(t)\ e^{-j2\pi ft} dt$$

$$Y(f) = \int_{-\infty}^{\infty} y(t)\ e^{-j2\pi ft} dt \qquad (5\text{-}7)$$

The reciprocal relationships are given by the inverse FT, defined by

$$x(t) = \int_{-\infty}^{\infty} X(f)\ e^{j2\pi ft} df$$

$$y(t) = \int_{-\infty}^{\infty} Y(f)\ e^{j2\pi ft} df \qquad (5\text{-}8)$$

These relationships cannot be implemented directly in practice since they require information about the signals for all time or at all frequencies. Methods of evaluating the FT with a finite time sample of a signal are given in Section 3.2.3.1.

Equation (5-7) shows that X(f) and Y(f) are complex quantities due to the presence of the complex exponential in the integral. They are of the form

$$X(f) = |X(f)|\ e^{j\alpha(f)}$$

$$Y(f) = |Y(f)|\ e^{j\beta(f)} \qquad (5\text{-}9)$$

The phase of each is relative to an arbitrary definition of when t = 0 in Equation (5-7). The frequency response function H(f) is then given by Equations (5-1) through (5-6).

The physical quantities used to represent the frequency response function of systems in acoustics and vibration depend

on the characteristics of the system. For acoustic systems, the response variable is usually a pressure P and the excitation is a volume velocity U. (U is defined as the average velocity V of a surface multiplied by the surface area S, or U = VS.) The corresponding frequency response function is called the acoustic impedance Z,

$$Z = P/U \tag{5-10}$$

The inverse of the acoustic impedance is called the admittance.

For vibratory systems the excitation variable is usually a force F and the response variable is one of the following three motion variables: displacement X, velocity V, or acceleration A. The following table summarizes the terminology used for the various frequency response functions related to these variables as presented in Reference [5-1].

VARIABLE RATIO	NAME	PREFERRED SYMBOL
A/F	Accelerance	---
V/F	Mobility (Mechanical Admittance)	Y
X/F	Compliance	---
F/A	Dynamic Mass	---
F/V	Mechanical Impedance	Z
F/X	Dynamic Stiffness	---

The frequency response function of a physical system is also dependent on the location of the measured quantities. When the excitation and response variables are measured at the same location, the ratio is called an input (or drive point) frequency

response function. When the response is measured at a location remote to the excitation, the ratio is called a transfer frequency response function. Other transfer ratios can be defined in terms of either an excitation or response quantity measured at two locations, such as: force, F_2/F_1; pressure, P_2/P_1; acceleration, A_2/A_1; etc.

For the remainder of this chapter the symbols x and y will continue to be used as two arbitrary, measured quantities with no relationship to particular physical quantities.

5.1.2 Relationship to the Impulsive Response Function

An alternative descriptor of the characteristics of a dynamic system is the impulse response function, h(t). Physically, this function is the time history of the response of a system excited by a unit impulse. Mathematically, it is related to the frequency response function by the Fourier Transform as follows:

$$h(t) = \int_{-\infty}^{\infty} H(f)\ e^{j2\pi ft} df$$

$$H(f) = \int_{-\infty}^{\infty} h(t)\ e^{-j2\pi ft} dt \qquad (5\text{-}11)$$

The relationship between the impulse response function and the input and output time signals is given by

$$y(t) = \int_{-\infty}^{\infty} h(\tau)\ x(t-\tau)\ d\tau \qquad (5\text{-}12)$$

This relationship is more complicated than the corresponding one for the frequency function given in Equation (5-1).

The integral in Equation (5-12) is called the convolution integral. The value of the output y(t) is a weighted sum of the values of the input x(t) at all times, and the weighting function is h(t). For physically realizable systems, the value of h(t) equals zero for $t < 0$. This means the system is causal (it does not start responding until it is excited). Then y(t) is the weighted sum of the current values of x(t).

Figure 5-2 illustrates in an approximate manner how the convolution integral works. The input x(t) can be divided up into a sequence of narrow pulses. Each pulse causes the system to respond with the time history similar to the impulse response function scaled by the amplitude of the pulse and beginning at the occurrence of the pulse. The output y(t) is then the sum of all these individual impulse responses. The convolution integral is the limit of this operation as the duration of each pulse approaches zero. For the special case of an input consisting of an ideal unit impulse, the output is precisely equal to the impulse response function.

5.1.3 Effect of the Transducers on the Frequency Response Function

Thus far in this chapter it has been assumed that the input and output variables have been measured with perfect accuracy by the transducers. That is, the output voltage signal from the transducer is an exact copy of the original dynamical variable sensed by the transducer. This is never the case in practice. Not only are errors introduced due to the presence of the transducer, but there are errors as well due to noise which may be present anywhere in the system and signal processing errors in the A/D converter and in the spectral analysis. A general discussion of noise is given in Section 5.1.4 and specific

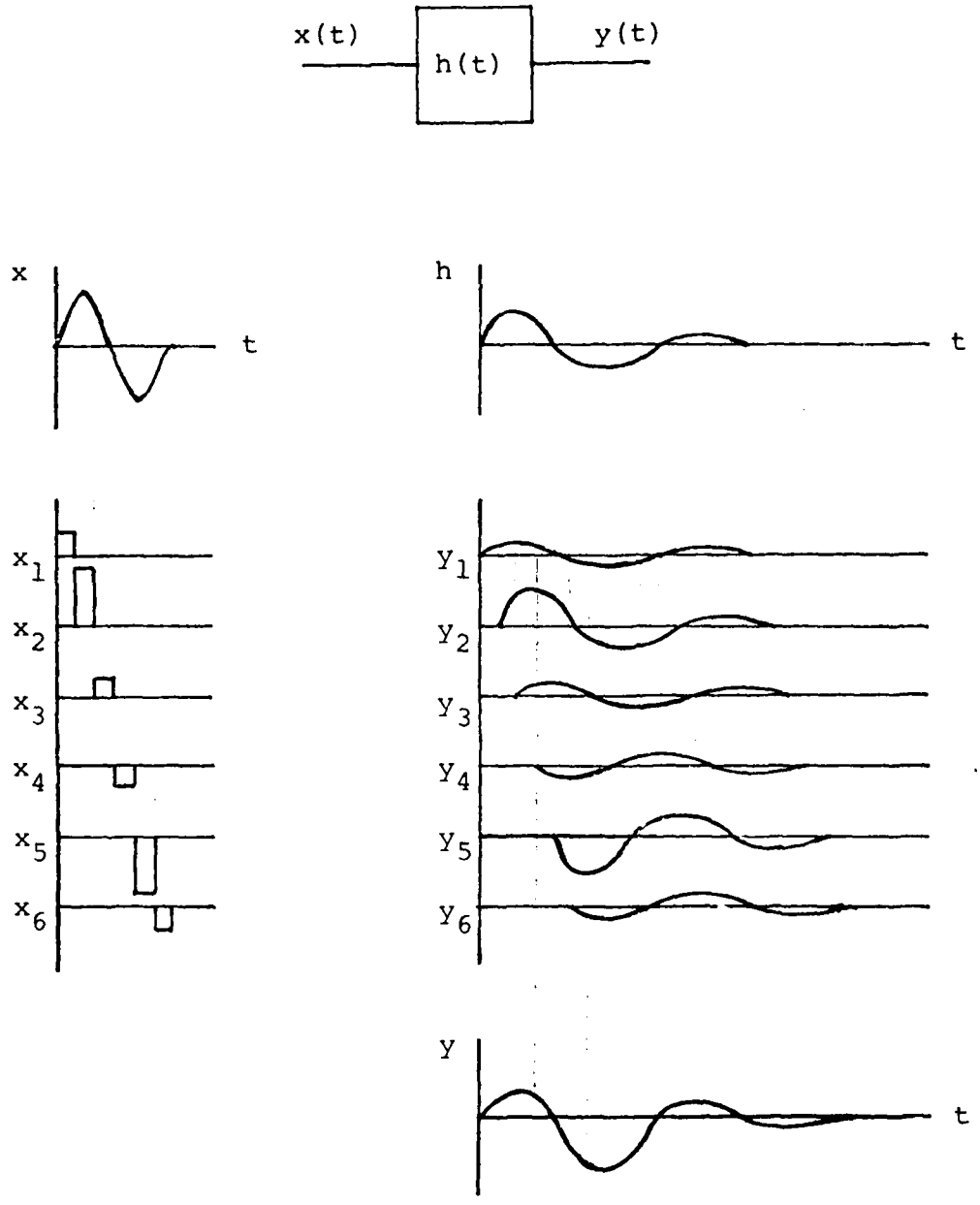


Figure 5-2. Impulse Decomposition of System Excitation and Response

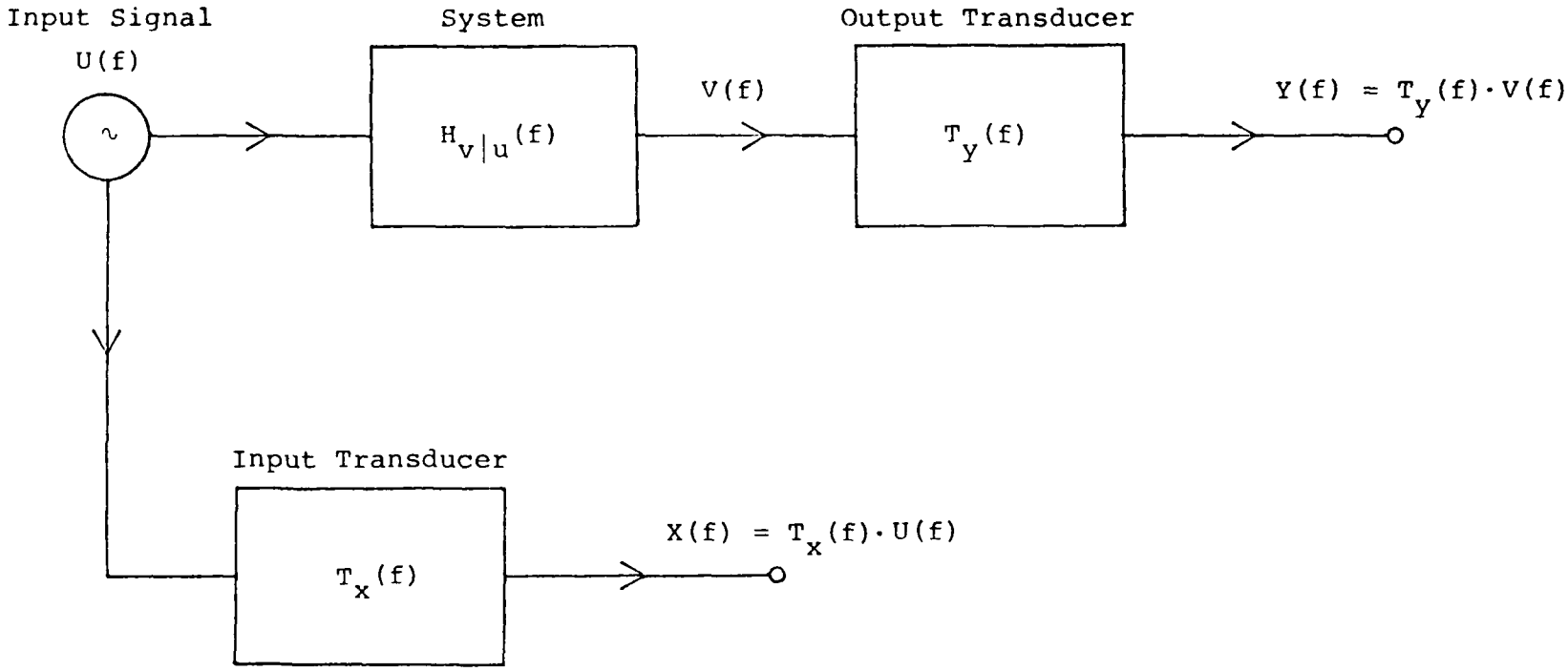


Figure 5-3. Single Input, Single Output System with Transducers

applications are presented in Sections 5.2 and 5.3. For present purposes, it is assumed that there is no noise present in the transducer, even though any transducer has a "noise floor" which inhibits the detection of sufficiently small signals. Section 5.1.4 shows how the presence of noise in the transducer may be represented as an additional input source of noise to the system. Errors in the measurement of the frequency response function due to signal processing are discussed in Chapter 3.

Figure 5-3 illustrates the same single input, single output as in Figure 5-1, except that the transducers are no longer assumed to be ideal (i.e., flat frequency response and sensitivity normalized to unity). There is a frequency response function T_x associated with the input transducer and T_y associated with the output transducer. The simplest type of transducer frequency response function is a numerical constant at all frequencies within the bandwidth of interest. This constant is the sensitivity of the transducer and typically gives the ratio of output voltage for a given input excitation (due to acoustical pressure, light intensity, acceleration of a structure, etc.). If the frequency response of the transducer is not constant, then the frequency response function T is a function of frequency.

Examination of Figure 5-3 shows that the true frequency response function is

$$H_{v|u}(f) = V(f)/U(f) \qquad (5\text{-}13)$$

whereas the measured frequency response function $H_{y|x}$, which is the ratio of the measured output signal to the measured input signal, is given by,

$$H_{y|x} = Y(f)/X(f) \qquad (5\text{-}14)$$

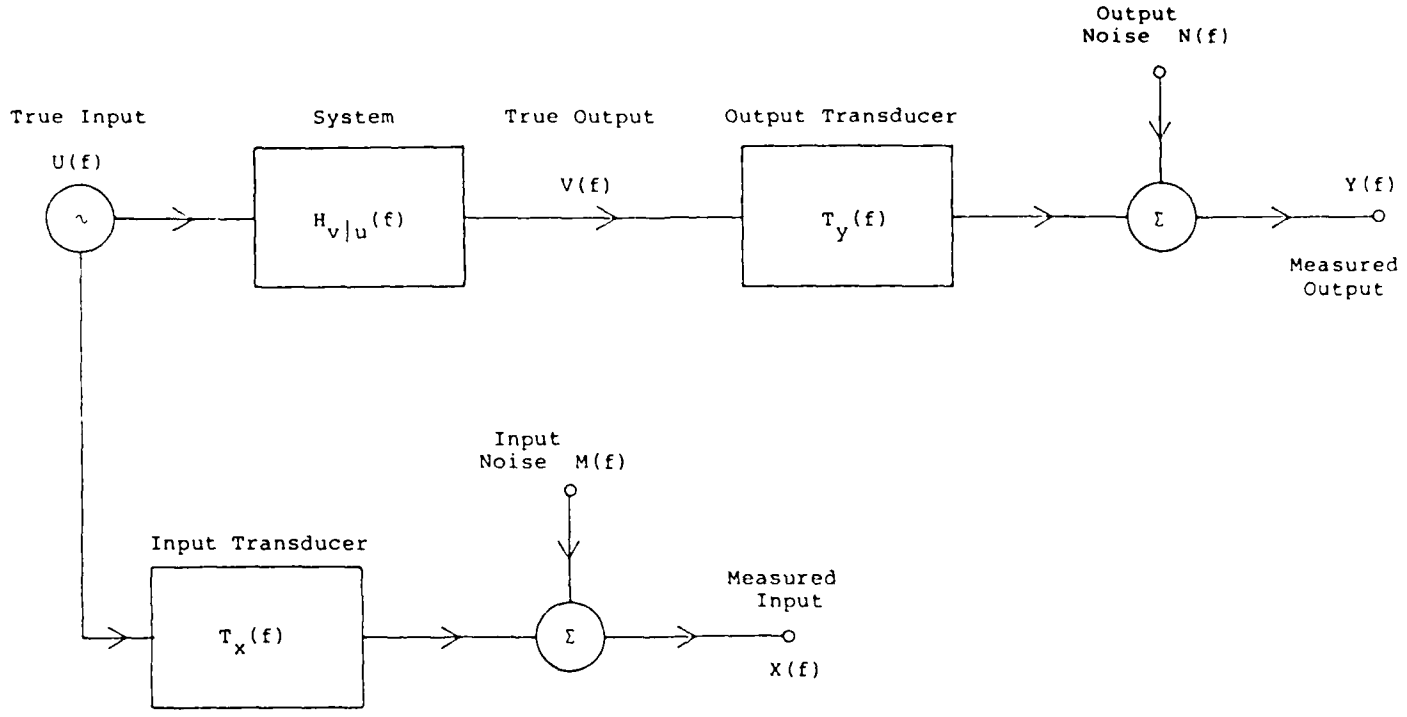


Figure 5-4. Single Input, Single Output System with Noisy Transducers

The frequency response function $H_{v|u}$ of the system is related to the measured function $H_{y|x}$ by

$$H_{v|u}(f) = H_{y|x}(f) \ [T_x(f)/T_y(f)] \qquad (5\text{-}15)$$

Therefore, one has to multiply the observed frequency response function of the system by the frequency response function associated with the input transducer, and divide by the output transducer frequency response function in order to determine the true frequency response function of the system. If the signal processing equipment in the laboratory includes a computer, then these frequency response functions associated with the tranducer calibration can be stored in memory and the appropriate operations can be carried out on the measured data. Additional information on the types of transducers and how to choose the right one for the desired type of measurement is found in Chapter 2.

5.1.4 Effect of Noise on the Frequency Response Function

One of the most common problems with which the experimenter or technician has to contend is the presence of noise in the measuring apparatus or signal processing equipment. This section is concerned with the presence of noise in the measuring apparatus, that is, the input and output transducers, and its effect upon the interpretation of frequency response function measurements. Errors in the digital signal processing are discussed in Chapter 3, Digital Data Analysis Technology, and in Section 5.2.

Each of the transducers is assumed to have a frequency response function associated with it (see Section 5.1.3) and there is assumed to be noise present in each of the transducers as well. Figure 5-4 illustrates the system. These noise signals

are summed with the input and output signals exciting the system, and the signals X(f) from the input transducer and Y(f) from the output transducer are:

$$X(f) = T_x(f)\ U(f) + M(f)$$

$$Y(f) = T_y(f)\ V(f) + N(f) \qquad (5\text{-}16)$$

where the input and output noise spectra are M(f) and N(f), respectively, and the input and output ideal spectra are U(f) and V(f), respectively. If the noise source is located within the system itself, then the output noise is summed with the system noise before being sensed by the transducer. For clarity, the noise has been assumed to be within the transducers as expressed by Figure 5-4 and Equation (5-16).

An important practical simplification is that the noise sources are uncorrelated with the ideal input and output signals, and with each other. For example, there is normally no causal connection between the true input signal u(t) to the system and the input noise signal m(t). This condition that u and m are statistically independent signals is expressed mathematically in the statement that the cross-correlation function of u and m is zero (see Section 3.3.2 for a discussion of cross-correlation functions between two time series).

For present purposes, it is convenient to define the auto-correlation and cross-correlation functions associated with continuous variables. The corresponding functions for time series, that is, for discrete functions are discussed in Chapter 3. The auto-correlation function R_{xx} associated with the signal x(t) is defined to be

$$R_{xx}(\tau) = \lim_{T\to\infty} (1/T) \int_0^T x(t)\ x(t+\tau)\ dt \qquad (5\text{-}17a)$$

The auto-correlation function is a measure of how the values of $x(t)$ are related to values of the signal at a later time $(t + \tau)$. The cross-correlation function R_{xy} associated with the signals $x(t)$ and $y(t)$ is defined by

$$R_{xy}(\tau) = \lim_{T\to\infty} (1/T) \int_0^T x(t)\ y(t + \tau)\ dt \qquad (5\text{-}17b)$$

It follows from these definitions that the auto-correlation function is an even function of time delay, $R_{xx}(\tau) = R_{xx}(-\tau)$. The spectral density function $G_{xx}(f)$ of the signal $x(t)$ defined only for positive f is related to the auto-correlation function by

$$G_{xx}(f) = 2 \int_{-\infty}^{\infty} R_{xx}(\tau) \cos 2\pi f\tau\ d\tau \quad (f > 0) \qquad (5\text{-}18a)$$

The cross-spectral density function $G_{xy}(f)$ defined to be non-zero only for non-negative frequencies f is related by the Fourier Transform to the cross-correlation function by

$$G_{xy}(f) = 2 \int_{-\infty}^{\infty} R_{xy}(\tau)\ e^{-j2\pi f\tau}\ d\tau \quad (f > 0) \qquad (5\text{-}18b)$$

Now consider an arbitrary signal z which is the sum of two signals x and y as follows:

$$z(t) = x(t) + y(t) \qquad (5\text{-}19)$$

It follows from the linearity of the definitions of the auto-correlation and the spectral density functions that

$$R_{zz}(\tau) = R_{xx}(\tau) + R_{yy}(\tau) + R_{xy}(\tau) + R_{yx}(\tau)$$

and

$$G_{zz}(f) = G_{xx}(f) + G_{yy}(f) + G_{xy}(f) + G_{yx}(f) \qquad (5\text{-}20)$$

We return to the system in Figure 5-4. The statement that the noise signals m(t) and n(t) are uncorrelated with either of the ideal input or output signal is that

$$R_{mu}(\tau) = R_{nv}(\tau) = 0 \qquad (5\text{-}21a)$$

and the fact that the cross-spectral density is the Fourier Transform of the cross-correlation function implies

$$G_{mu}(f) = G_{nv}(f) = 0 \qquad (5\text{-}21b)$$

Consideration of Figure 5-4 shows the frequency response function H to be the ratio of the ideal output V to to the ideal input U as follows:

$$V(f) = H_{v|u}(f)\ U(f) \qquad (5\text{-}22)$$

The measured input signal X from the noisy input transducer as well as the measured output signal Y are related to the ideal signals and noise signals as follows:

$$X(f) = T_x(f)\ U(f) + M(f) \qquad (5\text{-}23a)$$

$$Y(f) = T_y(f)\ V(f) + N(f) \qquad (5\text{-}23b)$$

The corresponding autospectral and cross-spectral densities under the conditions that both of the noise signals are uncorrelated with both the true input and output signals are

$$G_{vv}(f) = |H_{v|u}(f)|^2 \, G_{uu}(f)$$

$$G_{uv}(f) = H_{v|u}(f) \, G_{uu}(f)$$

$$G_{xx}(f) = |T_x(f)|^2 \, G_{uu}(f) + G_{mm}(f)$$

$$G_{xy}(f) = T_x^*(f) \, T_y(f) \, G_{uv}(f) + G_{mn}(f)$$

$$G_{yy}(f) = |T_y(f)|^2 \, G_{vv}(f) + G_{nn}(f). \qquad (5\text{-}24)$$

Note that the input and output noise signals may be correlated; i.e., G_{mn} is not assumed to be zero. The measured frequency response function of the system $H_{y|x}(f)$ is defined by (see Section 3.3.3.6):

$$H_{y|x}(f) = \frac{G_{xy}(f)}{G_{xx}(f)} \qquad (5\text{-}25a)$$

whereas the true frequency response function of the system is expressed in terms of the spectral density functions of the ideal input and output as follows:

$$H_{v|u}(f) = \frac{G_{uv}(f)}{G_{uu}(f)} \, . \qquad (5\text{-}25b)$$

Use of Equations (5-24) and (5-25) implies the following relation between the true and the measured system frequency response function:

$$H_{y|x}(f) = H_{v|u}(f) \, \frac{T_y(f)}{T_x(f)} \left\{ \frac{1 + \frac{G_{mn}(f)}{G_{uv}(f)} \, [T_x^*(f) \, T_y(f)]^{-1}}{1 + \frac{G_{mm}(f)}{G_{uu}(f)} \, [T_x(f)]^{-2}} \right\} \qquad (5\text{-}26)$$

Equation (5-26) is examined under several conditions. When no noise is present at either the input or output, then this equation reduces to:

$$H_{y|x}(f) = [T_y(f)/T_x(f)]\ H_{v|u}(f) \qquad (5\text{-}27)$$

The above relation is exactly the same as Equation (5-15). Now suppose there is noise at the input only; that is, n(t) in Figure 5-4 is equal to zero but m(t) is non-zero. In this case the frequency response function $H_{y|x}$ in Equation (5-26) simplifies to:

$$H_{y|x}(f) = H_{v|u}(f)\ \left\{\frac{T_y(f)/T_x(f)}{1 + \dfrac{G_{mm}(f)}{G_{uu}(f)}\ |T_x(f)|^{-2}}\right\} \qquad (5\text{-}28)$$

This equation is an important relation and shows that the presence of noise at the input always results in a measured frequency response function which is smaller in magnitude at all frequencies than if there were no noise present in the system. Equation (5-28) is valid under the more general condition that output noise is present, but that the output noise is uncorrelated with the input noise (so that $G_{mn} = 0$).

The error enters because of a biased estimate of the input autospectral density function; that is, G_{xx} is not equal to G_{uu}. The measured cross-spectral density G_{xy} is equal to the true cross-spectral density G_{uv}, because the input noise is uncorrelated with the output of the system. The magnitude of the bias in the estimate of the frequency response function is seen to be proportional to the the input noise-to-signal level $\alpha(f)$, defined as

$$\alpha(f) = G_{mm}(f)/G_{uu}(f) \qquad (5\text{-}29)$$

The coherence function $\gamma^2_{xy}(f)$ discussed in Section 3.3.3.5 is a measure of the linear relationship of two arbitrary signals x and y, and is defined to be

$$\gamma^2_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)\ G_{yy}(f)} \qquad (5\text{-}30)$$

The coherence function is a real, non-negative number whose value lies between 0 and 1. The coherence function between two uncorrelated signals is always zero, so that the coherence function between any pair of variables listed in Equation (5-21) is zero.

For example, the coherence function between the two noise signals is

$$\gamma^2_{mn}(f) = \frac{|G_{mn}(f)|^2}{G_{mm}(f)\ G_{nn}(f)} \qquad (5\text{-}31)$$

The cross-spectral density function G_{mn} between the input and output noise signals and hence the coherence function $\gamma_{mn}{}^2$ is zero as long as the noise signals are statistically independent. This is the case when each noise signal is independently produced. On the other hand, the noise signals are highly correlated if each transducer is sensing electrical line noise at 60 Hz; if both noise signals are caused by a common source then the coherence function equals unity.

The coherence function $\gamma^2{}_{xy}$ between the measured input and output signals is a coherence function which is often encountered. The coherence function corresponding to the system in Figure 5-4 is

$$\gamma_{xy}{}^2(f) = \frac{\left[1 + \frac{G_{mn}(f)}{G_{uv}(f)}\ (T_x^*(f)\ T_y(f)^{-1}\right]^2}{\left[1 + \frac{\alpha}{|T_x|^2}\right]\ \left[1 + \frac{\beta}{|T_y|^2}\right]} \qquad (5\text{-}32)$$

where β is the output noise to signal ratio defined by

$$\beta(f) = \frac{G_{nn}(f)}{G_{vv}(f)} \qquad (5\text{-}33)$$

Many dual channel digital spectrum analyzers which are now available compute this coherence function. The presence of input or output noise reduces its value from unity. However, the extent of the reduction critically depends on the coherence between the two noise signals.

5.2 DIRECT ESTIMATION OF FREQUENCY RESPONSE FUNCTIONS

5.2.1 Single Excitation Point

5.2.1.1 Single Response Point

This section describes the procedures used to measure a single input, single output frequency response function, the associated errors incurred, and the slight generalization to include the case of a single input and multiple outputs.

The first stage in the estimation process is to define the system of interest, as well as the inputs and outputs to that system. Next, one chooses input and output transducers appropriate to the types of signals of interest (see Chapter 2). The selection of instrumentation is complete with the choice of (digital) signal processing equipment. The continuous signals from the transducers are sampled and converted into a dual time series (see Chapter 3). Let the input time signal be x(t) and the output time signal be y(t). Then the digital signal processing estimate $H_{y|x}(f_k)$ of the frequency response function at the k-th sample frequency f_k is computed by

$$H_{y|x}(f_k) = G_{xy}(f_k)/G_{xx}(f_k) \qquad (5\text{-}34)$$

where G_{xx} is the autospectral density function estimate and G_{xy} is the cross-spectral density function estimate. The relation of the spectral density functions to the system frequency response function is discussed in Section 5.1, and the digital signal processing estimation of the spectral density functions is discussed in Sections 3.2.3 and 3.3.3. The effect of the non-ideal behavior of the transducers is discussed in Section 5.1.3, and Equation (5-15) shows the appropriate transducer correction factor to apply to the estimated frequency response function in order to estimate the true frequency response function H(f) of the system.

A more troublesome problem is caused by the presence of noise at some point in the system. The most important noise sources are those located at (or within) the input and output transducers. The input and output transducer signals x(t) and y(t), respectively, are

$$x(t) = u(t) + m(t)$$

$$y(t) = v(t) + n(t) \qquad (5\text{-}35)$$

where the true input and output signals are u(t) and v(t), and the input and output noise signals are m(t) and n(t), respectively. The presence of either input or output noise signals produces a systematic error in the estimation of the frequency response function. This error is given in Equation (5-26) of Section 5.1.4. There is always noise present in any laboratory measurement. The important question is whether the noise affects the estimate in a significant way. This question is most easily answered by computation of the coherence function between the input and output signals. Most dual channel digital spectrum

analyzers supply the coherence function. The coherence function is defined in Section 3.3.3.5 and discussed in relation to frequency response function in Section 5.1.4. The coherence function is equal to unity for linear systems with no noise present.

The presence of noise reduces the coherence function. Therefore, during the measurement procedure one should examine the estimate of the coherence function at the frequencies f_k. A value of the coherence function between 0 - 0.5 denotes that the noise signal is comparable to or larger than the signals of interest. The estimates of the frequency response function at these frequencies will be poor. Values of the coherence function near unity implies that the estimates of the frequency response function are not significantly degraded by the presence of noise. If there are large noise signals somewhere in the system, the experimenter or technician must find their source and make suitable changes in the apparatus and signal conditioning equipment in order to make subsequently meaningful estimates of the frequency response function.

The accuracy of the frequency response function estimate is also degraded by statistical errors. These errors arise from the limited validity of assumptions which underlie the digital signal processing. These statistical errors are more fully discussed in Sections 3.2.3.4 and 3.3.3.7. The errors can either be random or bias errors. Random errors are due to the scatter in the various estimates of a single variable and these errors are reduced by ensemble averaging. Bias errors involve a systematic shift in the estimate of any variable. For example, ignoring the non-flat frequency response of any transducer introduces a bias error into the frequency response function estimate. Reference [5-2] has a full discussion of bias and random errors in spectral estimation.

A statistical bias error in the digital signal processing is introduced due to the finite filter bandwidth of the FFT (see Sections 3.2.3.4 and 3.3.3.7). The bias error B(G) in the

estimate of the autospectral or cross-spectral density function G is

$$B[G(f)] = \frac{f_b^2 \, G''(f)}{24} \qquad (5\text{-}36)$$

where G''(f) is the second derivative of the spectral density function with respect to frequency and f_b is the effective filter bandwidth of the FFT. Since G'' is proportional to the curvature of the spectral density function, then the bias error is largest in frequency regions where the slope of G is rapidly changing - near maxima and minima. The normalized bias error $\varepsilon_b[G(f)]$ of the auto or cross-spectral density function G is

$$\varepsilon_b[G(f)] = B[G(f)]/G(f) = -(1/3)\left(\frac{f}{\Delta f_{1/2}}\right)^2 \qquad (5\text{-}37)$$

where $\Delta f_{1/2}$ is the half-power bandwidth (see Section 5.4.4) for the case of an autospectral density function maxima and the half-peak amplitude bandwidth for the cross-spectral density function.

There are also random errors in the evaluation of the auto-spectral and cross-spectral density functions associated with the digital signal processing. The variance $V[G_{xx}(f)]$ in the estimate of the autospectral density is

$$V[G_{xx}(f)] = \frac{G_{xx}^2(f)}{MK} \qquad (5\text{-}38)$$

where the number of ensemble averages is M and the number of averages at adjacent frequencies is K (see Section 3.2.3.4). In this case, the normalized random error $\varepsilon_r[G_{xx}(f)]$ for the autospectral density G_{xx} is

$$\varepsilon_r[G_{xx}(f)] = 1/\sqrt{MK}. \qquad (5\text{-}39)$$

Thus, increasing the number of averages diminishes the random error to any desired level. If, in addition, third-octave averaging is used, then each estimate of the autospectral density is an average over the third-octave of the estimates for all the sample frequencies contained in that interval. The factor K in Equation (5-38) refers to the number of data points in the third-octave. Hence, averaging the estimates over a wide bandwidth tends to smooth out the random error in the resulting estimate at the expense of overlooking the detailed structure of the spectrum.

The variance $V[|G_{xy}(f)|]$ in the estimate of the magnitude of the cross-spectral density function assuming white noise excitation is

$$V[|G_{xy}(f)|] = \frac{|G_{xy}(f)|^2}{2MK}\left(1 + \frac{1}{\gamma^2_{xy}}\right) \qquad (5\text{-}40)$$

where γ^2_{xy} is the coherence function. At frequencies where the coherence function between the input and output is low, the random error of the cross-spectral density function increases. For values of the coherence function close to unity, the relative random error in the magnitude estimation of the cross-spectral density function can be approximated by,

$$\varepsilon_r[|G_{xy}(f)|] = \frac{1}{|\gamma_{xy}(f)|\sqrt{MK}} \qquad (5\text{-}41)$$

Increasing the number of averages reduces the random error.

Errors in estimates of the spectral density functions produce error in the estimate of the input-output frequency response function $H_{y|x}(f)$. This is because the frequency response function is computed via Equation (5-34) as the ratio of the

cross-spectral density to the input autospectral density function. The relative random error in the magnitude $|H_{y|x}|$ and the phase error in radians of the frequency response function phase ϕ_{xy} are approximately equal to

$$\varepsilon_r[|H_{y|x}(f)|] = \frac{\{1 - \gamma_{xy}^2(f)\}^2}{|\gamma_{xy}(f)|\sqrt{2MK}} \qquad (5\text{-}42a)$$

$$\Delta\phi_{xy} = \varepsilon_r[|H_{y|x}(f)|] \qquad (5\text{-}42b)$$

That is, the random errors associated with the magnitude and phase estimates of the frequency response function are equal and they tend to decrease with the number of averages as $1/\sqrt{M}$.

The bias in the estimate of the magnitude of the frequency response function may be computed in terms of the bias error in the spectral density functions and is

$$\varepsilon_b[|H_{y|x}(f)|] = \{\varepsilon_b^2[G_{xx}(f)] + \varepsilon_b^2[G_{xy}(f)]\}^{1/2} \qquad (5\text{-}42c)$$

EXAMPLE: We conclude this section with an illustrative example shown in Figure 5-5. The system under investigation is a flat aluminum plate of thickness h = 1.6 mm and area A = 1.5 m^2. Bending waves in the plate are excited by a shaker using band-limited white noise from 10 - 1000 Hz. The input variable is the force excitation at the driving point F, which is measured by a force gauge mounted between the shaker and the plate. The output variable is the acceleration response of the structure at the driving point A, measured by an accelerometer with a flat frequency response in the range of interest. The frequency

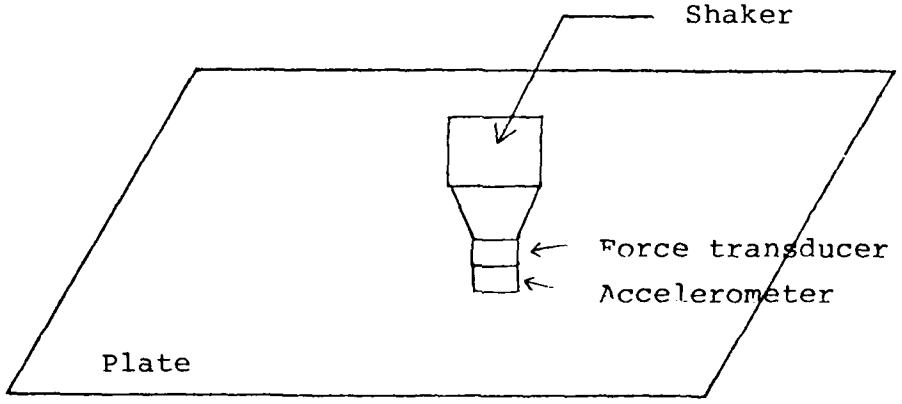


Figure 5-5. Experimental Setup for Drivepoint Measurement

response function over the entire frequency range is desired in 2.5 Hz bandwidths.

The shaker is driven with a white noise signal which is bandpass filtered from 10 - 1000 Hz. Both measured signals are also bandpass filtered from 10 - 1000 Hz. The signals are sampled digitally and processed using the methods described in Chapter 3. A sample rate of 2500 Hz is chosen for each data channel using the criterion of Equation (3-2).

In order to have a frequency resolution of 2.5 Hz it is necessary to obtain sample records of 0.4 sec according to Equation (3-7). This means that each record must have 1000 data samples. (For convenience of the digital hardware, this number would be adjusted to 1024 samples. Then either the frequency resolution would change to 2.44 Hz or the sample rate would be raised to 2560 Hz.)

The power spectral densities G_A and G_F and the cross-spectral density G_{AF} are computed according to Sections 3.2.3.1 and 3.3.3.1 (using a Hanning window function). The random error of the measurement is reduced by repeating and averaging this measurement over 100 sample records. Using Equation (3-80) the standard deviation is then 10% of the measured values.

The accelerance of the plate ($H_{A|F}$) is then computed by Equation (5-25a) and the results are shown in Figure 5-6. The sharp peaks and dips in the accelerance magnitude indicate estimates with low coherence.

5.2.1.2 Multiple Response Points

Suppose that a system is excited by a single input, but now there are multiple outputs y_i (i = 1 to n) rather than a single output (see Figure 5-7). In terms of the example presented at the end of the previous section, this would be the case where the plate is excited at a single point and the acceleration of the plate is measured at several points. We define a frequency

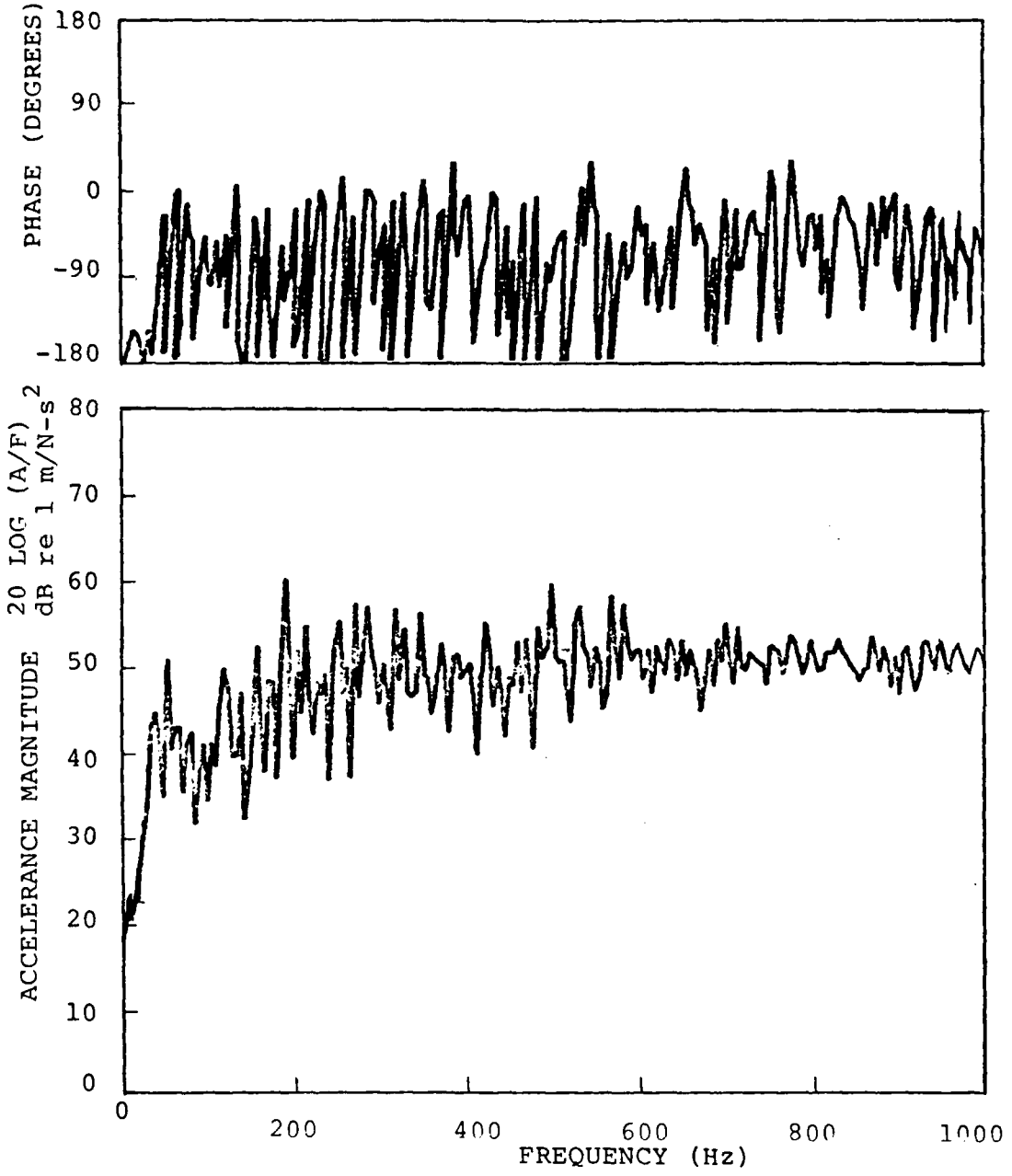


Figure 5-6. Measured Drive Point Accelerance of Flat Plate

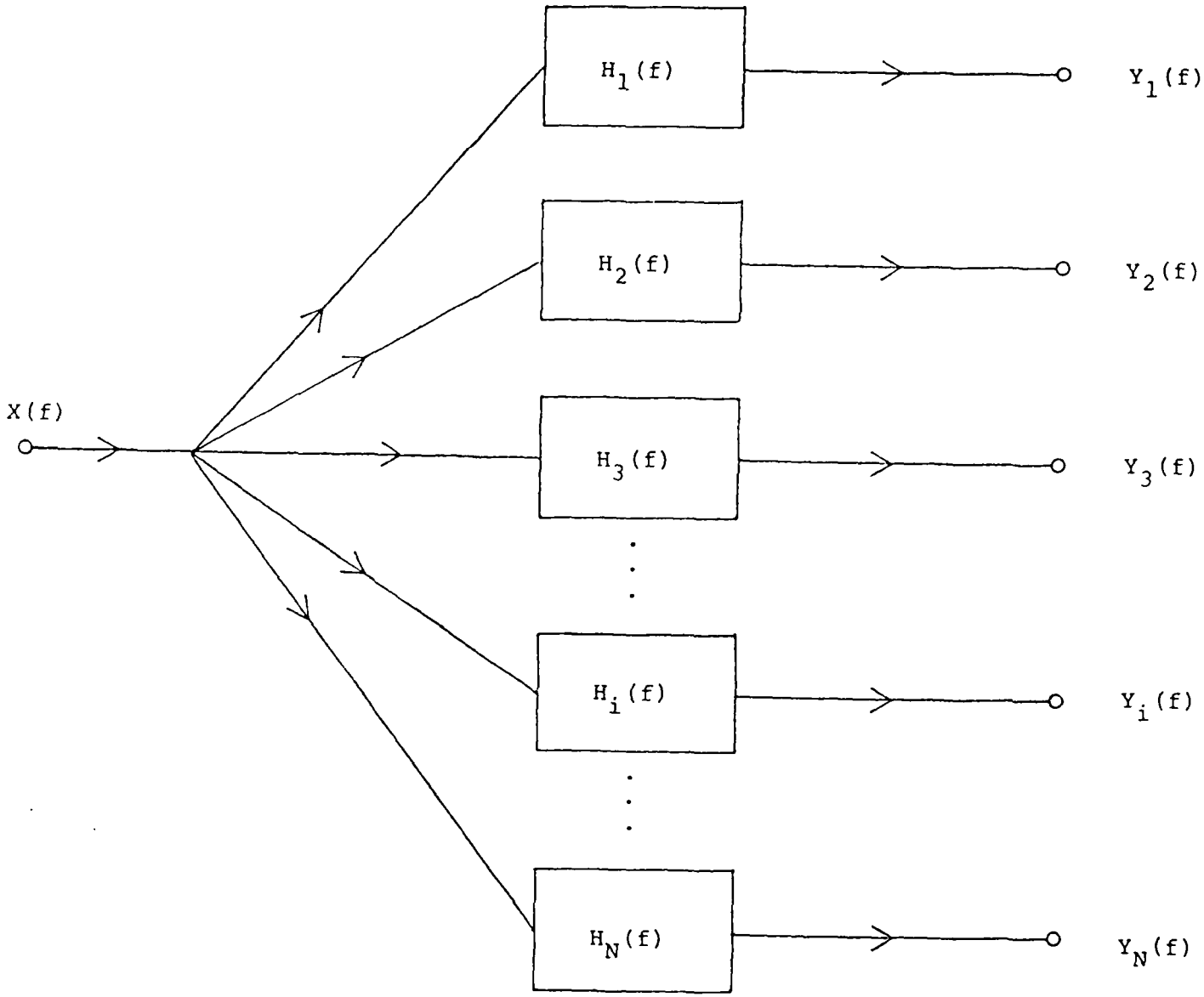


Figure 5-7. Single Input, Multiple Output System

response function $H_i(f)$ which is the ratio of the Fourier Transform $Y_i(f)$ of the i-th output to the transform of the input as follows:

$$H_{y_i|x}(f) = Y_i(f)/X(f) \qquad (5\text{-}43)$$

This definition is completely analogous to Equation (5-1) for the single input-single output case. All of the material on this section concerning digital signal processing estimation of the frequency response function, the effect of transducers, and the presence of noise is directly applicable to the case of multiple outputs. One need only replace the single output variable y by any of the multiple output variables y_i in any of the formulas. In particular, the estimate $H_{y_i|x}(f)$ of the frequency response function between the input and the i-th output is computed using the spectral density functions measurements as follows:

$$H_{y_i|x}(f) = \frac{G_{xy_i}(f)}{G_{xx}(f)} \qquad (5\text{-}44)$$

Single input, multiple output systems contain no new problems that are not reducible to single input, single output systems.

5.2.2 Multiple, Incoherent Excitations

Consider a system which is excited by multiple inputs $x_i(t)$, $i = 1, \ldots, n_x$. The single output variable y(t) measures the response of the system due to all of the input excitations. This section is concerned with the special situation in which the inputs are incoherent. This means that the coherence function $\gamma_{ij}^2(f)$ between the i-th and j-th input is zero for each pair of inputs as follows:

$$\gamma_{ij}^2(f) = \frac{|G_{ij}(f)|^2}{G_{ii}(f)\ G_{jj}(f)} = 0 \qquad (5\text{-}45)$$

Equivalently, the two sources are incoherent when their cross-spectral density function or cross-correlation function is zero. The case where all the inputs in the system are coherent ($\gamma_{ij}^2(f) = 1$) is discussed in Section 5.2.3. Finally, the more complicated situation where there exists partial coherence ($0 < \gamma_{ij}^2 < 1$) between the inputs is discussed in Section 5.3.

As in the earlier sections of this chapter, the system is assumed to be linear and the inputs are assumed to be stationary, ergodic excitations with zero mean value. Since the effect of the transducers on the signal processing for the multiple input system is handled in an identical fashion to that for the single input system (see Section 5.1.3), then without loss of generality we can assume that the transducers are ideal.

All of the essential effects of multiple inputs can be considered using a system of two inputs and a single output, as shown in Figure 5-8. The frequency response function $H_{v|1}(f)$ is associated with the input $U_1(f)$, and similarly $H_{v|2}(f)$ with $U_2(f)$. In addition, there are noise signals $M_1(f)$ and $M_2(f)$ present at each of the inputs and noise $N(f)$ present at the output. The measured input and output signals are $X_i(f)$ and $Y(f)$, respectively. It is assumed that all of the three noise signals are uncorrelated with respect to the true input signals $U_1(f)$ and $U_2(f)$ and the true output signal $V(f)$.

The basic relations of the system in Figure 5-8 are

$$X_1(f) = U_1(f) + M_1(f)$$

$$X_2(f) = U_2(f) + M_2(f)$$

$$Y(f) = H_{v|1}(f)\ U_1(f) + H_{v|2}(f)\ U_2(f) + N(f) \qquad (5\text{-}46)$$

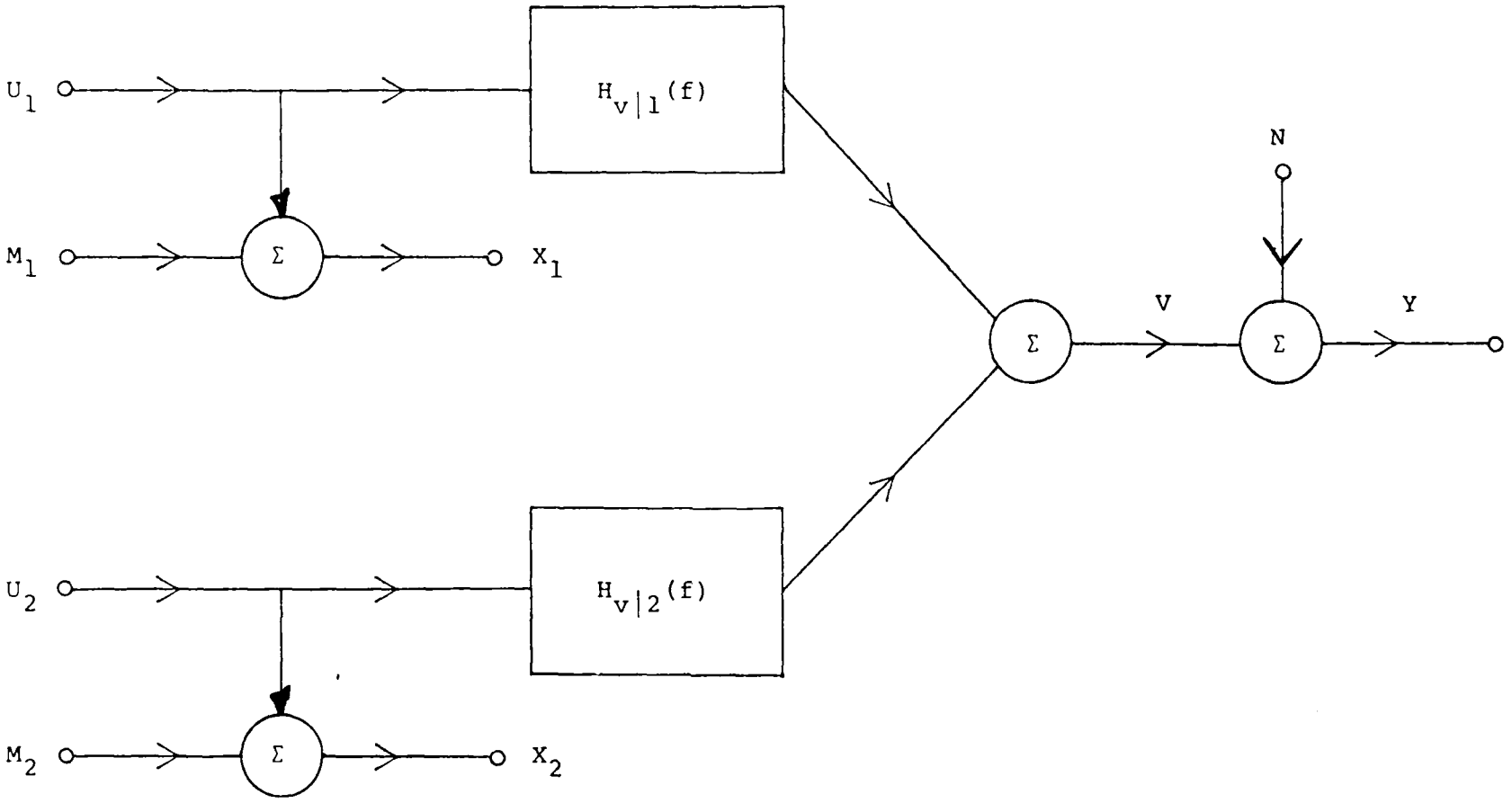


Figure 5-8. Two Input, Single Output System

The cross-spectral density function $G_{1y}(f)$ between the measured input $X_1(f)$ and measured output $Y(f)$ is calculated using the definitions of the spectral densities given in Section 5.1.4 with the result

$$G_{1y}(f) = H_{v|1}(f)\ G_{u_1u_1}(f) + H_{v|2}(f)\ G_{u_1u_2}(f) + G_{m_1n}(f) \qquad (5\text{-}47)$$

A similar equation for $G_{2y}(f)$ may be written by interchanging the subscripts 1 and 2. The system equations are written below with general indices i and j, rather than 1 and 2 in order to make the transition to the case of more than two system inputs. The cross-spectral density function $G_{u_iv}(f)$ between the true i-th input $U_i(f)$ and true output $V(f)$ is

$$G_{u_iv}(f) = H_{v|j}(f)\ G_{u_iu_j}(f). \qquad (5\text{-}48a)$$

This equation with i=1 is,

$$G_{u_1v}(f) = H_{v|1}(f)\ G_{u_1u_1}(f) + H_{v|2}(f)\ G_{u_1u_2}(f) \qquad (5\text{-}48b)$$

The above equations involve the unknown frequency response functions which we want to measure, but they also involve the cross-spectral density functions $G_{u_iu_j}$ between the various inputs to the system.

If all the inputs to the system are incoherent, then the cross-spectral density functions for each pair of inputs are zero. For the example in Figure 5-8, this implies

$$G_{u_1u_2}(f) = 0 \qquad (5\text{-}49)$$

Hence Equations (5-47, 5-48 and 5-49) imply

$$G_{iy}(f) = G_{u_iv}(f) + G_{m_in}(f) \qquad (5\text{-}50)$$

The autospectral density function G_{ii} associated with the measured input $X_i(f)$ is

$$G_{ii}(f) = G_{u_i u_i}(f) + G_{m_i m_i}(f), \tag{5-51a}$$

and the output autospectral density computed using Eq. (5-46) under the condition that the inputs are incoherent with one another as well as with the noise is

$$G_{yy}(f) = |H_{v|1}(f)|^2 G_{u_1 u_1}(f) + |H_{v|2}(f)|^2 G_{u_2 u_2}(f) + G_{nn}(f). \tag{5-51b}$$

The measured i-th frequency response function $H_{y|i}(f)$ is defined by analogy with Equation (5-25) and its value is

$$H_{y|i}(f) = \frac{G_{iy}(f)}{G_{ii}(f)} = H_{v|i}(f) \left\{\frac{1 + G_{m_i n}/G_{u_i v}}{1 + \alpha_i}\right\} \tag{5-52}$$

where α_i is the noise-to-signal ratio of the i-th input defined as follows:

$$\alpha_i(f) = [G_{m_i m_i}(f)/G_{u_i u_i}(f)] \tag{5-53}$$

This equation relates the measured frequency response function to the true frequency response function $H_{v|i}$ for the i-th subsystem. This equation for the incoherent multiple input case is equivalent to Equation (5-26) for the single input case. Although the discussion was only given for two incoherent inputs, it is precisely equivalent for any number of incoherent inputs.

This result is the important point of this section. If one

has a system which is acted on by multiple incoherent inputs, the resulting frequency response functions of the system are measured in a manner exactly analogous to the single input system. It does not even matter if there are multiple outputs. These multiple outputs are considered at the end of Section 5.2.1, where it is seen that the multiple output case may be reduced to a combination of single output systems. The resulting error analysis is identical to that considered in Section 5.2.1.

The main practical question which must be determined is whether or not the inputs are incoherent. This question is answered by computing the various coherence functions between the input signals. If all the coherence functions are zero, then the system can be reduced to an equivalent combination of single input, single output subsystems.

Any measurement of the i-th frequency response function $H_{y|i}(f)$ is almost always combined with a measurement of the coherence function $\gamma_{iy}^2(f)$ between the i-th input and the output defined as follows:

$$\gamma_{iy}^2(f) = \frac{|G_{iy}|^2}{G_{ii}G_{yy}} \tag{5-54}$$

For the two input system shown in Figure 5-8 the coherence function between $X_1(f)$ and $Y(f)$ is found from Eqs. (5-48,50,51 and 53) to be,

$$\gamma_{1y}^2(f) = \frac{|H_{v|1}|^2 G_{u_1u_1} \left|1 + \frac{G_{m_1n}}{G_{u_1v}}\right|^2}{(1+\alpha_1)[|H_{v|1}|^2 G_{u_1u_1} + |H_{v|2}|^2 G_{u_2u_2} + G_{nn}]} \tag{5-55}$$

A corresponding equation to the above for input 2 is obtained by interchanging the subscripts 1 and 2. It was shown in Section 5.2.1 that the coherence function between the measured input and

output signals is unity if there is no input or output noise present. If there is no noise present in the two input case, then the coherence function γ_{1y}^2 is not equal to unity as can be seen from Equation (5-55). In the absence of noise, the coherence functions satisfy the following relation:

$$1 = \gamma_{1y}^2(f) + \gamma_{2y}^2(f) \qquad (5\text{-}56)$$

The physical meaning of these coherence functions is that $\gamma_{1y}^2 G_{yy}$ is the output spectral density due to input 1, and similarly for input 2. Equation (5-56) is the statement that the output is caused either by input 1 or input 2. Measurement of the coherence functions provides information concerning which input is producing the largest response at the output measurement location. Similar considerations pertain for the case of more than two incoherent inputs.

The relative random error in the estimate of the magnitude of the frequency response function is given by a formula similar to Equation (5-42) for the single input/single output case as follows:

$$\varepsilon_r[|H_{y|i}|] = \frac{(1-\gamma_{iy}^2)^{1/2}}{|\gamma_{iy}|\sqrt{2MK}} \qquad (5\text{-}57)$$

The bias errors are calculated in a manner analogous to the single input case.

5.2.3 Multiple, Coherent Excitations

Now consider the case of a multiple input, single output system in which all the inputs are (perfectly) coherent. Without loss of generality, the case of two inputs is considered (see Figure 5-9a). Two inputs are said to be coherent when the coherence function $\gamma_{12}^2(f)$ between the two inputs is equal to

unity at all frequencies in the bandwidth of interest. This is the opposite case to the one considered in Section 5.2.2 in which the coherence function between the inputs is zero. The general case of partial coherence between the inputs is considered in Section 5.3.

A coherence function of unity implies that the two inputs are completely dependent. One method of stating this is to say that there must exist a frequency response function $H_{2|1}(f)$ (see Fig 5-9b) which connects the two inputs. In this way, the output signal depends only on the value of one of the inputs. Thus the two input system reduces to a single input system when the inputs are coherent, and similarly for any number of coherent inputs.

The statement that the coherence function between the inputs of a two input system equals unity implies

$$G_{21} = G_{12}^{*} = \frac{G_{11}\, G_{22}}{G_{12}} \qquad (5\text{-}58)$$

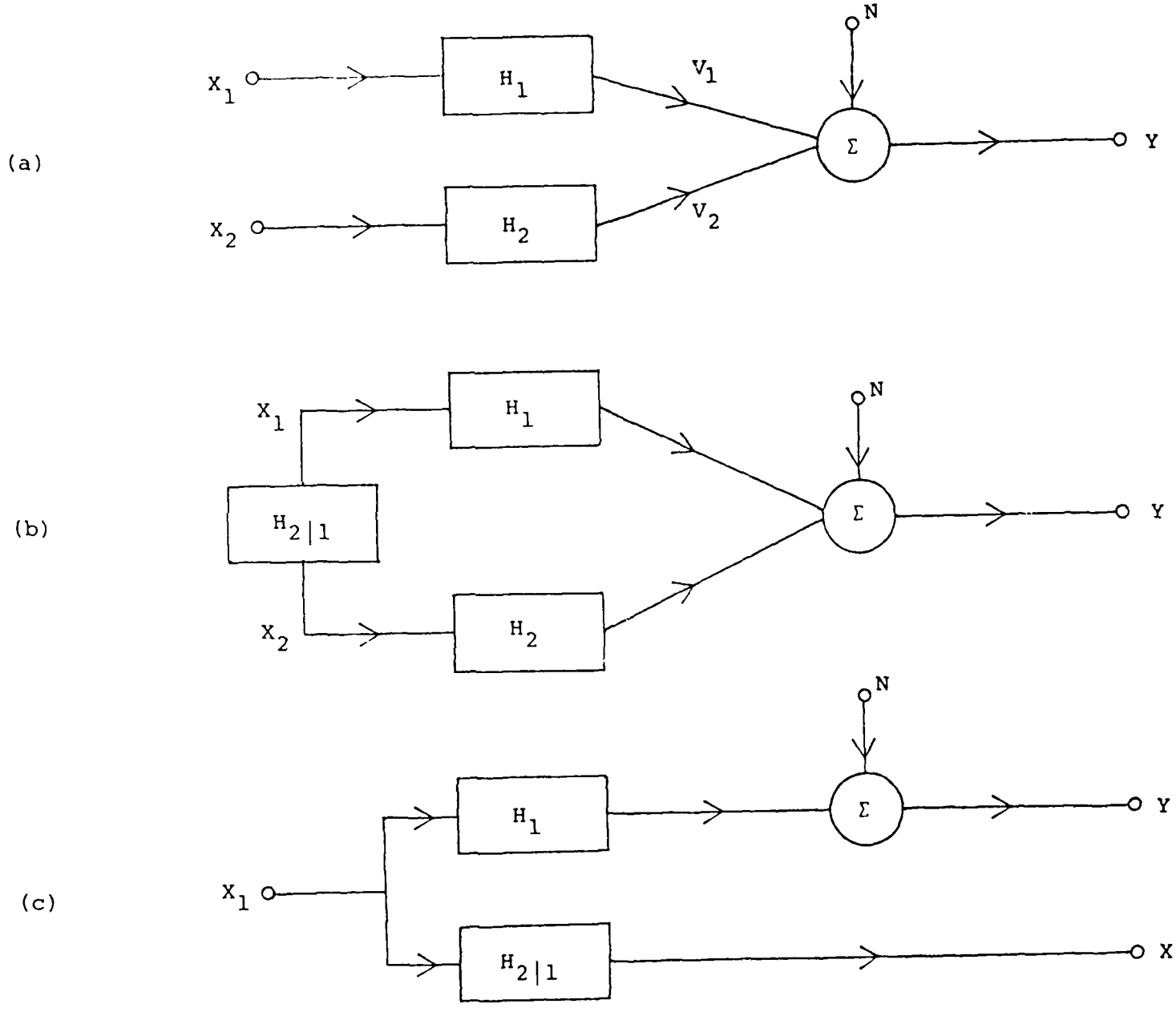
The cross-spectral density between the inputs and output is computed with the help of the above equation to be

$$G_{1y} = H_1 G_{11} + H_2 G_{12}$$

$$G_{2y} = H_1 G_{21} + H_2 G_{22}$$

$$= \left(\frac{G_{22}}{G_{12}}\right) G_{1y} \qquad (5\text{-}59)$$

These equations, along with the definition of the input/output coherence functions, imply


(a)
X_1
H_1
V_1
N
Σ
Y
X_2
H_2
V_2
(b)
X_1
H_1
N
$H_{2|1}$
Σ
Y
X_2
H_2
(c)
N
H_1
Σ
Y
X_1
$H_{2|1}$
X_2

$$\gamma_{1y}^2 = \gamma_{2y}^2 \tag{5-60}$$

The above equation means that each input contributes equally to the output. Since both inputs are coherent, this is to be expected.

Figure 5-9a shows the general two input system with noise present at the output and Figure 5-9b shows the modification in the system when the two inputs are coherent. In this case the frequency response function $H_{2|1}$ representing the deterministic link between the two inputs is present. This figure shows that a single frequency response function $H_{y|1}$ fully describes the system response as follows:

$$H_{y|1} = H_1 + H_{2|1}H_2 \tag{5-61}$$

Figure 5-9c shows a third way of describing the system. It represents the coherent input system as a single input, double output system. This example clearly points out some of the pitfalls in interpreting multiple input/output systems. It is not always obvious whether a given system response should be classified as an input or an output. An alternative formulation exists which eliminates the oftentimes arbitrary distinction between what is an input or output [5-3].

The errors for the multiple coherent input systems are calculated [5-2] in a manner analogous to the single input case.

5.3 INDIRECT ESTIMATION OF FREQUENCY RESPONSE FUNCTION

5.3.1 Decomposition of Multiple, Partially Coherent Excitations

This section describes how to analyze a multiple input system in which the inputs are partially coherent. The two special cases in which the inputs are fully coherent or incoherent are

discussed in Section 5.2. Two inputs are said to be partially coherent (or partially correlated) when the coherence function between the two inputs defined in Equation (5-45) has a value intermediate between zero and unity. The same assumptions regarding the inputs as those described near the beginning of Section 5.2.2 apply also in this section. In addition, it is assumed that the only noise present in the system is at the output of the system, and this noise is uncorrelated with any of the inputs.

In general practice, one should compute the coherence functions between the various inputs and the output. If any such coherence function is zero, then that input makes no contribution to the output and thus can be dropped from the multiple input model. Along the same lines, the coherence functions between all pairs of inputs should be calculated. If any such coherence function is unity, then the two inputs are causally related and can be replaced by a single input.

Reduction of a multiple input system with partially coherent excitations to an equivalent system with incoherent inputs can sometimes be accomplished without recourse to the analytic methods of this section. This may be true, for example, in the case of a panel excited by a turbulent boundary layer. The pressure fluctuations are distributed in space with a certain characteristic space-time correlation [5-4]. The excitation at points on the panel which are separated by a distance larger than the correlation length are statistically independent. That means that transducers located at distances larger than the correlation length give rise to incoherent excitations. Thus, prior knowledge of the statistical character of the distributed source excitation can be used to determine the most efficient spacing of transducers. In this situation the techniques described in Section 5.2 may be used to evaluate the system frequency response functions rather than the more powerful but also more complicated

techniques described in this section.

Another related factor is that the desired system response parameter is frequently not for a single point, but rather a spatial average of a response parameter over some part of a structure. This is particularly true when one is using a statistical analysis in a frequency range of high modal density of a structure, and therefore one is also averaging the response over a frequency band (such as third-octave). It has been shown that by averaging the response of a dynamic system over frequency and space the correlated response of the system averages to zero. This leads to a very simple model of the system (see Chapter 8 on S.E.A.).

The analytical techniques used in this section transform the original set of inputs to an equivalent set of incoherent inputs. This decomposition is not unique. These calculational procedures are much more efficient than older methods of treating partial coherence using least-squares prediction techniques [5-5, 5-6].

The underlying concepts are introduced by considering a two input/single output system. The system is illustrated in Figure 5-10 and the linear system equations are

$$Y(f) = H_1(f)\ X_1(f) + H_2(f)\ X_2(f) + N(f) \qquad (5\text{-}62a)$$

$$G_{yy} = |H_1|^2 G_{11} + |H_2|^2 G_{22} + G_{nn} + H_1 H_2^* G_{21} + H_2 H_1^* G_{12} \qquad (5\text{-}62b)$$

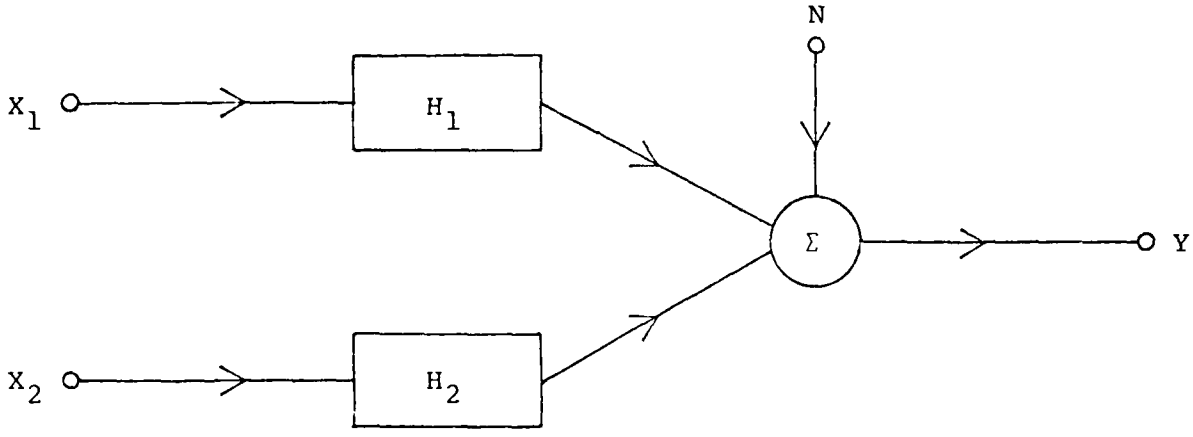


Figure 5-10. General Two Input System

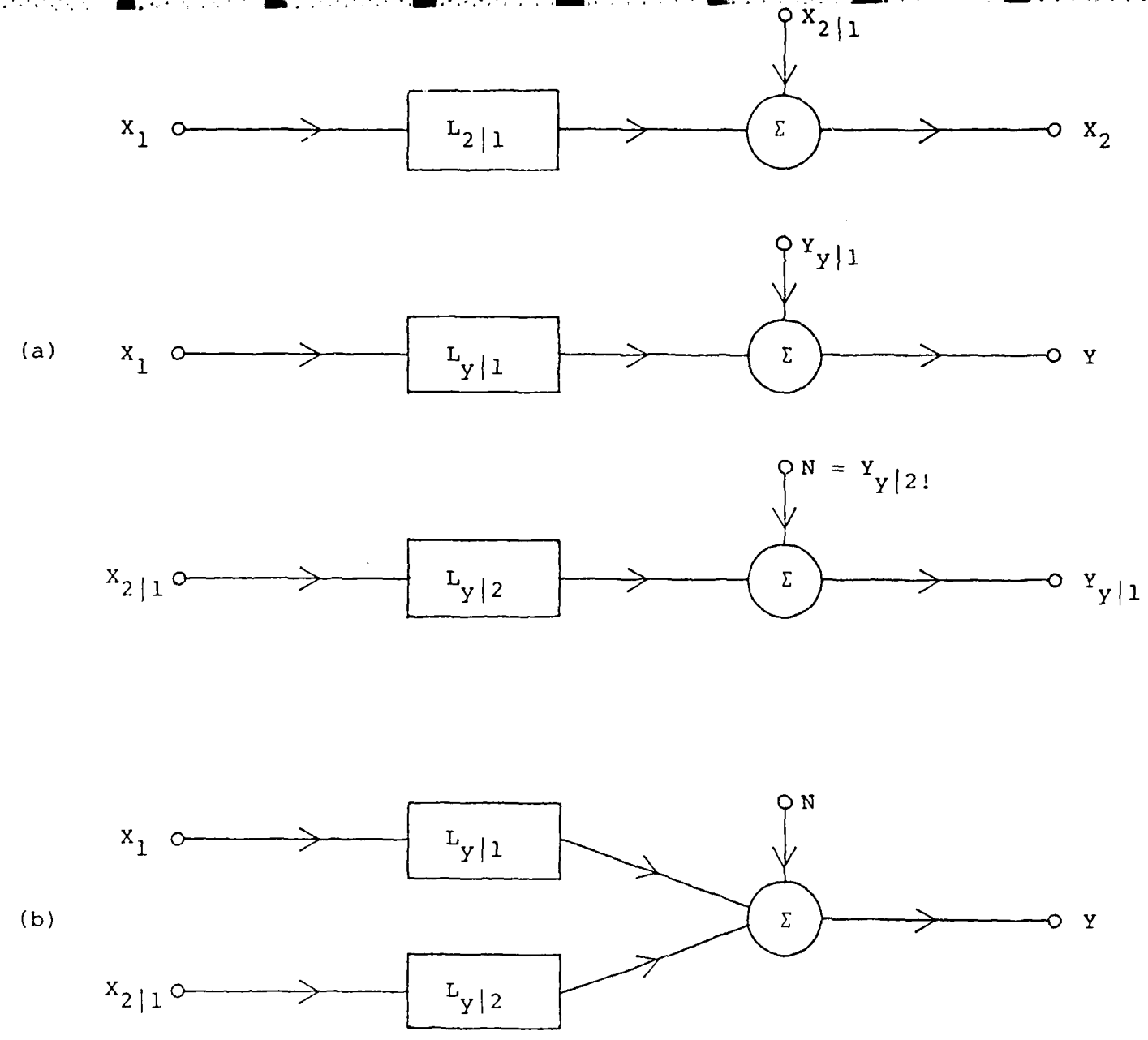


Figure 5-11. Equivalent Two Input System

The two inputs are $X_1(f)$ and $X_2(f)$, the output is $Y(f)$, the output noise signal is $N(f)$, and $H_1(f)$ and $H_2(f)$ are the system frequency response functions which are to be determined in terms of the various spectral and cross-spectral density functions.

The decomposition of the inputs into two equivalent incoherent inputs proceeds as follows. Write input X_2 as a sum of two terms, one of which is coherent with X_1 and the other which is incoherent. This relation is

$$X_2(f) = L_{2|1}(f)\ X_1(f) + X_{2|1}(f) \qquad (5\text{-}62c)$$

A more specialized notation for the subscripts is used in this section as follows: $X_{2|1}$ is that part of X_2 which is <u>incoherent</u> with X_1. Section 5.2.3 demonstrates that there exists an equivalent frequency response function between any pair of fully coherent inputs. Thus, the portion of X_2 which is coherent with X_1 may be written as $L_{2|1}X_1$. The equivalent subsystem is illustrated in Figure 5-11.

Similarly, the output signal Y may be written as a sum of two signals $Y_{y|1}$, which is incoherent with X_1, and $L_{y|1}X_1$, which is coherent with X_1, as follows:

$$Y(f) = L_{y|1}(f)\ X_1(f) + Y_{y|1}(f) \qquad (5\text{-}62d)$$

This equivalent subsystem is also illustrated in Figure 5-11. The equivalent frequency response functions $L_{2|1}$ and $L_{y|1}$ are computed from the measured spectral densities with the result

$$L_{2|1}(f) = G_{12}(f)/G_{11}(f) \qquad (5\text{-}63a)$$

$$L_{y|1}(f) = G_{1y}(f)/G_{11}(f) \qquad (5\text{-}63b)$$

Proceeding along similar lines, the signal $Y_{y|1}$ may be written as a sum of a signal $L_{y|2}X_{2|1}$ which is coherent with $X_{2|1}$, and a signal $Y_{y|2!}$ which is incoherent with both X_1 and X_2. The factorial notation (2!) in the subscript refers to a signal which is incoherent with inputs 1 and 2. The only excitation left in the original system of Figure 5-10 which is incoherent with both input signals is the output noise N. Thus, we write

$$N(f) = Y_{y|2!}(f) = Y_{y|1}(f) - L_{y|2}(f)\ X_{2|1}(f) \qquad (5\text{-}64)$$

Let $G_{22|1}$ be the spectral density function of the portion of the input X_2 which is incoherent with X_1, and let $G_{2y|1}$ be the cross-spectral density between $X_{2|1}$ and $Y_{y|1}$. Then the equivalent frequency response function $L_{y|2}$ is

$$L_{y|2}(f) = G_{2y|1}(f)/G_{22|1}(f) \qquad (5\text{-}65)$$

Inspection of Figures 5-10 and 5-11b show that $L_{y|2} = H_2$.

The representation of the output in the equivalent system analogous to the original expression Equation (5-62b) is

$$Y = L_{y|1}X_1 + L_{y|2}X_{2|1} + N \qquad (5\text{-}66a)$$

$$G_{yy} = |L_{y|1}|^2 G_{11} + |L_{y|2}|^2 G_{22} + G_{nn} \qquad (5\text{-}66b)$$

$$G_{nn} = G_{yy}\ [(1 - \gamma_{1y}^2)(1 - \gamma_{2y|1}^2)] \qquad (5\text{-}66c)$$

There are no troublesome cross-terms in Equation (5-66b) as there were in the original expression for G_{yy}. The coherence functions in the above are defined in the usual way by

$$\gamma_{1y}^2 = \frac{|G_{1y}|^2}{G_{11}G_{yy}} \tag{5-67a}$$

$$\gamma_{2y|1}^2 = \frac{|G_{2y|1}|^2}{G_{22|1}G_{yy|1}} \tag{5-67b}$$

Equation (5-66c) can be expressed intuitively. The total output which is incoherent with X_1 is $G_{yy}(1 - \gamma_{1y}^2)$, and similarly the output incoherent with both X_1 and $X_{2|1}$ is $G_{yy}(1 - \gamma_{1y}^2)(1 - \gamma_{2y|1}^2)$. This portion of the output signal is precisely the noise signal.

The equivalent frequency response function may be used to calculate the system frequency response functions as follows:

$$H_2(f) = L_{y|2}(f) \tag{5-68a}$$

$$H_1(f) = L_{y|1}(f) - L_{2|1}(f)\ H_2(f) \tag{5-68b}$$

Therefore, the equivalent frequency response functions $L_{2|1}$, $L_{y|1}$ and $L_{y|2}$ are evaluated using Eqs. (5-63a) through experimental measurements of the spectral and cross-spectral densities, and the true system frequency response functions may then be obtained using Equation (5-68).

The multiple coherence function $\gamma_{y:x}^2$ is defined as the ratio of the output energy due to all the inputs to the total output energy, as follows:

$$\gamma_{y:x}^2(f) = \frac{G_{yy} - G_{nn}}{G_{yy}} \tag{5-69}$$

The multiple coherence function is found from Equation (5-66c) to be

$$\gamma_{y:x}^2 = 1 - [(1 - \gamma_{1y}^2)(1 - \gamma_{2y|1}^2)] \qquad (5\text{-}70)$$

The discussion of the two input system is concluded by pointing out the operational significance of the equivalent frequency response functions shown in Figure 5-11a. The spectral output at Y due to all the system inputs which are coherent with X_1 is $|L_{y|1}|^2 G_{11}$. This output includes not only all of input 1, but also the part of input 2 which is coherent with input 1. The spectral output at Y due to all the remaining system inputs which are coherent with input 2 is $|L_{y|2}|^2 G_{22|1}$. That this equivalent decomposition of the original two input system is not unique may be seen by interchanging the subscripts 1 and 2 in Figure 5-11 and all the above formulas in this section. Nevertheless, the consistent use of either decomposition leads to estimates of the true system frequency response functions. This concludes the discussion of the two input/single output system.

The general multiple input/single output system [5-2] is illustrated in Figure 5-12. The reduction of the original system to an equivalent system with incoherent outputs proceeds exactly as for the two input system. The first step is to decompose all q inputs (1, 2, 3, ..., q) into a part which is coherent with input 1 and a part which is incoherent, exactly as was carried out in Equation (5-62c) for the two input system. After this step the system consists of input X_1 and (q-1) inputs which are incoherent with X_1; these are $X_{2|1}$, $X_{3|1}$, ..., $X_{q|1}$. The next step is to decompose the (q-1) inputs into the part coherent with X_2 but incoherent with X_1, and (q-2) inputs which are incoherent with both X_1 and X_2; namely, $X_{3|2!}$, $X_{4|2!}$, ..., $X_{q|2!}$. This procedure continues until an equivalent q-input system, all of whose inputs are incoherent, is obtained as illustrated in Figure 5-13. The resulting so-called <u>conditioned inputs</u> are X_1, $X_{2|1}$, $X_{3|2!}$, ..., $X_{i|(i-1)!}$, and $X_{q|(q-1)!}$.

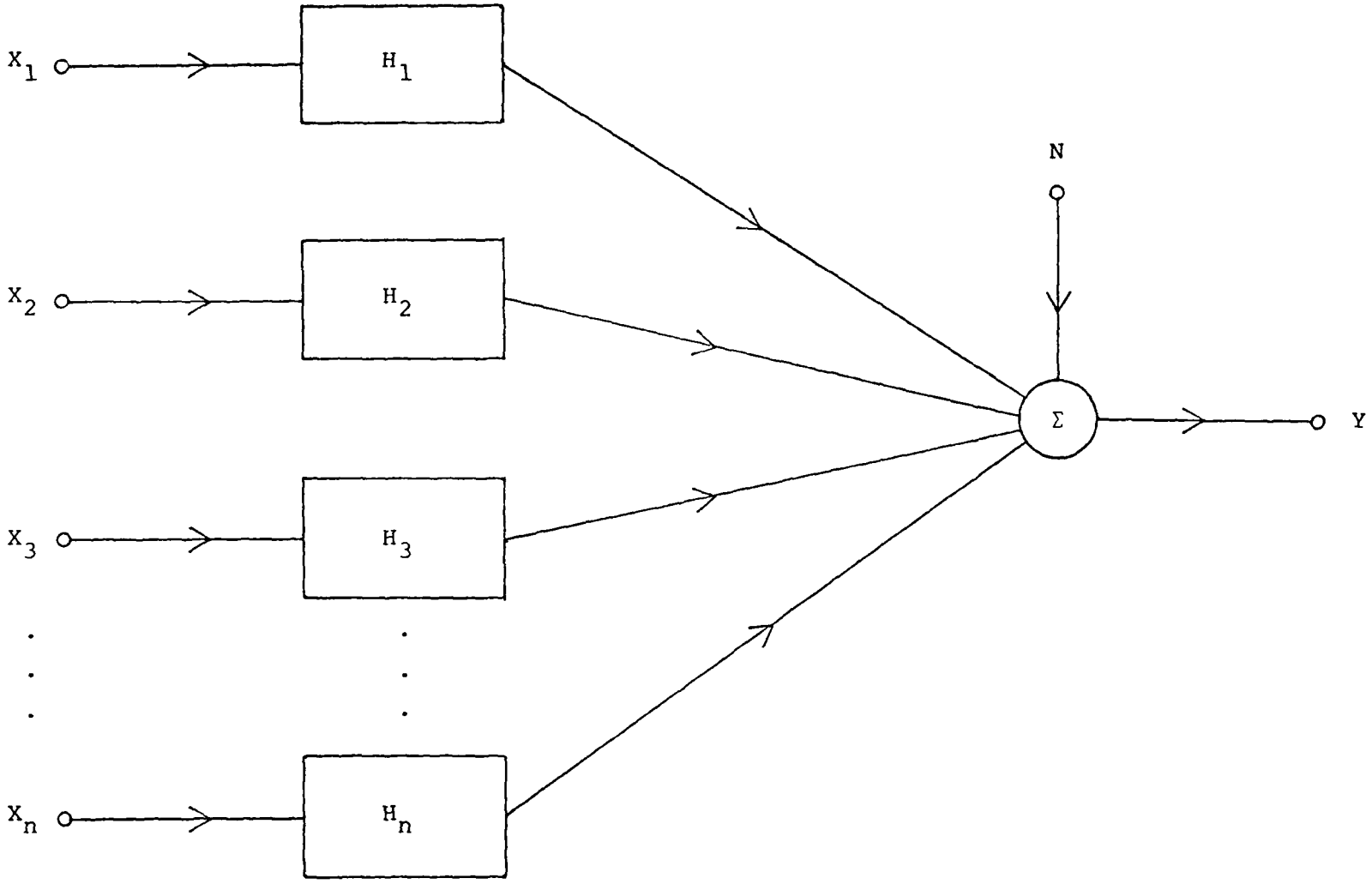


Figure 5-12. General Multiple Input/Single Output Model

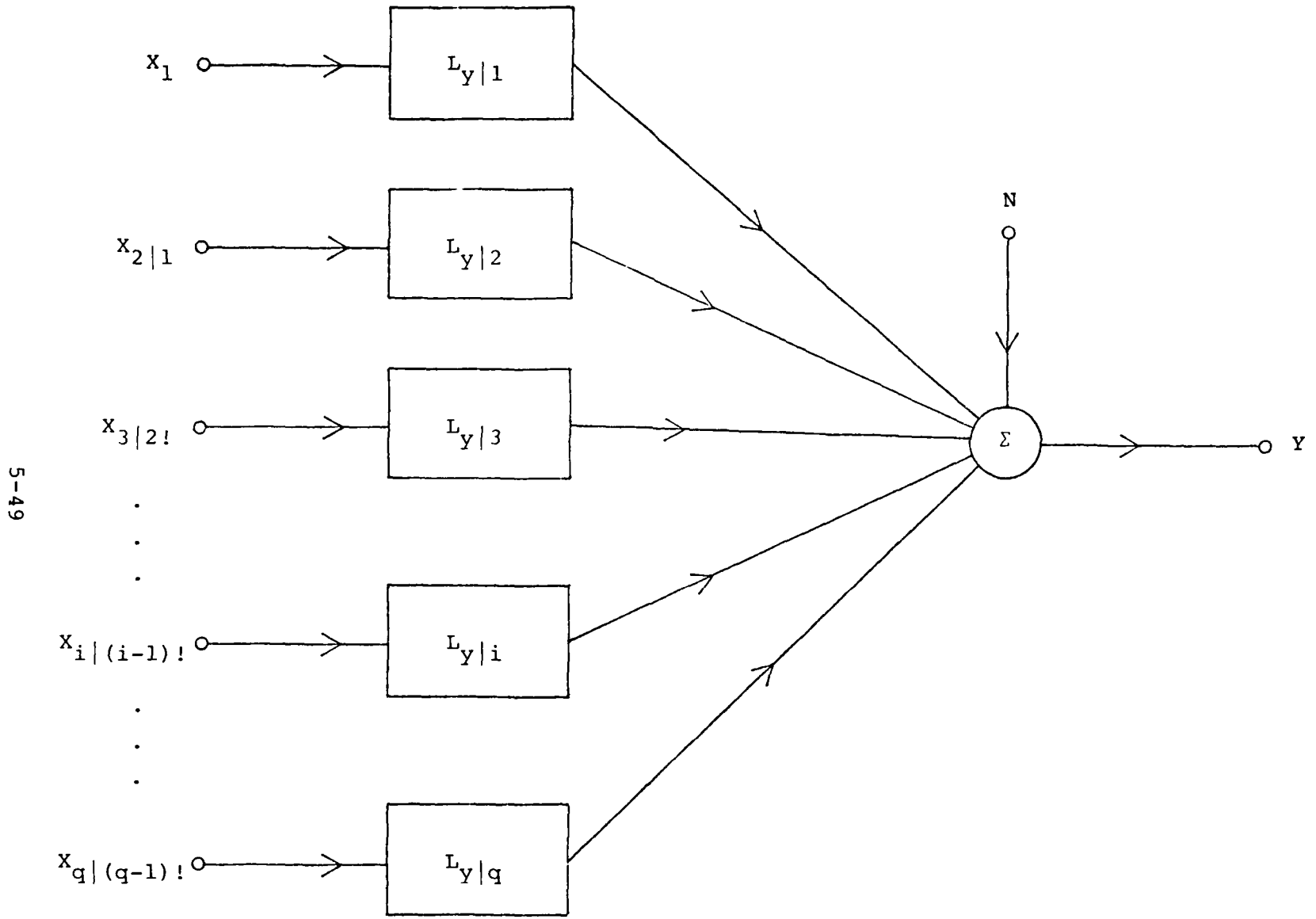
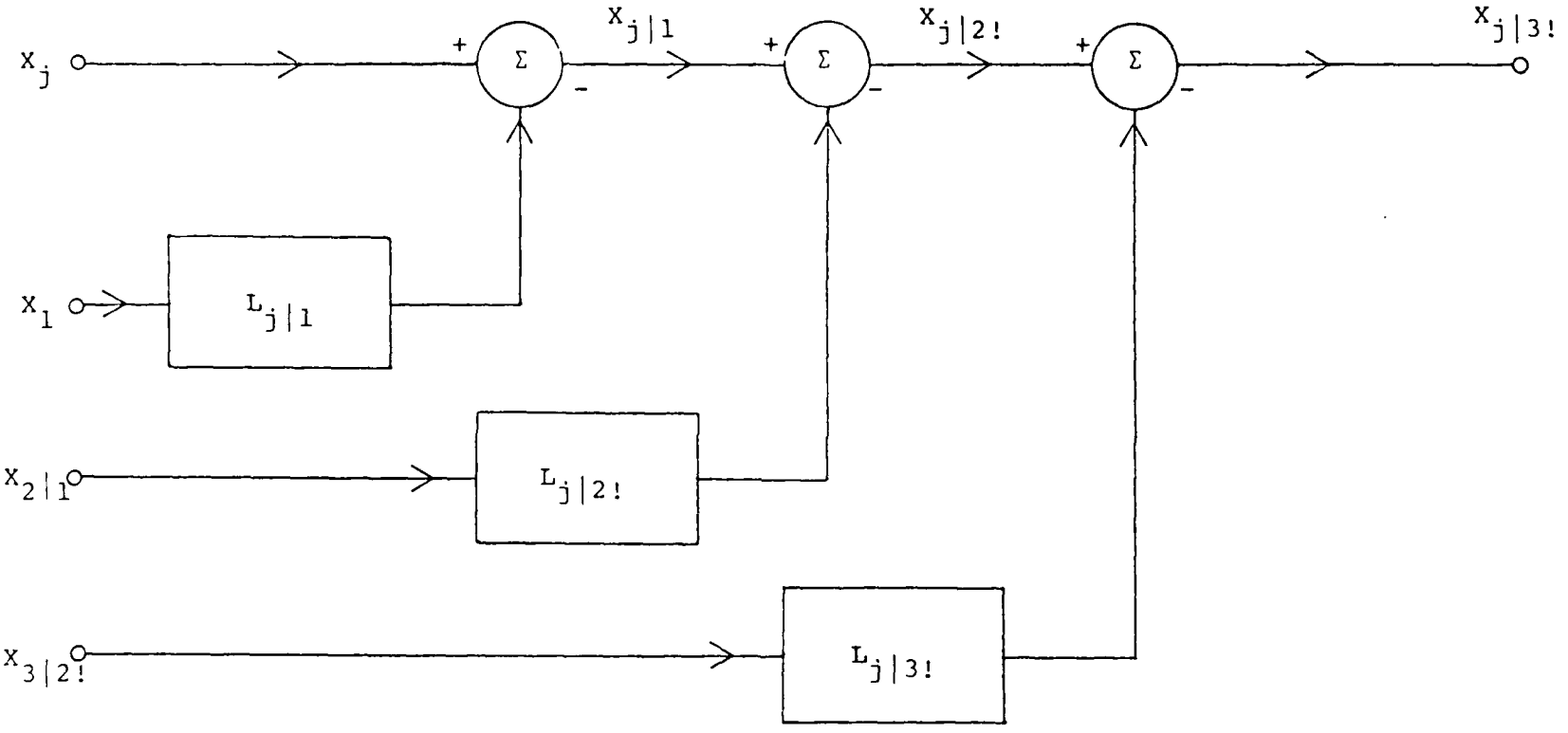


Figure 5-13. Multiple Input/Single Output System for Equivalent Incoherent Inputs



(Example for $j \geq 4$)

Figure 5-14. Reduction of System Input to Equivalent Input

The equivalent frequency response functions $L_{y|i}$ in Figure 5-13 are calculated from

$$L_{y|i} = G_{iy|(i-1)!}/G_{ii|(i-1)!} \quad i = 1, \ldots, q \tag{5-71}$$

For i = 1, 2 these equations reduce to Equations (5-63b) and (5-65). The spectral density function $G_{ii|(i-1)!}$ corresponds to the i-th equivalent input which is constructed to be incoherent with inputs 1, 2, ..., (i-1), and similarly for the cross-spectral density function $G_{iy|(i-1)!}$. Algorithms for computing these conditioned spectral density functions in terms of the measured ones are given below.

Any true input X_j to the system goes through a procedure in which the coherent parts of the first (j-1) inputs are subtracted off from the j-th input as illustrated in Figure 5-14. The analogous procedure for the two input system is illustrated in Figure 5-11a, and Equation (5-62c) is the algorithm for that case. The equivalent frequency response function $L_{j|i}$ connects the coherent part of the i-th and j-th inputs conditioned on the first (i-1) inputs for i < j. For j = 4, the equations represented by Figure 5-13 are

$$X_{4|1} = X_4 - L_{4|1}X_1 \tag{5-72a}$$

$$X_{4|2!} = X_{4|1} - L_{4|2!}X_{2|1} \tag{5-72b}$$

$$X_{4|3!} = X_{4|2!} - L_{4|3!}X_{3|2!} \tag{5-72c}$$

where the conditioned inputs in the above are

$$X_{2|1} = X_2 - L_{2|1}X_1 \tag{5-72d}$$

$$X_{3|2!} = X_{3|1} - L_{3|2!}X_{2|1}$$

$$= [X_3 - L_{3|1}X_1] - L_{3|2!}[X_2 - L_{2|1}X_1] \qquad (5\text{-}72e)$$

This example shows how all the conditioned inputs ($X_{2|1}$, $X_{4|3!}$, etc.) may be expressed in terms of the true system inputs and the equivalent frequency response functions $L_{j|i!}$. The general algorithm connecting the conditioned inputs to the true inputs is

$$X_{j|k!} = X_{j|(k-1)!} - L_{j|k!}X_{k|(k-1)!} \qquad (5\text{-}73)$$

It is necessary to determine the $L_{j|i!}$ functions in order to make use of the above algorithm. The iterative algorithm which expresses these functions in terms of the spectral density functions is given below as follows:

$$L_{j|i!} = \frac{G_{ij|(i-1)!}}{G_{ii|(i-1)!}} \qquad \text{for } i < j \qquad (5\text{-}74)$$

If $i > j$, then $L_{j|i!} = 0$, and $L_{i|i!}$ always equals unity for all inputs. Figure 5-15 shows the algorithm connecting the measured spectral density functions G_{ij} with the conditioned spectral density functions $G_{ij|k!}$ by means of the $L_{j|i!}$. The first few terms of this algorithm are

$$G_{12|1} = G_{12} - L_{2|1}G_{11} \qquad (5\text{-}75a)$$

$$G_{13|1} = G_{13} - L_{3|1}G_{11} \qquad (5\text{-}75b)$$

$$G_{13|2!} = G_{13|1} - L_{3|2!}G_{12|1} \qquad (5\text{-}75c)$$

$$G_{23|1!} = G_{23} - L_{3|1}G_{21} \qquad (5\text{-}75d)$$

$$G_{22|1} = G_{22} - L_{2|1}G_{21} \qquad (5\text{-}75e)$$

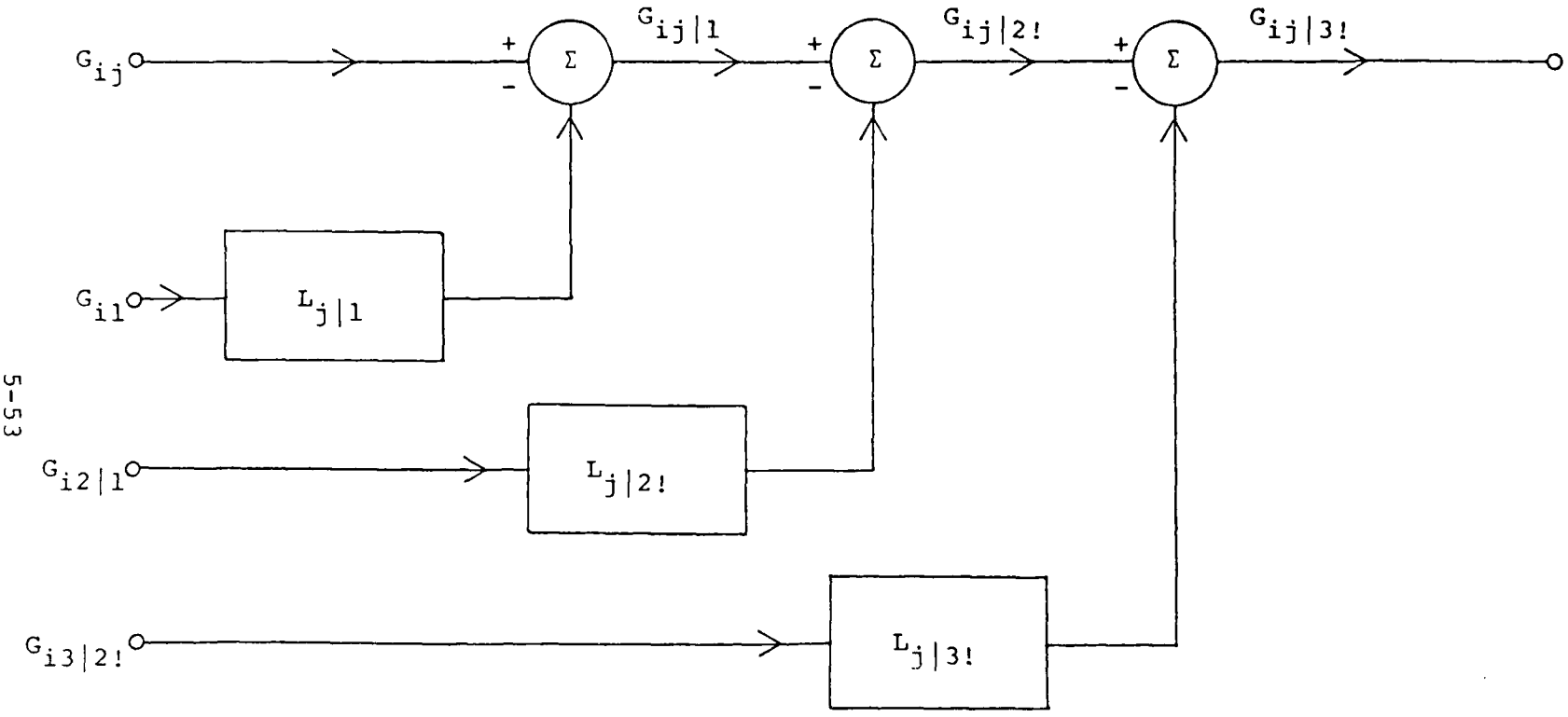


Figure 5-15. Computation of Conditional Spectral Density Functions

The general algorithm of which the above equations are special cases is

$$G_{ij|k!} = G_{ij|(k-1)!} - L_{j|k!}G_{ik|(k-1)!} \tag{5-76}$$

Each higher order conditioned spectral density may ultimately be determined in terms of the measured spectral densities G_{ii} and G_{ij} by this algorithm. Consequently, each $L_{j|i!}$ in Equation (5-74) is calculated from the lower order conditioned spectral densities.

It remains to calculate the conditioned cross-spectral density $G_{iy|(i-1)}$ between the i-th input and the output which occurs in Equation (5-71) for the equivalent frequency response function $L_{y|i!}$. The algorithm is similar to Equation (5-76) with the subscript j replaced by y; that is

$$G_{iy|k!} = G_{iy|(k-1)!} - L_{y|k}G_{ik|(k-1)!} \tag{5-77}$$

The lowest order example of the above is

$$G_{2y|1} = G_{2y} - L_{y|1}G_{21} \tag{5-78}$$

where $L_{y|1}$ is given from Equation (5-71) for i = 1 to be

$$L_{y|1} = G_{1y}/G_{11} \tag{5-79}$$

Given $G_{2y|1}$ and $G_{22|1}$ (from Equations 5-75e and 5-63a), one can calculate $L_{y|2}$ using Equation (5-71) with i = 2.

This process continues until all the equivalent frequency response functions $L_{j|i!}$ and $L_{y|i}$ have been calculated using Equations (5-71) and (5-74) starting from i = 1 and j = 2. The true system frequency response functions are related to the equivalent functions by

$$H_q = L_{y|q} \tag{5-80a}$$

$$H_i = L_{y|i} - \sum_{j=i+1}^{q} L_{j|i!} H_j \qquad i = (q-1), (q-2), \ldots, 1 \tag{5-80b}$$

Equation (5-68) is the corresponding pair of relations for the two input system.

The partial coherence function $\gamma_{iy|k!}^2$ between the i-th input conditioned on the first k inputs and the output is

$$\gamma_{iy|k!}^2 = \frac{|G_{iy|k!}|^2}{G_{ii|k!}\, G_{yy|k!}} \qquad \text{for } k < i \tag{5-81}$$

In order to compute this partial coherence function it is necessary to compute the output autospectral density function $G_{yy|k!}$ conditioned on the first k inputs, which is equal to

$$G_{yy|k!} = G_{yy|(k-1)!} - |L_{y|k}|^2 \, G_{kk|(k-1)!} \tag{5-82}$$

The autospectral density G_{nn} of the output noise signal is by definition that part of the output signal density which is incoherent with all q of the inputs. It is expressed in terms of the measured output autospectral density and the partial coherence functions of all q inputs (i = 1, 2, ..., q) conditioned on the first (i-1) inputs, respectively, by

$$G_{nn} = G_{yy|q!} = G_{yy}[1-\gamma_{1y}^2][1-\gamma_{2y|1}^2] \ldots [1-\gamma_{qy|(q-1)!}^2] \tag{5-83}$$

This is an important relation, since it allows one to estimate the output noise power spectral density. If this signal is large relative to the total output, then that implies there are additional inputs to the system which are not being accounted for.

The multiple coherence function $\gamma_{y:x}^2$ defined in Equation (5-69) is a direct measure of the presence of additional inputs, since it is the ratio of the output signal density due to all the

inputs considered and the total output signal density. Use of Equations (5-69) and (5-83) implies

$$\gamma_{y:x}^2 = \gamma_{y|q!}^2 = 1 - [(1-\gamma_{1y}^2)(1-\gamma_{2y|1}^2) \cdots (1-\gamma_{qy|(q-1)!}^2)] \quad (5\text{-}84)$$

If there is no noise present in the system which is not accounted for by any of the inputs, then the multiple coherence function is equal to unity.

5.3.1.1 Errors

The random error in the estimate of the multiple coherence function [5-2] is

$$\varepsilon_r[|\gamma_{y:x}^2|] = \frac{\sqrt{2[1 - \gamma_{y:x}^2]}}{|\gamma_{y:x}| \sqrt{M-q}} \quad (5\text{-}85)$$

The important feature this and other of the random error formulas is that the relative error is reduced by a factor $\sqrt{M-q}$ in the denominator rather than $\sqrt{M}$. The relative random error in the estimate of the magnitude of the equivalent frequency response function $L_{y|i}$ is

$$\varepsilon_r[|L_{y|i}|] = \frac{[1 - \gamma_{iy|(i-1)!}^2]^{1/2}}{|\gamma_{iy|(i-1)!}| \sqrt{2(M+1-i)}} \quad (5\text{-}86a)$$

whereas the error in radians in the phase estimate of the equivalent frequency response function is

$$\Delta\phi_{iy} = \varepsilon[|L_{y|i}|] \quad (5\text{-}86b)$$

This equation is very similar to Equation (5-42b) for the single

input/single output system. Equation (5-86) has the additional property that for a given number of averages M, the random error increases as i goes from 1 to q. This error can be made suitably small by increasing the number of statistically independent averages.

5.3.1.2 Example

The coherence method of analyzing multiple input systems was applied to the study of noise generation by a diesel engine [5-10]. The six inputs to the system were the pressures in each of the six cylinders of the engine. The output was considered to be the engine noise acoustic pressure measured at a point three feet from the engine surface. Figure 5-16 shows the multiple coherence function between the six inputs to the system and the single output over the frequency range 500-3500 Hz. It can be seen that the multiple coherence is large over the range 500-2500 Hz which implies that the output noise is produced by the six identified inputs of the system. The coherent output power $\gamma_{y:x}^2 G_{yy}$ is illustrated in Figure 5-17 along with the total output power G_{yy}. This figure shows that the coherent power, produced by sources coherent with at least one of the six inputs, is within 3 dB of the total power over most of the frequency range between 500-2500 Hz.

A second example involving partially coherent sources is described in Section 7.1.3.2.

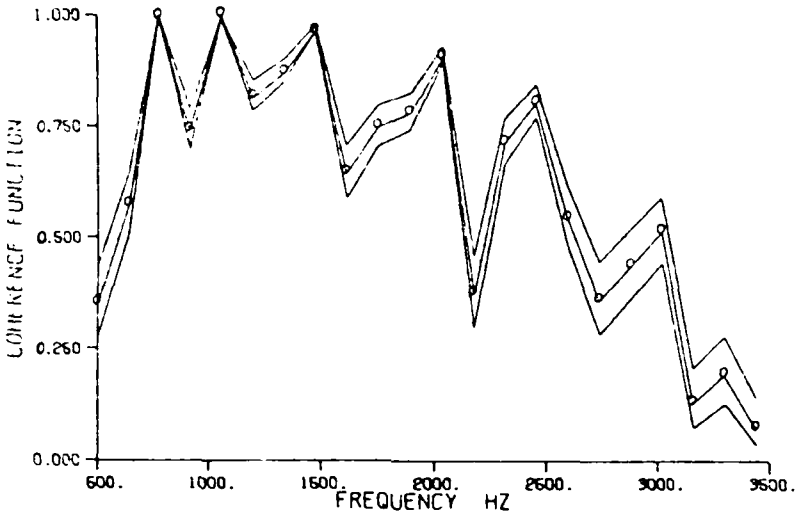


FIG. 5-16

MEASURED
MULTIPLE COHERENCE FUNCTION

Data from Ref. 5-10

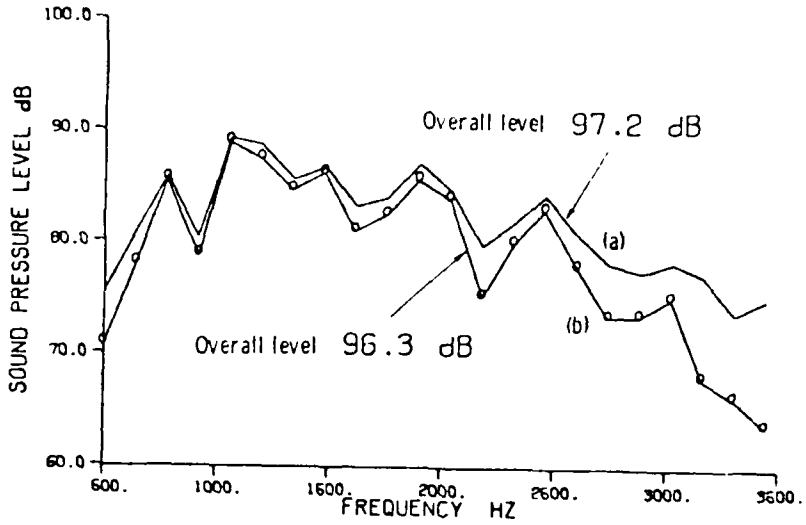


FIG. 5-17

MEASURED COHERENT OUTPUT POWER

a) TOTAL ENGINE NOISE

b) COHERENT ENGINE NOISE

Data from Ref. 5-10

5.4 ESTIMATION OF DAMPING

5.4.1 Measures of Damping

5.4.1.1 Introduction

Damping is an important parameter in the frequency response functions of dynamic system. Because it is difficult to determine damping values analytically, there is a strong dependence on experimental evaluations of the damping. Typically, one wants to determine the value of the damping as a function of frequency. In certain situations where nonlinear processes such as turbulence are important, the damping is measured also as a function of excitation level. The effect of damping on frequency response functions depend on whether a drive point or transfer frequency response function is being considered. In drive point response functions the resonant peaks and dips are reduced in amplitude and broadened in frequency by increased damping, but the average value of the response function does not change. However, with transfer frequency response functions the damping is often the controlling factor of the average response level.

Damping measurements on flight structures are complicated by the fact that damping values often change with the flight environment. Internal damping of some materials change with temperature and age. Air pumping at riveted joints depends on ambient pressure. Also, for lightweight surface shells the radiation damping of the vibration is a significant factor which changes with the conditions of the surface air flow. Standard damping measurements of mechanical structures using a measure of the decay of energy with time is usually only feasible in ground tests because a fairly low background noise level is required. These methods are most appropriate for measurements of the damping of interior elements of a flight structure and must be

used with caution for elements which are subject to temperature variations or fluid loading.

This section is concerned with the various measures of damping and their inter-relationships. The remainder of Section 5.4 is concerned with the various methods of experimentally estimating the damping.

5.4.1.2 Experimental Methods

In its most fundamental form, damping refers to the loss of energy of a system over time. The energy E(t) of a lightly damped oscillator decays exponentially with time with a relaxation time τ as follows:

$$E(t) = E_0 \, e^{-t/\tau} \tag{5-87}$$

The energy at time t = 0 is E_0, and the initial energy has decreased to 37% of its inital value after a time interval of $t = \tau$. A measurement of the relaxation time of this simple system completely characterizes the description of the damping.

In most realistic systems the damping varies with the frequency of excitation. It is convenient to describe the energy decay in terms of the loss factor η by

$$E(t) = E_0 \, e^{-\eta\omega t} \tag{5-88a}$$

where the angular frequency ω is related to the frequency f by

$$\omega = 2\pi f \tag{5-88b}$$

Then

$$\eta = (1/\omega t) \tag{5-88c}$$

One simple method of measuring the loss factor is to measure the time T_{60} it takes for the energy to decay to 10^{-6} of its initial

value; equivalently, the so-called reverberation time T_{60} is the time it takes for the energy level to decrease by 60 dB. The reverberation time is related to the loss factor by

$$\eta = \frac{\ln (10^6)}{\omega T_{60}} \approx \frac{2.2}{f T_{60}} \qquad (5\text{-}89)$$

If the energy decays exponentially with time then the energy level measured in decibels decays linearly with time (its slope is a straight line). More complex systems do not typically exhibit true exponential decay. It is often convenient to define a reverberation time based on the initial straight line slope of the energy level decay curve (see Figure 5-18).

Another common measure of damping in a simple oscillatory system is the half-power bandwidth $\Delta f_{1/2}$. As implied by its name, the half-power bandwidth is most simply defined by a frequency domain representation of the power loss in the system. Resonance in a simple mechanical oscillator occurs at a frequency f_0 when the mobility, defined as the ratio of the velocity of the oscillator to the driving force, is a maximum. If the force excitation is independent of frequency, then the power injected into the system is proportional to the real part of the mobility. Thus, the half-power bandwidth is the frequency range over which the real part of the mobility falls to half its value. This half-power bandwidth is related to the loss factor and resonance frequency of the oscillator by

$$\Delta f_{1/2} = \eta f_0 \qquad (5\text{-}90)$$

The half-power bandwidth is a useful measure of damping in systems with many degrees of freedom as long as this bandwidth is small compared to the mean frequency spacing between the resonances (see Section 5.4.4).

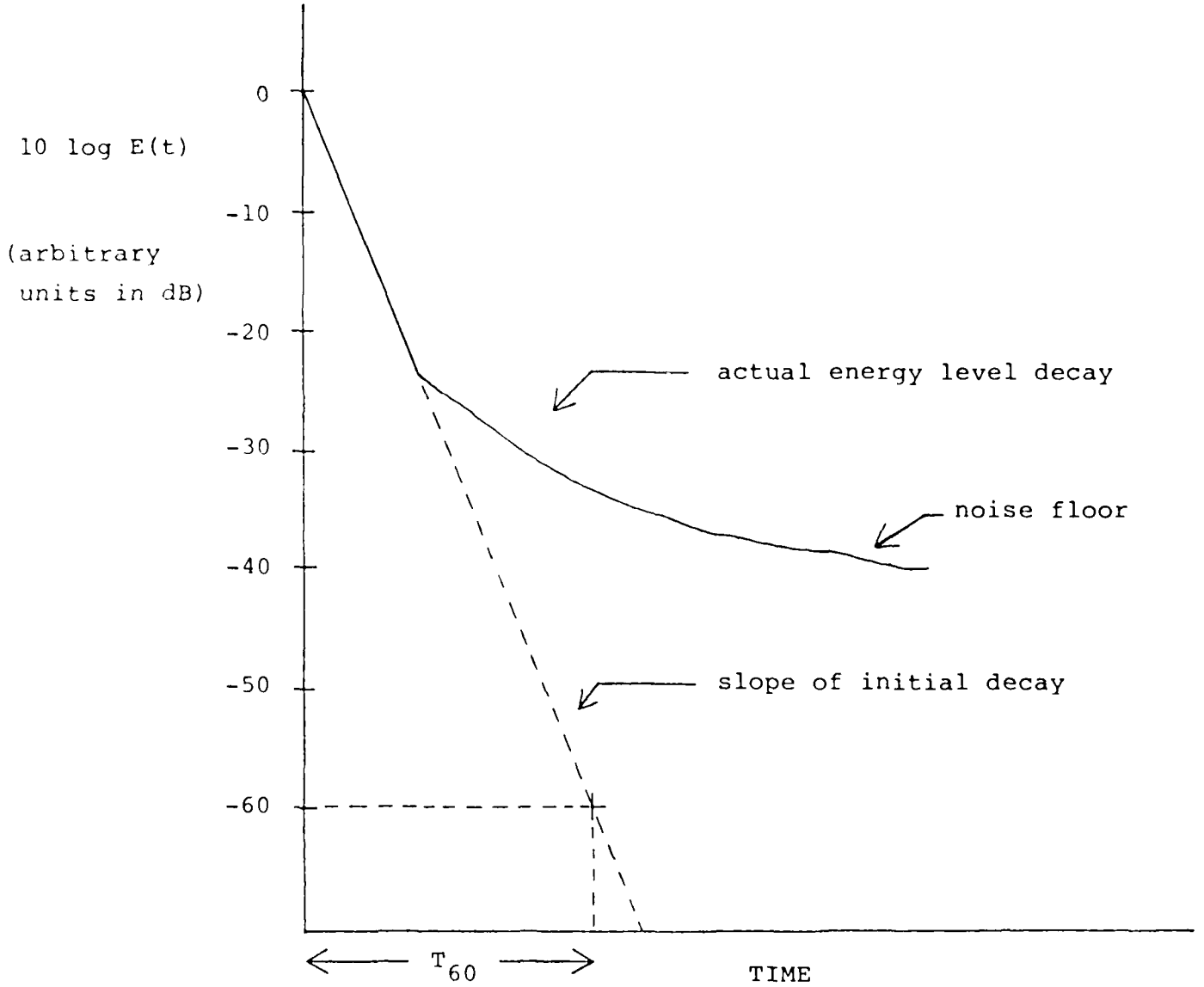


Figure 5-18 Energy Level Decay and Reverberation Time

Other measures of damping are sometimes used in frequency response function measurements on spatially extended systems. One such measure is the energy attenuation coefficient α, which is defined to be the increment in the energy loss over a small increment of distance. The relation of the attenuation coefficient to the loss factor of a system whose resonances do not significantly overlap is rather complicated and depends on the spatial separation of the two measurement locations and the impedances at these locations. In a system with significant modal overlap, the damping may be described in a statistical manner which naturally leads to a consideration of an average attenuation coefficient (see Chapter 8).

The equivalent viscous damping coefficient β and the damping ratio ζ are sometimes used to define the damping. These are defined in terms of the loss factor η by

$$\eta = 2\beta/\beta_c = 2\zeta \tag{5-91}$$

where the critical damping coefficient is the smallest value of β such that the oscillation is critically damped. For a simple harmonic oscillator of mass m and spring constant k described by the relation,

$$\ddot{x}(t) + 2\beta\dot{x}(t) + (k/m)\, x(t) = 0 \tag{5-92a}$$

the critical damping coefficient is

$$\beta_c = \sqrt{k/m} \tag{5-92b}$$

The superscript ($\cdot$) denotes a time derivative. For a bending wave displacement y on a beam of density ρ, bending stiffness B, cross-section area A, and bending wave number k_B described by the relation,

$$\ddot{y} + 2\beta \dot{y} + B k_B^4 / (\rho A)\, y = 0 \qquad (5\text{-}93a)$$

where the critical damping coefficient is

$$\beta_c = (B k_B^4 / \rho A)^{1/2} \qquad (5\text{-}93b)$$

5.4.1.3 Example

We discuss as an example of a damping measurement the equivalent viscous damping coefficients associated with each of the widely-spaced resonant modes of a lightly damped beam [5-11]. The beam was excited by a shaker mounted at one point and the acceleration was measured at another point. The damping information was determined using a computer via the Modal Analysis and Modeling Systems (MODAMS) software package. Given the measured transfer function, the computer program chooses a least-squares fit for the resonance frequency and viscous damping coefficient associated with each of the widely-spaced modes.

Figure 5-19 shows the measured transfer function magnitude along with the computed least squares fit associated with one of the resonant modes. The estimate of the resonant frequency is 265 Hz, and the estimate of the damping ratio ζ is 0.01396.

5.4.2 Energy Decay Method

The energy decay method is used to estimate the reverberation time of the system from which the loss factor may be inferred by Equation (5-89). Perhaps the simplest method is to use an impulse as the excitation. One measures the total RMS amplitude decay of the system as a function of time. The reverberation time is the time it takes for this response level to fall by 60 dB. One does not usually measure this time directly. Instead one measures the initial slope of the response

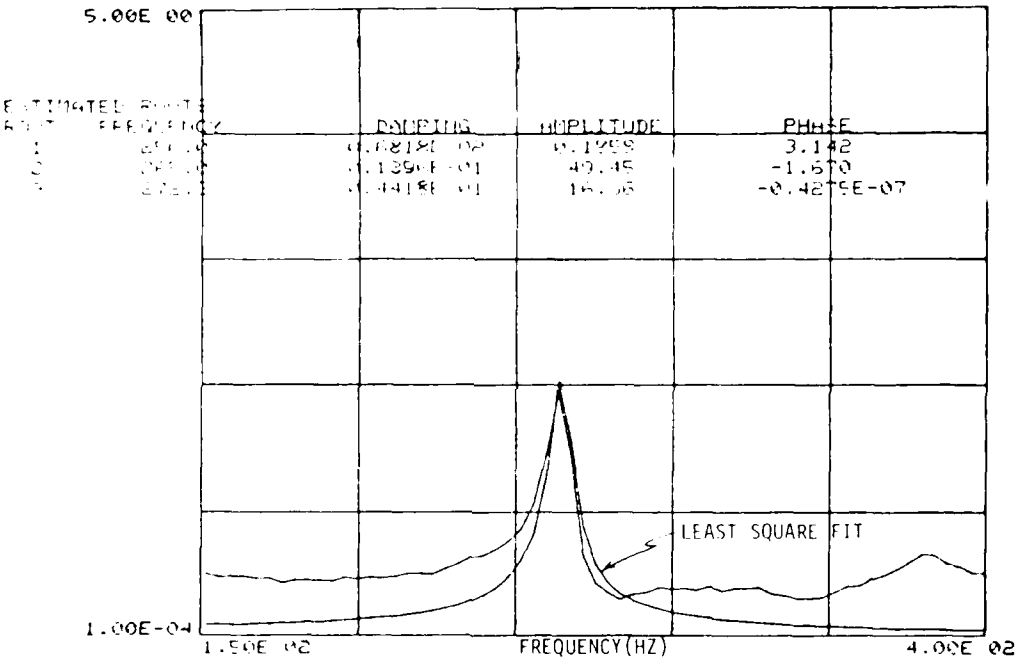


FIG. 5-19 MEASURED TRANSFER FUNCTION OF BEAM

Data from Ref. 5-11

level as a function of time and continues this slope downwards until the response is 60 dB down (see Figure 5-18).

In a measurement of this sort all the frequency dependent information on the damping is lost. One can determine the frequency dependence of the reverberation time by using narrow band (typically third-octave) filtering of the measured response. The output of the filter may be connected to a graphic level recorder or storage oscilloscope in order to determine the initial slope.

The method of generating the impulse is typically a hammer impact on a structure, the bursting of a balloon, firing of a gun, or a spark plug discharge for the case of an acoustic excitation. The limitations of any of these methods is that the frequency range over which data may be acquired is limited by the finite duration of the impulse. A finite duration impulse of duration Δt has an approximately uniform frequency spectrum for frequencies small compared to $(1/\Delta t)$. One may also excite the system with a broad band noise excitation and measure the amplitude fall-off immediately after the source is turned off. This type of input excitation has the advantage of better repeatability of the stimulus as well as the fact that the vibrational reverberant field is fully established before the excitation is removed. One can also use bandpass limited noise in order to investigate the frequency dependence of the reverberation time, thus eliminating the need for careful filtering of the output response. A practical drawback is that more instrumentation is required to carry out the measurement. The main source of computational error is associated with the determination of the initial slope of the decay.

A practical source of error associated with any impulse technique is that the short duration of the impulse degrades the signal to noise level. Increasing the duration degrades the frequency bandwidth of the measurement. The very wide dynamic range of impulse measurements forces the technician to be very

aware of the dynamic range of the transducers and other instrumentation.

A different source of error arises in measuring the reverberation time of any system with a high degree of modal overlap. The implicit assumption in a measurement of reverberation time is that an ensemble average is carried out over a large number of identical systems. As long as the system is ergodic, an ensemble average may be achieved by spatial averaging over the system. In the absence of this averaging, there can be large variability in estimates of the reverberation time. Very precise estimates of the reverberation time are possible if one statistically averages the decay curves [5-7].

In measurements of reverberation time on structures one must take into account the possible effect of the mass loading of the driver on the structure being excited. Whether one is using an impulse excitation or a steady state excitation which is switched off, the decaying excitation involves the motion of the structure coupled with the driver. One may assume that the resulting reverberation time of the structure plus the driver is the reverberation time of the structure alone as long as the mass of the driver is negligible.

5.4.3 Autocorrelation and Random Decrement Methods

The damping of a single degree of freedom oscillator may be measured using the autocorrelation function of the system response variable or using the random decrement method. By use of appropriate bandpass filtering, these methods may also be used for a system whose resonances are widely separated relative to their half-power bandwidth.

The autocorrelation function $R_{xx}(t)$, defined by Equation (5-17a) is proportional to the free vibration energy decay of a simple harmonic oscillator when the excitation is a stationary, Gaussian random process (see Chapter 3). Thus, the amplitude

envelope of the autocorrelation function is given by Equation (5-88a). Estimates of the autocorrelation function give a direct estimate of the loss factor.

The Randomdec signature $D_{x_o}(t)$ is generally defined to be the average of the time histories of the ensemble of samples whose initial vibratory amplitude and slope are identical. An alternative formulation assumes a less restrictive requirement that all members of the ensemble have the same amplitude (trigger level) X_0 at time t = 0, without restricting the slope. It is assumed in Randomdec analysis that the amplitude envelope of the Randomdec signature is proportional to the free decay of the system given by Equation (5-88a). When the linear system is excited by a stationary, Gaussian random process, the Randomdec signature, based upon the ensemble of time histories whose trigger level X_0 is identical, is in fact proportional [5-8] to the normalized autocorrelation function as follows:

$$D_{x_o}(t) = \frac{R_{xx}(t)}{R_{xx}(0)} X_0 \qquad (5\text{-}94)$$

Note that the Randomdec signature is also proportional to the trigger level.

Techniques for estimating the autocorrelation function are discussed in Sections 3.2.2 and 3.2.3.2. The statistical errors in estimating the autocorrelation function are discussed in Section 3.2.2.1. The variance of the estimate of the Randomdec signature is given by

$$V[D_{x_o}(t)] = \frac{R_{xx}(0)}{M} \{1 - [\frac{R_{xx}(t)}{R_{xx}(0)}]^2 \} \qquad (5\text{-}95)$$

under the condition that the excitation is stationary, Gaussian random process with zero mean. The variance decreases with the

number M of ensemble averages. This estimate of the variance is valid under the condition that individual samples of the Randomdec time histories are uncorrelated. This implies that successive samples are taken at time intervals exceeding the time for which the autocorrelation function $R_{xx}(t)$ is non-negligible. The variance increases when the samples are correlated. It follows that convergence of the estimate can be obtained with fewer, more widely spaced samples.

5.4.4 Spectral Bandwidth Method

Section 5.4.1 identified the half-power bandwidth of a single resonance peak as a measure of damping. This is strictly true for a single degree of freedom oscillator. It is also approximately true for a multi-degree of freedom system if the resonances are well separated in frequency. The average separation between resonant frequencies is called the mean modal frequency spacing δf. The resonances are considered to be well separated when

$$\Delta f_{1/2} << \delta f \qquad (5\text{-}96)$$

This condition is also called "low modal overlap." In the vicinity of the resonant frequency of any resonance of such a system, the mobility of a structure or the impedance of an acoustical system resembles the analogous quantities for a single degree of freedom system. Therefore, the half-power bandwidth of such a resonance can be used to estimate the loss factor associated with that particular mode of the system.

If the resonance peaks do overlap, then the apparent bandwidth of a resonance is broadened. When the peaks are spaced even more closely, then it is no longer possible to resolve two or more peaks. The response of the system becomes smoothed out.

The vibrational field within this type of system is called a reverberant field. The usual criterion that a field be reverberant is that

$$\Delta f_{1/2} > 3\ \delta f \qquad (5\text{-}97)$$

The autocorrelation method of estimating damping requires that the resonant peaks be widely spaced. Since the transform of the autocorrelation function is the autospectral density and since the half-power bandwidth involves the autospectral density of the absorbed power, then these two methods are necessarily closely interrelated.

A simple method of measuring the half-power bandwidth is to examine the power spectral density as a function of frequency with a narrow band analog filter whose bandwidth is much less than the half-power bandwidth of interest. The response level is noted at the resonant frequency, and the analyzer is slightly detuned off resonance both above and below the resonant frequency until the response level is 3 dB down. The 3 dB bandwidth of the response level is the half-power bandwidth of the resonance peak. There is a bias error associated with this measurement since the measurement of the response level at the resonance frequency underestimates the actual value for reasons identical to those given in conjunction with Equation (5-36). This error is reduced by reducing the analysis bandwidth.

One can also measure the absorbed power of the system using digital signal processing. For example, the mobility of a structure may be estimated as the ratio of the cross-spectral density of the force and velocity to the autospectral density of the force. The half-power bandwidth of the resulting frequency response function may be measured as long as the frequency bandwidth of adjacent bins is small relative to the bandwidth of the resonance. Methods for measuring this frequency response

function are discussed in Section 5.2 along with the associated error analysis.

5.4.5 Integrated Impulse Method

The integrated impulse method is used to measure the reverberation time in a single measurement, thus eliminating the need for averaging. It was first developed in an application to room acoustics [5-9], but the technique is generally applicable. In the reverberation decay method a system is excited by a random noise source. A particular source and receiver arrangement is chosen. The source is abruptly switched off at time t=0 and the slope of the decay curve of the output response gives the reverberation time of the system as discussed in Section 5.4.2. The ensemble average of squared output response decay curves y(t) is related to the integral witched off at time t=0 and the slope of the decay curve of the output response gives the reverberation time of the system as discussed in Section 5.4.2. The ensemble average of squared output response decay curves y(t) is related to the integral of the square of the impulse response function as follows:

$$y^2(t) = N \int_t^\infty h^2(\tau)\, d(\tau) \qquad (5\text{-}98)$$

The noise power per unit injected into the system in the reverbertion decay experiment is N (this is the average autospectral) density of the input in the filter bandwidth. The integral is from arbitrary time t to infinity where the impulse occurs at t=0. The impulse response function is the Fourier transform of the frequency response function relating the output response to the input signal and is discussed in Seciton 5.1.2. In the impulse response method the damping is evaluated based on

the right hand side of Equation (5-98) if the impulse response function is known. If the impulse response function is not known, it is most conveniently measured by using a finite duration impulse. Ideally, the measured response y(t) to an impulse is the impulse response function. The finite duration of the impulse limits the high frequency resolution of the measurement system in a manner discussed in Section 5.4.2. The measured impulse response is filtered (for example, in a third-octave band), analogous to the filtering of the random noise-source in the reverberation decay method. This filtered response is then rectified (corresponding to calculating $y^2(t)$) and integrated from time t out to later times. The upper limit of temporal integration is determined by the background noise level in the system. The process of rectification and integration may be carried out using a full-wave rectifier voltmeter, op amp integrating circuit, and strip chart recorder. Alternatively, one may record the signal and subsequently analyze it on a computer, or digitally sample and process the data in real time.

One feature of the integrated impulse method is that the resulting reverberation decay curve is applicable for a particular source and receiver spatial configuration. A drawback of this technique relative to the averaging of energy decay curves described in Section 5.4.2 is that the dynamic range is less in the integrated impulse method. This is because the energy decay method uses a continuous random signal which is abruptly switched off, whereas the integrated impulse method uses an impulse excitation. The advantage is that the integrated impulse technique is very simple in terms of instrumentation requirements, and the ensemble average is obtained with a single measurement.

REFERENCES

5-1. American National Standard Nomenclature and Symbols for Specifying the Mechanical Impedance of Structures, ANSI S2.6 (R 1976), Proposed Revision S2.21 - 1979.

5-2. Bendat, J.S. and Piersol, A.G., Engineering Applications of Correlation and Spectral Analysis, John Wiley and Sons, Inc., New York, NY, 1980.

5-3. Dodds, C.J. and Robson, J.D., "Partial Coherence in Multivariate Random Processes," J. Sound Vib., Vol. 42, No. 2, 1975.

5-4. A Review of Methods for Estimation of Aeroacoustic Loads on Flight Vehicle Surfaces, AFFDL-TR-76-91, Vol. III.

5-5. Jenkins, G.M. and Watts, D.G., Spectral Analysis and Its Applications, Holden-Day, Inc., San Francisco, CA, 1968.

5-6. Bendat, J.S. and Piersol, A.G., Random Data: Analysis and Measurement Procedures, John Wiley and Sons, Inc., New York, NY, 1965.

5-7. Chu, W.T., "Comparison of Reverberation Measurements Using Schroeder's Impulse Method and Decay Curve Averaging Method," J. Acoust. Soc. Am., Vol. 63, 1978, p. 1444-1450.

5-8. Vandiver, J., Dunwoody, A., Campbell, R., and Cook, M., "A Mathematical Basis for the Random Decrement Vibration Signature Analysis Technique," ASME Paper No. 81-DET-13.

5-9. Schroeder, M.R., "New Method of Measuring Reverberation Time," J. Acoust. Soc. Am., Vol. 37, 1965, p. 409-412.

5-10. Chung, J. Y., Crocker, M. J., and Hamilton, J. F., "Measurement of frequency responses and the multiple coherence function of the noise-generation system of a diesel engine," J. Acoust. Soc. Am., Vol. 58, 1975, p. 635-642.

5-11. Lee, J. H., Obal, M. W., Brown, D. L., "Prediction of the Angular Response Power Spectral Density of Aircraft Structures," AFFDL-TR-78-188, AF Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio, December 1978.

CHAPTER 6. IDENTIFICATION OF TRANSMISSION PATHS

By

D. H. KEEFE

CHAPTER 6. IDENTIFICATION OF TRANSMISSION PATHS

TABLE OF CONTENTS

		Page
	LIST OF FIGURES	6-iii
	LIST OF SYMBOLS	6-iv
6.1	TRANSMISSION PATH MODEL	6-1
6.2	NONDISPERSIVE WAVE PROPAGATION	6-2
	6.2.1 Definitions	6-2
	6.2.2 Time Delay and Correlation Analysis	6-4
	6.2.2.1 Time Delay Procedure	6-4
	6.2.2.2 Cross-Correlation Procedure	6-5
	6.2.3 Impulse Response Analysis	6-9
	6.2.4 Cepstral Analysis	6-10
	6.2.4.1 Introduction	6-10
	6.2.4.2 Procedure	6-11
	6.2.4.3 Example	6-14
	6.2.5 Time Delay Measurements in the Frequency Domain	6-17
	6.2.5.1 Cross-Spectral Density Technique	6-17
	6.2.5.2 Complex Coherence Technique	6-19
	6.2.5.3 Smoothed Coherence Transform Technique	6-20

6.3 DISPERSIVE WAVE PROPAGATION..... 6-23

6.3.1 Introduction..... 6-23

6.3.2 Cross-Correlation Analysis..... 6-23

6.3.2.1 Introduction..... 6-23
6.3.2.2 Procedures..... 6-24
6.3.2.3 Example..... 6-28

6.3.3 Frequency Domain Methods..... 6-31

6.3.3.1 Introduction..... 6-31
6.3.3.2 Cross-Spectral Density Technique.. 6-31
6.3.3.3 Recompression Technique Using Spatial Impulse Response Function. 6-33

REFERENCES..... 6-40

LIST OF FIGURES

Figure		Page
6-1.	Cross-Correlation Function for Multiple Path System.....	6-7
6-2.	Cross-Correlation Function for Dispersive System.....	6-7
6-3.	Measured Cepstrum of Jet Engine Noise.....	6-15
6-4.	Use of Measured Cepstrum in Dereverberation.....	6-18
6-5.	Measured Smoothed Coherence Transform and Cross-Correlation Function Comparison.....	6-22
6-6.	Measured Cross-Correlation for Ship Hull Transmission.....	6-29
6-7.	Measured Phase of Cross-Spectral Density for Diesel Engine.....	6-34
6-8.	Measured Standard Impulse Response Function of Beam.....	6-38
6-9.	Measured Spatial Impulse Response Function of Beam.....	6-39

LIST OF SYMBOLS

C(f)	Smoothed coherence transform (SCOT)
C(z)	Normalized correlation coefficient
Cg	Group velocity
Ci(z), Si(z)	Fresnel cosine and sine functions
Cy(q)	Power Cepstrum
c	Phase speed
d	Distance
f	Frequency, cycles/sec
G(f)	Spectral density function
h	Plate thickness
h(t)	Impulse response function
j	-1 complex number
k	Wavenumber
k(q)	Complex cepstrum
m(t)	Noise at input
n(t)	Noise at output
q	Time delay, sometimes referred to as quefrency
R( )	Correlation function
s(t)	Signal time history
t	Time
V	Volume
W(f)	Weighting function
x(t), X(t)	Generalized input variable and its transform
y(t), Y(t)	Generalized output variable and its transform
z	Dummy variable
γ	Complex coherence function
Φ	Phase
ω	Radian frequency, radians/sec
τ	Delay time

CHAPTER 6. IDENTIFICATION OF TRANSMISSION PATHS

This chapter presents methods for applying measured data to the task of transmission path identification. Following preliminary definitions in Section 6.1, nondispersive wave transmission is described in Section 6.2 and dispersive wave transmission is described in Section 6.3. Transmission path identification methods may be grouped into two classes; (i) time domain methods such as time delay measurement, cross-correlation analysis, and impulse response analysis, and (ii) frequency domain methods such as the cross-spectral density method and the smoothed coherence transform technique. This chapter of the compendium describes the use of such techniques.

6.1 TRANSMISSION PATH MODEL

The identification of transmission paths in dynamic systems is useful both for the identification and diagnosis of the sources of excitation of the system. It is also important in order to understand why the system responds the way it does and how to solve the problems that exist. After the transmission paths have been identified, they can be modelled (see Chapter 8) in some manner to predict changes in the system response to changes in the system inputs and system design (involving the system frequency response functions).

The simplest type of transmission path is that corresponding to propagation of a signal. Examples include the propagation of an acoustic wave in free space, and wave excitation in a very large homogeneous structure produced by a pulse excitation over sufficiently brief durations that the excitation has not yet arrived at any boundary of the structure. One usually regards a propagating wave as one in which there is negligible reflection of energy at any point in the medium.

A propagating wave is said to be attenuated if the energy of the wave is dissipated as heat. The energy of an acoustic wave is attenuated by viscous and thermal losses in air, while attenuation of propagating waves on plates is primarily due to the internal friction of the material comprising the plate. The term attenuation is sometimes used to refer to any loss of energy by a system. This may include radiated energy, for energy which is transmitted to a different system.

A discontinuity in the impedance of a structure or any system which can support wave motion produces a reflected wave. This reflected wave can be re-reflected at a system boundary or other impedance discontinuity, so that the vibrational field becomes more complex than the field of a propagating wave. One of the greatest difficulties in constructing transmission path models is to separate out the direct field (due to the propagating wave) of a source excitation from the multiple reflections which occur. Various techniques discussed in this chapter are used to effect this decomposition, and thereby isolate the various transmission paths.

The presence of multiple reflections in a system with low attenuation leads to the creation of a reverberant field. A reverberant field is one in which sufficiently many modes of the system are excited that the resonances associated with these modes overlap in the frequency domain. Examples include a sound field within an enclosure or the excitation of a structure at sufficiently high frequencies.

6.2 NONDISPERSIVE WAVE PROPAGATION

6.2.1 Definitions

A nondispersive wave is one in which the phase velocity of the propagating wave is independent of frequency. In practice,

there are many examples of wave motion for which this condition is approximately satisfied. Sound in air at acoustic frequencies propagates in a nearly nondispersive manner, as long as the small frequency-dependent attenuation is ignored. Similarly, longitudinal and shear waves in solids tend to be nondispersive. One of the most familiar examples of dispersive wave motion is that of transverse bending waves on a plate, whose phase velocity increases with frequency. Dispersive wave motion is considered in Section 6.3.

A system can be considered to be nondispersive if the frequency bandwidth of interest is sufficiently narrow that the phase velocity remains nearly constant. For example, electromagnetic radiation has a slowly varying phase velocity mainly due to variations with frequency in the dielectric constant of the medium in which the wave propagates. However, over a sufficiently narrow bandwidth the variations in the phase velocity with frequency may be neglected.

An important characteristic of nondispersive waves is that the shape of the temporal waveform, moving at the phase velocity, does not change with time or space. In other words, there is no wave-form distortion. Thus, the origin of the name nondispersive wave signifies that an initial temporal pulse does not spread out or disperse with increasing time. This is because all the spectral components of the pulse propagate at the same speed. Conceptually, a disturbance in a nondispersive system can be followed around in time to identify which paths it takes to go from the source of the disturbance to the response point of interest without the problem of having the disturbance change its shape except at boundaries.

6.2.2 Time Delay and Correlation Analysis

6.2.2.1 Time Delay Procedure

The most direct method of identifying transmission paths of a non-dispersive system is by exciting the system with impulses (using either an externally applied signal or internal impulsive sources within the system). One monitors the response at a point and detects the arrival time of the impulse. Since the wavespeed is constant, then a measurement of the travel time leads to an estimate of the propagation distance between the points if the wavespeed is known, or an estimate of the wavespeed if the spatial path separation between the points is known. Examination of later impulses received give information on the first few reflections within the system. Thus, various paths can be identified and their relative amplitudes measured.

The measurement procedure for time delay is as follows. The time delay between the signals may be measured by a timer clock. The clock is turned on by the first pulse and turned off by the second pulse, and the time delay between the two signals is displayed. The main consideration in reducing the error is to set the threshold levels for the on and off-gating circuits to a sufficiently high level that no spurious signals affect the timing.

If the source is externally driving the system, then gating techniques can be used to eliminate the reflected waves from the system, or alternatively gating can be used to investigate the pulse after a single reflection. A single tone burst, which is a rectangularly-gated sinusoid, is the excitation signal whose temporal duration is shorter than the time delay between successive reflections of the signal from the system boundaries. The signal at the receiving transducer is a series of pulsed sinusoids. The first pulse is the direct propagation of the signal from source to receiving point, and the later pulses

are the first, second, etc. reflected signals. The various time delays are easily measured by gated timing circuits, and hence the propagation path distance for the reflected as well as the direct propagation path may be obtained.

Problems arise with all time delay methods when the arrival times for successive propagation paths become closely spaced relative to the pulse or tone burst duration. For a three-dimensional, non-dispersive system of volume V and phase speed c, the mean time Δt between successive reflections is

$$\Delta t = (V/4\pi c^3 t^2) \qquad (6\text{-}1)$$

This mean time decreases as the total time t increases, so that at some future time the mean temporal spacing between pulses always becomes less than the pulse duration. Under this condition, it is not possible to separate out the various transmission paths without the use of signal processing techniques such as cepstral analysis (see Section 6.2.3).

6.2.2.2 Cross-Correlation Procedure

A second method of temporal discrimination involves the use of correlation analysis. The cross-correlation function is computed using methods discussed in Section 3.3.2 between the source and the response point if the source location is known, or between the two response points. If the source excitation is white noise or band-limited noise, then the cross-correlation function is a maximum at a delay time which is equal to the time delay for propagation between the two response points. The most important difference between the time delay measurement method and the correlation analysis method is that the time delay technique is useful for pulse excitation, whereas the correlation technique is useful for steady-state, random excitation. If the

source excitation is broad-band and produced within the system, then correlation analysis is often the more appropriate technique.

Suppose the excitation is band-limited white noise between the lower frequency f_L and upper frequency f_U. Then the cross-correlation function $R_{xy}(\tau)$ between the responses $x(t)$ and $y(t+\tau)$ whose path length difference is d is

$$R_{xy}(\tau) = XY \cos[\pi(\tau-d/c)(f_L+f_U)] \frac{\sin[\pi(\tau-d/c)(f_U-f_L)]}{\pi(\tau-d/c)(f_U-f_L)} \qquad (6\text{-}2)$$

where X and Y are the signal amplitudes. The cross-correlation function is a maximum at $\tau=d/c$, the propagation time between the two response points, and the envelope of the peak is more sharply defined as the noise bandwidth (f_U-f_L) increases. Therefore, determination of the peaks in the spatial cross-correlation function enable a measurement of the time delay between the signals.

Now suppose we have a system with N paths such that the path length d_i $(i=1,2,\ldots,N)$ is different for each path. The resulting cross-correlation function between the input and output has N peaks as shown in Figure 6-1. The peaks in the cross-correlation function have exactly the same form as in Equation 6-2, so that the delay time for each path corresponds to the time at which each peak of the cross-correlation function is located. The amplitudes of each peak depend upon the relative magnitudes of the transmission coefficients of the various paths.

Figure 6-1 illustrates that various transmission paths can be identified only if the delay times between any pair of paths are sufficiently different that the peaks can be distinguished. Let τ_1 and τ_2 be the delay times for two of the transmission paths between the two response points. The criterion [6-1] that

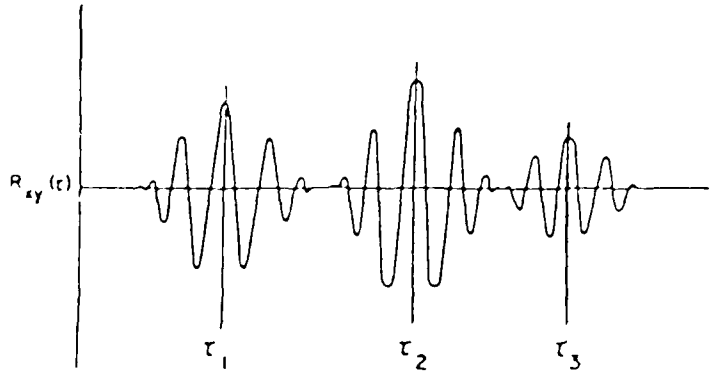


Figure 6-1. Cross Correlation Function for Multiple Path System

Data from Ref. 6-1

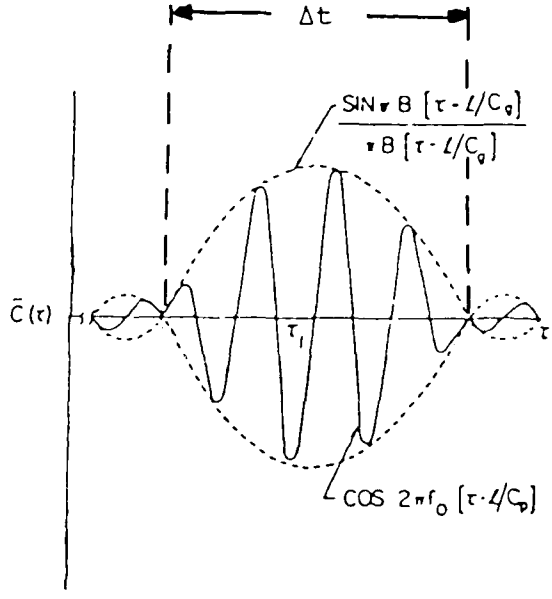


Figure 6-2. Cross Correlation Function for Dispersive System

Data from Ref. 6-1

these paths may be distinguished using correlation analysis is that the difference in these delay times satisfy the following inequality:

$$|\tau_2 - \tau_1| \ (f_U - f_L) \ > \ 1 \qquad (6\text{-}3)$$

The normalized correlation coefficient $C_{xy}(\tau)$ is defined to be,

$$C_{xy}(\tau) = R_{xy}(\tau)/[R_{xx}(0) \ R_{yy}(0)]^{1/2} \qquad (6\text{-}4)$$

where R_{xy} is defined in Equation (6-2) and discussed in Section 3.3.2, and R_{xx} and R_{yy} are the input and output autocorrelation functions (proportional to the energies of each of the signals) discussed in Section 3.2.2. The correlation coefficient for the time delay between transmission paths is given by Equation (6-2) with the signal magnitudes X and Y set equal to unity. Thus, the correlation coefficient evaluated at the delay time is equal to,

$$C_{xy}(\tau=d/c) = 1 \qquad (6\text{-}5)$$

It is important to note that the amplitude of the correlation coefficient is independent of the transmission path characteristics for nondispersive systems. This is not the case for dispersive systems. Figure 6-2 shows the correlation coefficient for a dispersive system. The time width Δt between the adjacent zeros of the correlation coefficient on either side of the central peak is illustrated on the figure, and this time width is equal to,

$$\Delta t = 1/(f_U - f_L) = 1/\Delta f \qquad (6\text{-}6)$$

where Δf is the frequency bandwidth of the excitation. The correlation coefficient peak is most clearly resolved by increasing the frequency bandwidth of the excitation.

6.2.3 Impulse Response Analysis

The impulse response function of a nondispersive system may be used to evaluate the delay time of propagation. The method is similar to the cross-correlation method described in Section 6.2.2. If the excitation has energy localized in narrow frequency bands then the impulse response function may be a better indicator of delay time than the cross-correlation function [6-2]. Let the two measured time histories be x(t) and y(t). What is desired is the computation of the single or multiple path delay times. The impulse response function h(t) (see Equations 5-11 and 5-25a of Chapter 5) of output y relative to x may be computed in terms of the cross-spectral density $G_{xy}(f)$ and autospectral density $G_{xx}(f)$ by

$$h(t) = \int_{-\infty}^{\infty} df\ e^{j2\pi ft}\ G_{xy}(f)/G_{xx}(f) \qquad (6\text{-}7)$$

If the spectral densities are computed using the FFT, then the integral is converted into a sum over all frequencies in the analysis bandwidth. Compare the impulse response function to the cross-correlation function expressed as follows:

$$R_{xy}(t) = \int_{-\infty}^{\infty} df\ e^{j2\pi ft}\ G_{xy}(f) \qquad (6\text{-}8)$$

If x(t) is the time history of a white noise input signal, then G_{xx} is constant and the impulse response and cross-correlation functions are equal to within a constant. The impulse response and cross-correlation functions are interchangeable in this case as regards delay time estimation. In effect, the autospectral density function $G_{xx}(f)$ in the denominator of Equation (6-7) is a

frequency domain-whitening function which can potentially resolve multiple path delay times. A related frequency domain weighting function is the smoothed coherence transform discussed in Section 6.2.5.

6.2.4 Cepstral Analysis

6.2.4.1 Introduction

This section describes a nonlinear filtering technique known as cepstral analysis. Due to the advent of small computers with extensive computational power, practical applications of cepstral analysis for filtering and signal recovery in the laboratory have become possible. Cepstral techniques have some advantages over the more traditional linear signal processing techniques in the area of transmission path analysis, as well as in the areas of de-reverberation or spectral smoothing, the analysis of periodic signals, and source identification.

Cepstral analysis may be used to detect and/or remove echoes from a nondispersive signal, and can thus be employed to compute the free-field response from the measured reverberant field response. Of particular interest are the capabilities of cepstral analysis for periodic signals. For example, the periodic components of the spectrum produced by harmonics of a particular gearbox rotation rate is one such periodic signal. All the harmonics corresponding to a given rotation rate are mapped by the cepstrum into a single peak. Filtering in the cepstral domain (an operation referred to as liftering) thus isolates the transmission characteristics of the various gearbox sources. The cepstrum may under certain conditions be used to separate source and transmission path effects. This is because the cepstrum converts the convolution of source/path characteristics in the time domain into simple addition in the cepstral domain.

The discussion in this section is limited to nondispersive systems. Future cepstral applications to systems exhibiting both dispersion and reverberation, such as the structural vibration of an aircraft frame, depend upon the development of more powerful techniques than exist at present. If the dispersive characteristics of the structure can be adequately modelled, then a combination of cepstral techniques with inverse filtering (used to remove dispersive effects) can be employed to study structural transmission paths. An indication of such an approach to dispersive structural path transmission is given in Section 6.3.3; however, cepstral techniques are not therein applied.

6.2.4.2 Procedure

As discussed in Section 3.3.4, the complex cepstrum $k_y(q)$ of a signal $y(t)$, whose Fourier transform is $Y(f)$, is the inverse Fourier transform of the logarithm of $Y(f)$. Since the transform $Y(f)$ is a complex quantity (see Section 3.2.3.1), then the appropriate logarithm is the complex logarithm. The transform $Y(f)$ is written in terms of its magnitude and phase as follows:

$$Y(f) = |Y(f)|\, e^{j\phi} \qquad (6\text{-}9a)$$

and the complex logarithm of $Y(f)$ is

$$\log Y(f) = \log |Y(f)| + j\phi \qquad (6\text{-}9b)$$

The real part of the above involves the logarithm of a real number, and this is well-defined. The imaginary part is not uniquely defined since addition of 2π radians to the phase of $Y(f)$ does not alter the complex value of $Y(f)$. Thus, it is necessary to restrict the range of the phase to 2π. Under this

condition the complex cepstrum $k_y(q)$ is defined to be the inverse Fourier transform (see Equation 5-8) of the logarithm of Y(f) as follows:

$$k_y(q) = \int_{-\infty}^{\infty} \log\ Y(f)\ e^{j2\pi fq}\ df \qquad (6\text{-}9c)$$

The parameter q plays a role analogous to that of time in the conventional Fourier transform. During the computation of the cepstrum it is necessary to unwrap the phase; i.e., the phase must be a continuous function of frequency. Information on the computation of the complex cepstrum is found in the sources referenced in Section 3.3.4.

The power cepstrum $c_y(q)$ is defined to be the inverse Fourier transform of the logarithm of the autospectral density $G_{yy}(f)$ of the signal Y(f) as follows:

$$c_y(q) = \int_{-\infty}^{\infty} \log[G_{yy}(f)]\ e^{j2\pi fq}\ df \qquad (6\text{-}10)$$

The autospectral density (or power spectral density) function is discussed in Section 3.2.3 and its application to transfer function measurements in Section 5.1. Since the autospectral density is a real, non-negative function, then the logarithm in Equation (6-10) is the real logarithm.

It is important to note that both the power and complex cepstrum are real quantities. This is obvious in the case of the power cepstrum, but the fact that the complex cepstrum is real depends upon the symmetry properties of Equation (6-9) (see reference [6-3]). The complex cepstrum is more general since the original time domain signal may be completely reconstructed, whereas this is not the case with the power cepstrum.

Cepstral analysis has usefully been applied to signal deconvolution. Consider the single input, single output linear

system illustrated in Figure 5-1. The input and output time domain signals x(t) and y(t), respectively, are related by the following convolution:

$$y(t) = h(t) * x(t) \qquad (6\text{-}11a)$$

$$y(t) = \int_{-\infty}^{\infty} h(\tau)\, x(t-\tau)\, d\tau \qquad (6\text{-}11b)$$

Equation (6-11a) is a short-hand notation for Equation (6-11b). The impulse response function h(t) of the system is discussed in Section 5.1.3.

Convolution in the time domain becomes addition in the cepstral domain. It is this property which makes cepstral analysis useful. One may filter in the cepstral domain to focus attention on either the source (x(t)) or the transmission (h(t)) characteristics. A periodic source excitation produces a cepstrum with a peak at a single value of q which is directly related to the repetition rate of the time domain signal.

One may thus use the same filtering techniques in the cepstral domain that are used in the frequency domain as described in Chapter 3. For example, assume that the input x(t) to the single input, single output system is a repetitive pulse. Assume also that the impulse response function of the system is slowly varying relative to the repetition rate of the pulse. The convolution of this pulse with the impulse response function produces a cepstrum whose high q values are predominantly due to the pulse, and whose low q values are due to the transfer function. Application of a low-pass lifter (i.e., a low pass filtering operation in the cepstral domain) isolates the transfer function contribution.

The time domain signal x'(t) after liftering may be recovered by use of the inverse cepstrum. Computation is strongly affected by phase considerations and the details of the filtering in the cepstral domain.

6.2.4.3 Example

Cepstral analysis can be used to isolate transmission paths by isolating the direct signal from the first few reflections, or from the reverberant field as a whole. The main property of the cepstrum which allows this is that a periodic time signal produces a cepstrum with a single peak. This method has been applied [6-4] to open air tests on the radiated sound from a jet engine. Since the investigators were interested in estimates of radiated power rather than reconstruction of the temporal waveform, they used the power cepstrum.

Figure 6-3a illustrates the radiated power spectrum, and the ripples in this spectrum are due to ground reflection. The power cepstrum corresponding to this signal is in Figure 6-3b, and the ground reflection is represented by the very strong peak in the cepstrum. This peak has been filtered out of the cepstrum in Figure 6-3c, and the resulting inverse cepstrum leads to the modified power spectrum shown in Figure 6-3d. This power spectrum is compared to the power spectrum measured under anechoic conditions (i.e., when measured in a special test facility which virtually eliminates all reflections in the bandwidth of interest), and the agreement is excellent. Syed et al. [6-4] remark that the cepstral technique is applicable even in the presence of wind, turbulence, or temperature gradients near the ground. They found it useful in both narrow-band and third octave averaging.

Cepstral analysis was also used in power measurements of an unheated jet used in conjunction with a spinning rig, an apparatus used to study in-flight effects on jet mixing noise [6-5]. The angular distribution of emitted jet noise varies

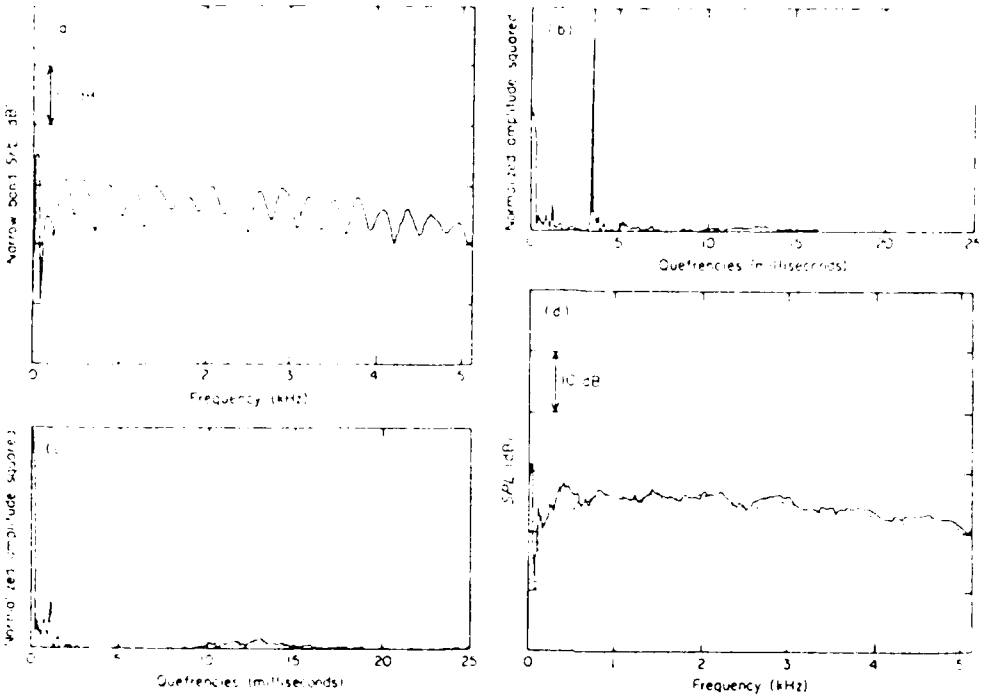


FIG. 6-3 MEASURED CEPSTRUM OF JET ENGINE NOISE

a) Autospectrum level with reflective ripple.

b) Power Cepstrum of (a) showing peak due to reflection.

c) Power Cepstrum with reflective peak liftered out.

d) Autospectrum level obtained via liftered cepstrum compared to autospectrum measured under anechoic conditions.

Data from Ref. 6-4

whether the test is conducted in-flight or by ground test. The spinning rig is used to create motion of the jet relative to the ground which reproduces some of the observed flight test power spectral data.

Figure 6-4a shows the power cepstrum for a microphone mounted 3 meters above the ground when the cold jet is mounted on the spinning rig. Since the propagation path is through air, then we are dealing with nondispersive propagation. The peak at B on the figure is due to a ground reflection and the peak at A is due to reflection from the foundation of the spinning rig. The power spectrum is illustrated in Figure 6.4b. The solid curve shows the ripples in the spectrum due to reflections, and the dotted curve shows the spectrum obtained from the inverse cepstrum after peaks A and B have been filtered out of the power cepstrum. The power spectrum thus obtained is free of ripple. The solid line of Figure 6-4c shows the power spectrum obtained from the cepstrum with peak A filtered out, and the dotted line shows the spectrum with Peak B filtered out. Both curves in Figure 6-4c still show ripple.

The last two parts of the figure contrast the elimination of reflections by cepstral processing with the direct elimination of reflection by using acoustic absorbers. The solid curve in Figure 6-4d shows the measured power spectra when the ground was acoustically treated to eliminate reflections. There is ripple in the curve due to reflections from the foundation, corresponding to peak A in Figure 6-4a. The dotted curve shows the power spectrum whose associated cepstrum has peak B, corresponding to the ground reflection, filtered out. The agreement is excellent. Figure 6-4e shows the constrasting power spectra when the foundation of the rig was covered with acoustic absorbers as compared to the spectrum obtained when peak A is filtered out of the cepstrum. The agreement is not as good as for Figure 6-4b since it is more difficult to acoustically treat a foundation

than the ground. This figure shows the utility of cepstral analysis, and suggests that ground test data may be acquired by a portable measuring test facility for which there is no need for acoustic treatment.

Inspection of the power cepstrum in Figure 6-4a shows that all of the useful information relating to the power spectrum is contained in the low-q region of the cepstrum. The signal degradation due to reflections is limited to high-q values. If the path length difference between the direct path and the reflected path were to decrease, then the high-q and low-q information would begin to overlap in the cepstrum and cepstral analysis would not be able to separate out the various transmission paths. In this particular case, a sufficiently long transmission path length difference is obtained by mounting the rotating jet much higher off the ground than the main body of the rig foundation.

6.2.5 Time Delay Measurements in the Frequency Domain

6.2.5.1 Cross-Spectral Density Technique

The time delay between various transmission paths may also be measured in the frequency domain. A common method is to use the cross-spectral density function. Consider a signal s impinging on each of two response points with excitations x(t) and y(t) such that a noise signal is present at each response point as follows:

$$x(t) = s(t) + m(t) \qquad (6\text{-}12a)$$

$$y(t) = s(t+\tau) + n(t) \qquad (6\text{-}12b)$$

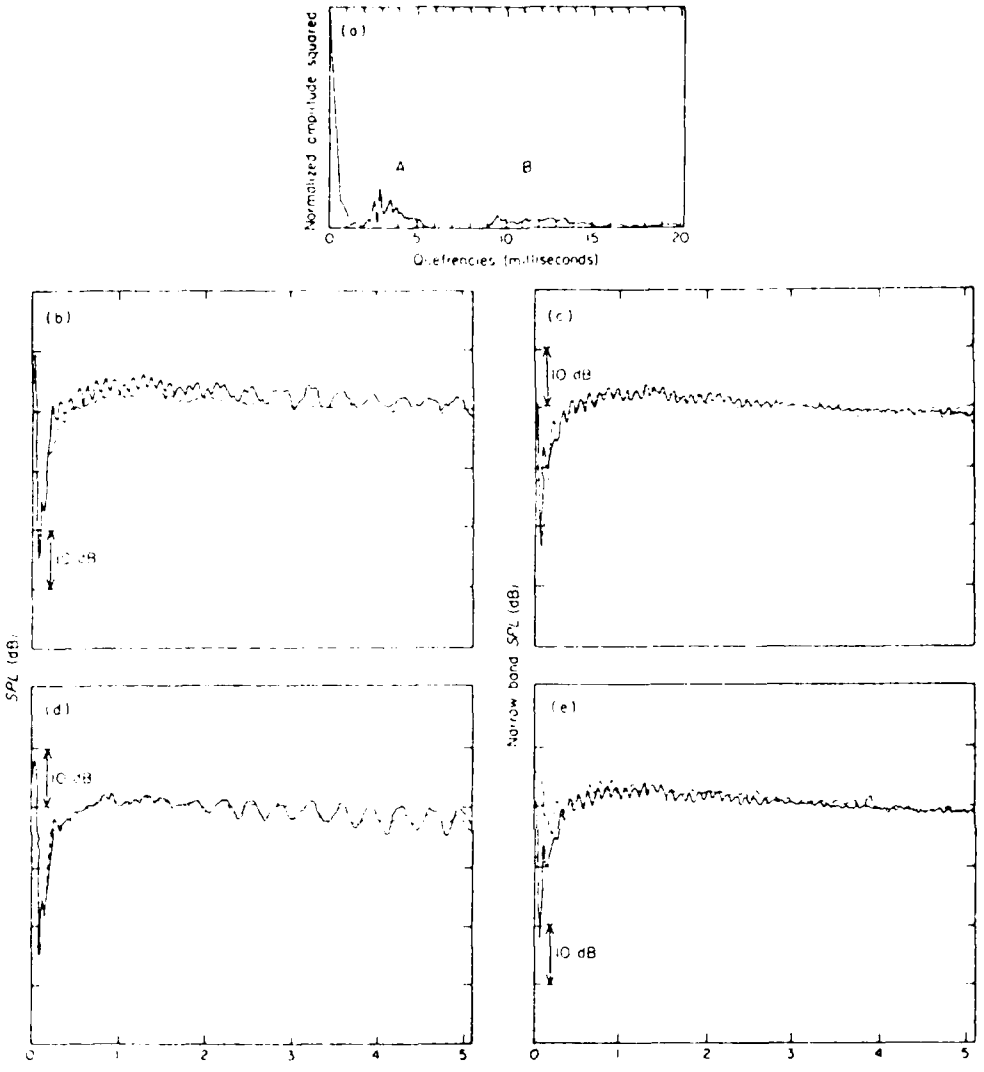


FIG. 6-4 USE OF MEASURED CEPSTRUM IN DEVERBERATION

a) Cepstrum from rig with reflected signals at A and B untreated for reflections from rig and ground.

b) Autospectra with reflections from A and B (solid line), and with reflections liftered out (dotted line).

c) Autospectra with A removed (solid line) and B removed (dotted line).

d) Autospectra with acoustically treated ground to remove reflections (solid line), and with untreated ground with B removed (dotted line).

e) Autospectra with acoustically treated rig to remove reflections (solid line) and with untreated rig with A removed (dotted line).

Data from Ref. 6-4

There is a time delay τ of the signal at y relative to x. The noise signals m(t) and n(t) are assumed to be uncorrelated with s(t) and with respect to each other. The cross-spectral density $G_{xy}(f)$ between x and y is

$$G_{xy}(f) = G_{ss}(f)\ e^{j2\pi f\tau} \tag{6-13}$$

The autospectral (power) density of signal s is G_{ss}, and it is a real non-negative quantity. Therefore, the phase Φ_{xy} of G_{xy} is,

$$\Phi_{xy} = 2\pi f\tau \tag{6-14}$$

The delay time between the transmission paths is simply obtained as the average slope of the graph of phase versus angular frequency.

6.2.5.2 Complex Coherence Technique

It is sometimes possible to improve the resolution of the delay time measurement by use of the coherence function. The complex coherence function γ_{xy} between the response at x and y is defined to be

$$\gamma_{xy}(f) = \frac{G_{xy}}{\sqrt{G_{xx}\ G_{yy}}} \tag{6-15}$$

The squared magnitude of the complex coherence function is the coherence function discussed in Chapters 3 and 5. The complex coherence function for the system defined by Equation (6-12) is,

$$\gamma_{xy}\ (f) = \frac{G_{ss}\ e^{j2\pi f\tau}}{\sqrt{(G_{ss} + G_{mm})(G_{ss} + G_{nn})}} \tag{6-16}$$

where G_{mm} and G_{nn} are the noise autospectral densities. The phase of the complex coherence function is precisely equal to the

phase defined in Equation 6-14), so that the complex coherence function or the cross-spectral density function may be used to estimate the path length difference.

The main distinction between the two frequency domain methods occurs when the noise sources are partially correlated with the signals. The signal s in practice may not be a white noise signal. It may for example be a signal with noise components as well as a few strong sinusoidal components. Such a signal will not be statistically independent relative to a noise signal. Thus the phase of such a cross-spectral density not proportional to the delay time, but includes phase error terms due to the correlation of the signal with the "noise sources" -- noise in any practical measurement situation meaning any background signal which is not of interest. The cross-spectral density function and the complex coherence function differ only in their magnitude response. The complex coherence function is the ratio of the cross-spectral density to the square root of the product of the autospectral densities of the two signals. This division by the autospectral densities has the effect of whitening the spectrum since the frequencies at which the autospectral densiites are small are boosted in the complex coherence function relative to the cross-spectral density. Conversely, Equation (6-16) shows that if the noise at either measurement point is large relative to the signal autospectral density G_{ss}, then the complex coherence function is reduced in magnitude. It is possible to weight the complex coherence function by appropriate low, high, or band-pass filtering to eliminate points where the coherence is low.

6.2.5.3 Smoothed Coherence Transform Technique

The inverse Fourier transform of the complex coherence function γ_{xy} weighted by a smooth weighting function W(f) is the smoothed coherence transform (SCOT) defined as follows:

$$C_{xy}(\tau) = \int_{-\infty}^{\infty} W(f)\ \gamma_{xy}(f)\ e^{-2\pi j f \tau} df \qquad (6\text{-}17)$$

The frequency domain weighting function is used to filter out frequencies at which the magnitude of the complex coherence is small relative to unity. As with the cross-correlation function (see Section 6.2), a peak in the SCOT occurs at a time equal to the time delay between the two signal paths. An advantage of the SCOT over the cross-correlation function is that the peak in the SCOT is oftentimes sharper, due to finite bandwidth considerations, when the signal power is larger than the noise densities at higher frequencies.

EXAMPLE: Figure 6-5 is an example from reference [6-6] which illustrates the potential benefit of the SCOT. The signal s(t) is a source with broad band noise and three sinusoids. The cross-correlation function, which is the inverse Fourier transform of the cross-spectral density, is illustrated in Figure 6-5a. As discussed in Section 6.2.2, a delay between two broad band noise signals produces a peak in the cross-correlation at a time equal to the delay time. The spectral and temporal characteristics of the source lead to a complicated cross-correlation such that the transmission path delay time cannot be distinguished with any degree of certainty. Figure 6-5b shows the SCOT. There is a clear peak at the correct delay time, normalized to zero on the horizontal axes of both plots. This figure clearly shows that the SCOT is a useful estimator or delay time for complex signals.

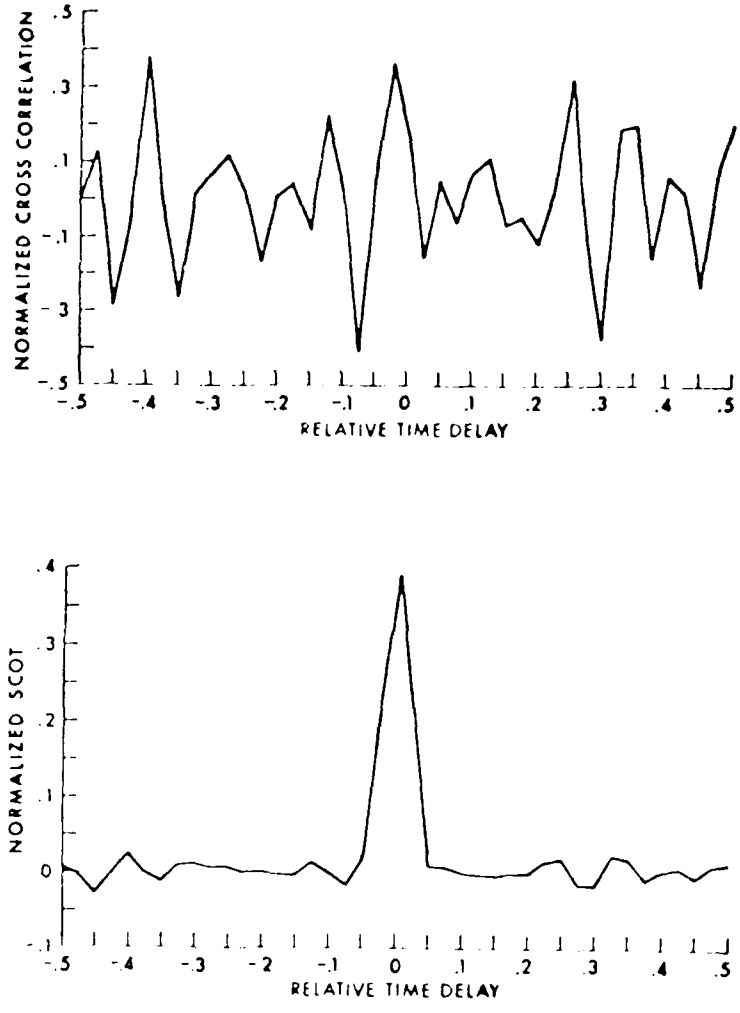


FIG. 6-5 MEASURED SMOOTHED COHERENCE TRANSFORM AND CROSS-CORRELATION FUNCTION COMPARISON

Data from Ref. 6-6

6.3 DISPERSIVE WAVE PROPAGATION

6.3.1 Introduction

The identification of transmission paths in dispersive systems such as flexural waves in structures is complicated by the fact that the phase speed of the wave excitation is not constant with frequency. As a result, different frequency components of a disturbance in the system propagate at different speeds so that the shape of the disturbance smears out and is difficult to follow in space. For flexural waves in structures, the phase speed increases as the square root of frequency. An immediate consequence of dispersion is that there is not a well-defined delay time between transmission paths.

This factor affects each of the methods used for non-dispersive systems discussed in Section 6.2. In order to relate measured travel times with the path length difference it is necessary to know the propagation speed. Since this changes with frequency in dispersive systems it is necessary to look at various narrow frequency bands individually. This prevents the use of the impulse technique since it is applicable only for broadband transient signals.

6.3.2 Cross-Correlation Analysis

6.3.2.1 Introduction

There are significant differences between the cross-correlation analysis of dispersive versus non-dispersive systems. The limitations of time-bandwidth considerations must be clearly recognized. It is remarked in Section 6.3.1 that the dispersive effects can be minimized by choosing a sufficiently small frequency bandwidth Δf in which the phase velocity is nearly constant. Such a system is nearly indistinguishable from a non-dispersive system. However, the discussion in Section

6.2.2 on the cross-correlation in non-dispersive systems notes that the time width of the peak of the cross-correlation function is inversely proportional to the frequency bandwidth (see Equation (6-6)). As the frequency bandwidth is narrowed, the peak in the cross-correlation function or correlation coefficient becomes spread out. Since the delay time is estimated by measurement of the time for which the cross-correlation function has its peak value, then it becomes difficult to estimate the delay time for excitations with a narrow frequency bandwidth. To summarize the difficulty, narrowing the frequency bandwidth to lessen the effects of dispersion increases the uncertainty in estimating the delay time.

This problem becomes more pronounced in a system with multiple transmission paths, since the individual peaks in the cross-correlation function corresponding each to a single transmission path delay time cannot be resolved as the frequency bandwidth is diminished.

6.3.2.2 Procedures

Suppose, as in Section 6.2.2, that the system is excited by band-limited noise with lower and upper cutoff frequencies f_L and f_U, respectively. The system is dispersive and is composed of a single transmission path. An example of such a system might be the propagation of (dispersive) bending waves on an aircraft fuselage from a source excitation point to a distant point on the fuselage. The cross-correlation coefficient $C_{xy}(\tau)$, defined below as the normalized cross-correlation function, is

$$C_{xy}(\tau) = R_{xy}(\tau)/[R_{xx}(0)\ R_{yy}(0)]^{1/2} \qquad (6\text{-}18a)$$

$$= (1/\Delta f) \int_{f_L}^{f_U} \cos[2\pi f(\tau - d/c_p)]\ df \qquad (6\text{-}18b)$$

where d is the transmission path length difference, Δf is the frequency bandwidth, and $c_p(f)$ is the dispersive phase velocity which varies with frequency.

Equation (6-18b) has been investigated [6-7 and 6-1] for bending waves. The phase velocity for a bending wave on a plate of thickness h and longitudinal phase velocity c_L is

$$c_p = [2\pi f\ hc_L/\sqrt{12}]^{1/2} \tag{6-19}$$

Define the center frequency f_0 of the frequency band as follows:

$$f_0 = \tfrac{1}{2}(f_L + f_U) \tag{6-20}$$

For this choice of phase velocity, the cross-correlation coefficient in Equation (6-18) is found to be,

$$C_{xy}(\tau) = [2f_0/(\Delta f^2 \tau_p)]\ \cos[\beta(Ci(z_U) - Ci(z_L))] + \sin[\beta(Si(z_U) - Si(z_L))] \tag{6-21a}$$

where

$$\beta = 2\pi f_0 \tau_p\ [(\tau/\tau_p - 3/4)^2 + 3/8] \tag{6-21b}$$

$$z = (8f_0\tau_p)^{1/2}\ [f/(4f_0) - \tau/\tau_p - 3/4] \tag{6-21c}$$

and the term z is evaluated at the upper and lower band frequencies in Equation (6-21a). The Fresnel cosine and sine integrals [6-8] are Ci(z) and Si(z), respectively.

The main features of the analysis do not require a detailed evaluation of the Fresnel integrals. The peak in the cross-correlation function (or cross-correlation coefficient) occurs at a time τ_g equal to the group delay defined as follows:

$$\tau_g = d/c_g \qquad (6\text{-}22)$$

The group delay is the time interval for energy to propagate at a velocity equal to the group velocity c_g between the points in the system at which signals X and Y are measured. The group velocity is the velocity at which energy propagates. The group velocity is not equal to the phase velocity in a dispersive system; it is given by,

$$c_g = \partial\omega/\partial k \qquad (6\text{-}23)$$

Where the radian frequency ω and wave number k are related by the dispersion relation for the system. For dispersive flexural waves in structures, the phase velocity is given in Equation (6-19) and the group velocity is,

$$c_g = 2c_p \qquad (6\text{-}24)$$

The peak of the cross-correlation function occurs at a time equal to the group delay time for a dispersive system. The peak for a nondispersive system occurs at a time equal to d/c_p, but this phase delay time is equal to the group delay since the phase and group velocities are equal in a nondispersive system.

The time width Δt of the cross-correlation function (as well as the cross-correlation coefficient) for a dispersive system is given by

$$\Delta t = \tau_p \Delta f/4f_0 > 1/\Delta f \qquad (6\text{-}25)$$

The center frequency of the excitation is f_0, and the phase delay τ_p is defined in terms of the phase velocity c_p and path length

difference d by,

$$\tau_p = d/c_p = 2\tau_g \qquad (6\text{-}26)$$

The width of the time window in Equation (6-25) has been defined to be the time interval in which the cross-correlation coefficient in Equation (6-21) is greater than or equal to one-half of its peak value. There are two differences between the width of the time window for dispersive and nondispersive systems. The width of the cross-correlation peak for a non-dispersive system is independent of the delay time (see Section 6.2.2) whereas for the dispersive system it depends on the phase delay time. Secondly, the peak width of the cross-correlation function is more spread out in a dispersive system than a nondispersive system.

The peak amplitude of the cross-correlation coefficient for a nondispersive system is equal to unity independent of time delay or frequency bandwidth. This is not the case for dispersive systems. The mean amplitude of the cross-correlation coefficient for a dispersive system within the window width corresponding to a single peak is [6-7]

$$\overline{C_{xy}} = 2\,(f_0\tau_p)^{-1/2}\,(f_0/\Delta f) \qquad (6\text{-}27)$$

Equation (6-27) is valid when the path length difference is large relative to a flexural wavelength. As the path length difference becomes small relative to a wavelength the cross-correlation peak value approaches unity.

The results of this section are summarized in the statement that the cross-correlation coefficient contains information on the dispersive system transmission path characteristics with respect to the time at which the peak occurs, the time width of

the peak, and the mean amplitude of the peak. The cross-correlation coefficient contains information on a nondispersive system only with respect to the time at which an individual peak occurs.

6.3.2.3 Example

An example of the use of the cross-correlation coefficient for a dispersive system is considered. The system under consideration is noise transmission between two widely separated points on the hull of a ship [6-7]. A third-octave noise source centered at f_0=3.15 kHz (Δf=0.23 f_0) is located at the propeller and the source wave form is measured by a hydrophone at the stern. The other location is d=108.8 meters away measured by a hydrophone at the bow. Possible transmission paths include flexural and longitudinal wave propagation through the hull, sound propagation through water along the hull, and propagation through water including a reflection off the sea bottom (61 meters below the ship).

Figure 6-6 shows the measured autocorrelation and cross-correlation coefficients for the system. The shape of the two autocorrelations are similar but not identical, which indicates that dispersive path propagation is involved. The cross-correlation coefficient has its peak at the group delay, and this peak is in the neighborhood of τ_g=126.8 msec. The hull thickness is approximately 6.35 mm corresponding to a flexural phase velocity (using Equation (6-19)) of c_p=400 m/s. If the peak in the cross-correlation coefficient is due to flexural propagation through the hull, then the measured group delay should be approximately equal to one-half of the computed phase delay for a plate of the same thickness as the hull. This phase delay is $\tau_p=d/c_p$=272 msec; this corresponds (see Equation (6-19) to a group delay of 136 msec, in good agreement with the measured value of 126.8 msec. This demonstrates that most of the energy

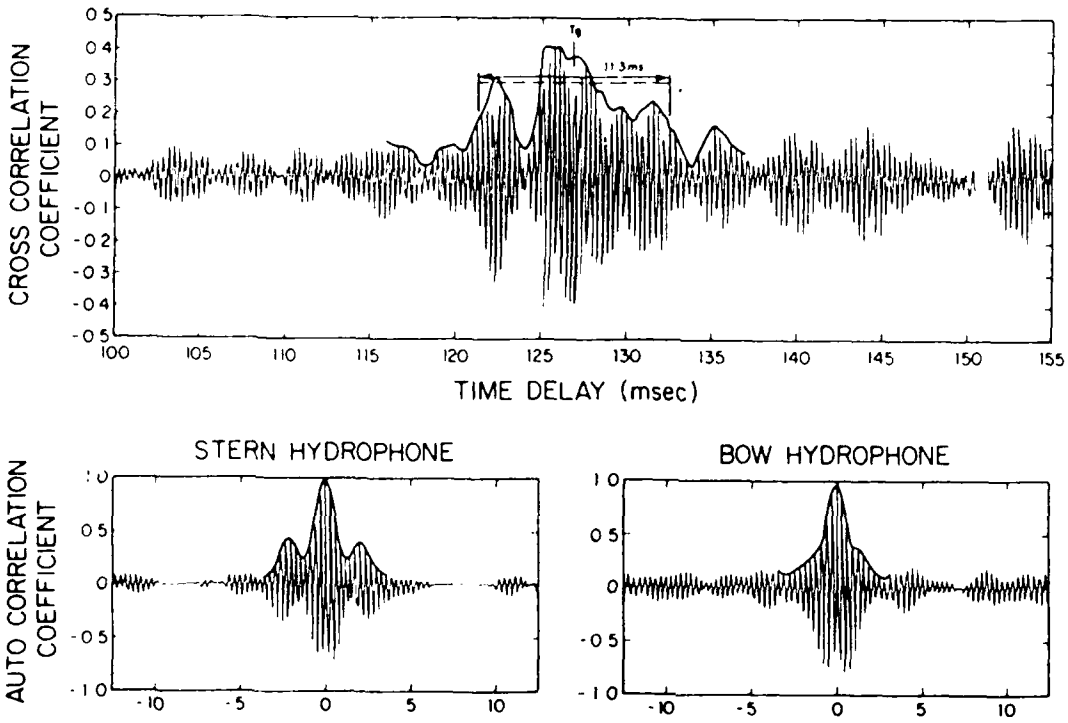


FIG. 6-6

MEASURED CROSS-CORRELATION FOR SHIP HULL TRANSMISSION

Data from Ref. 6-7

is propagating from the source to the receiver through the flexural motion of the hull. Computation of the expected delays for sound propagation through water with or without reflection off the sea bottom predict time delays which would produce peaks in the cross-correlation which are not observed.

The mean cross-correlation coefficient computed using Equation (6-27) is 0.30 using the computed phase delay and frequency bandwidth data. This is in agreement with the approximate mean of the cross-correlation coefficient as illustrated in Figure 6-6. The figure also shows that the measured time width of the peak within which it exceeds half its maximum value is 11.3 msec, whereas the predicted value from Equation (6-25) is 15.6 msec. The conclusion is that all three measures of path propagation associated with the cross-correlation coefficient imply that the dominant path is associated with flexural propagation through the hull. Correlation analysis is shown in this example to be a good method to separate out the structure-borne excitation from the fluid or air-borne excitation.

It is interesting to note that the time width of a cross-correlation peak associated with a nondispersive system excited by third-octave noise centered at 3.15 kHz is 1.38 msec. This is far less than the corresponding quantity in this particular example of dispersive propagation and underscores the differences between the use of correlation analysis of nondispersive and dispersive systems.

6.3.3 Frequency Domain Methods

6.3.3.1 Introduction

The main difficulty in determining the propagation path length in a dispersive system is that energy in different frequency bands of the signal travels at different speeds. The cross-correlation methods discussed in Section 6.3.2 are often limited in scope since the presence of dispersion produces a widening in the cross-correlation peak corresponding to a propagation path length. This difficulty is compounded in a system with multiple paths.

Two frequency domain methods are discussed in this section. The first is applicable for single path, dispersive systems. The second approach is a more general approach which depends upon a knowledge of the dispersive characteristics of the wave-bearing medium.

6.3.3.2 Cross-Spectral Density Technique

Let the time histories of a dispersive signal at two receivers be $x(t)$ and $y(t)$ such that,

$$x(t) = s(t) + m(t) \tag{6-28a}$$

$$y(t) = h(t) * s(t) + n(t) \tag{6-28b}$$

where the noise signals $m(t)$ and $n(t)$ are incoherent with one another as well as with the source signal $s(t)$, and the impulse response function between the two receiver points is $h(t)$. The convolution $h(t) * s(t)$ of impulse response function and the source signal $s(t)$ is

$$h(t) * s(t) = \int_{-\infty}^{\infty} h(\tau)\, s(t-\tau)\, d\tau \qquad (6\text{-}29)$$

as discussed in Section 5.1.2.

The cross-spectral density $G_{xy}(f)$ between the time histories $x(t)$ and $y(t)$ is equal to

$$G_{xy}(f) = H(f)\, G_{ss}(f) = |H(f)|\, e^{j\phi}\, G_{ss}(f) \qquad (6\text{-}30)$$

where the source autospectral density is $G_{ss}(f)$ and the transfer function $H(f)$ has a phase angle ϕ. Thus the phase angle ϕ_{xy} of the cross-spectral density is simply the transfer function phase angle ϕ. If the propagation path length between the receivers is d and if the medium is dispersive with group velocity c_g, then this phase is related to the propagation path length by,

$$\phi_{xy} = 2\pi f d/c_g \qquad (6\text{-}31)$$

This equation is equivalent to Equation (6-14) for the nondispersive case. In the nondispersive case the phase of the cross-spectral density increases linearly with frequency. The slope of this linear plot is 2π times the propagation delay time τ, and the delay time is the ratio of the propagation path length to the phase velocity (τ=d/c). The change in the dispersive case is that the phase velocity is replaced by the frequency-dependent group velocity.

Consider as an example the propagation of bending waves for which the group velocity is proportional to the square root of the frequency. It follows from Equation (6-31) that the phase of the cross-spectral density in such a system varies as the square root of frequency. Underlying the use of the cross-spectral density function is the condition that reverberation or other multiple scattering processes be negligible relative to the direct path. Suitable time-windowing of the time histories in

the cepstral domain (see Section 6.2.4) allow one to remove the effects of reverberation and focus attention on the direct path contribution. Figure 6-7 shows the phase of cross-spectral density of the time-windowed signals for vibration transmission in an engine structure. The phase in this figure indicates a transition in the propagation characteristics from a dispersive flexural wave at low frequencies ($\phi_{xy} \alpha \sqrt{f}$) to a nondispersive longitudinal wave at high frequencies ($\phi_{xy} \alpha f$) .

6.3.3.3 Recompression Technique Using Spatial Impulse Response Function

A second approach in the application of frequency domain methods to the task of propagation path length determination in dispersive systems is based upon the fact that dispersive signals may be recompressed if the dispersion relationship is known [6-9]. The impulse response function in a nondispersive system has a sharp peak corresponding to the time delay of each propagation path in the system. However, for dispersive systems the peak in the impulse response is spread out. Thus, the peaks corresponding to different path length differences overlap and can become indistinguishable. If the dispersive characteristics of the system are known, then the impulse response function may be processed to give sharply defined peaks. This processing is carried out through introduction of the wavenumber k and its associated transfer function H(k) as described below.

The Fourier transform of the impulse response function is the transfer function H(f) between the response points, and this transfer function is computed using the methods of Chapter 5. Suppose that the dispersion relation of the system is known; namely, that the phase velocity c_p and group velocity c_g are known as functions of frequency as discussed in Section 6.3.2. Then the frequency f is related to the spatial wave number k by,

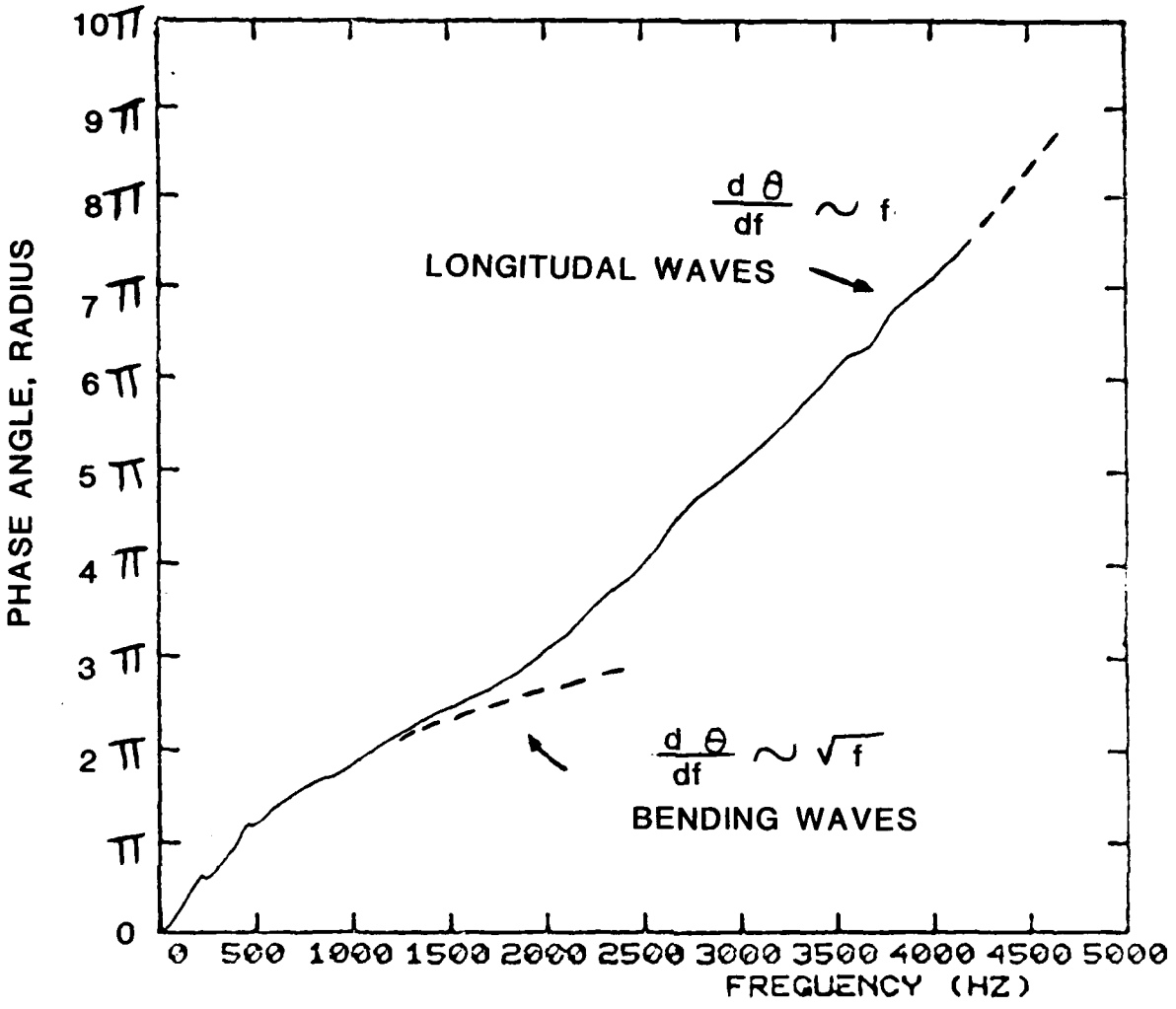


FIG. 6-7

MEASURED PHASE OF CROSS-SPECTRAL DENSITY FOR DIESEL ENGINE

$$k = 2\pi f/c_p \tag{6-32}$$

The transfer function is re-expressed in terms of the wave number k to give the spatial transfer function H(k). The inverse Fourier transform h(x) of the spatial transfer function is defined by

$$h(x) = \int_{-\infty}^{\infty} H(k)\ e^{jkx}\ dk \tag{6-33}$$

This spatial impulse response function can be computed in terms of frequency domain quantities by

$$h(x) = 2\pi \int_{-\infty}^{\infty} H(f)\ e^{j2\pi fx/c_p}\ c_g^{-1}\ df \tag{6-34}$$

The resulting spatial impulse response function h(x) has sharply defined peaks at lengths corresponding to the propagation path length differences between the two receivers.

The method of the spatial impulse response function is summarized in the following steps:

1. Measure the transfer function of the dispersive system using the techniques of Chapter 5.

2. Numerically evaluate the spatial impulse response function defined in Equation (6-34) using FFT techniques described in Chapter 3.

3. The peaks of the envelope of the spatial impulse response function occur at the propagation path length distances of each of the system transmission paths.

EXAMPLE: An example of the recompression technique is presented. The multipath reflections of flexural waves in a steel bar of thickness h=0.015 m and length 1.72 m were studied [6-10]. A shaker was attached at one end and a piezoelectric accelerometer signal h(t) was measured at the other. Under the assumption that rotary inertia and shear effects are negligible, the phase and group velocities for flexural wave propagation are,

$$c_p = [2\pi f c_L h/\sqrt{12}]^{1/2} \tag{6-35a}$$

$$c_g = 2c_p \tag{6-35b}$$

where c_L is the longitudinal phase velocity in the bar.

The transfer function H(f) was estimated using FFT methods, and the spatial impulse response function h(x) was computed using Equation (6-34) and the assumed dispersion characteristics in Equations (6-35). Figure 6-8 shows the measured impulse response function h(t) as a function of time. There are three clearly resolved peaks, whereas dispersion degrades the ability to resolve later peaks. The envelope of the spatial impulse response function h(x) is illustrated in Figure 6-9a as a function of the propagation path length. There are six readily identifiable peaks corresponding to the direct path and the first five reflections. Thus the spatial impulse response function measures propagation path length much better than the temporal impulse response function measures delay time. The propagation path length d and the delay time τ are related in terms of the phase velocity c_p by

$$d = \tau c_p \tag{6-36}$$

When the higher order effects of rotary inertia and shear are included, the phase velocity is found to vary with frequency

as $f^{0.54}$ rather than as the square root of frequency; the group velocity is also slightly altered. If this more correct dispersion relation is used, the resulting spatial impulse response function envelope shown in Figure 6-9b has more clearly defined peaks and the first seven arrivals are visible.

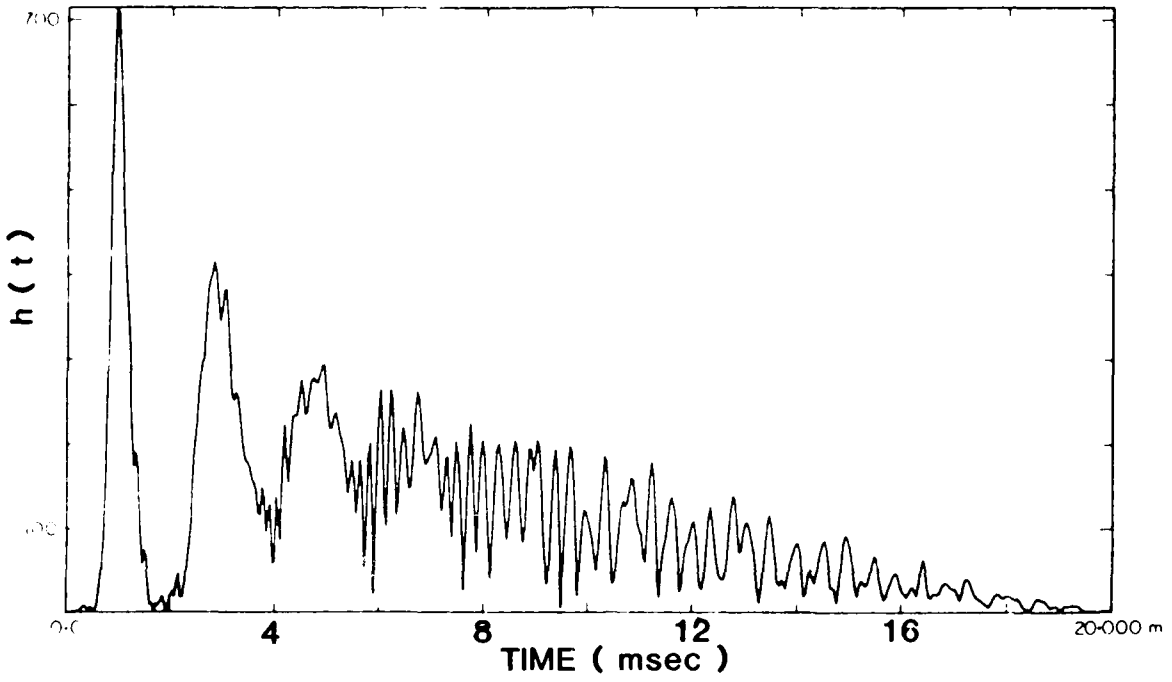


FIG. 6-8

MEASURED STANDARD IMPULSE RESPONSE FUNCTION OF BEAM

Data from Ref. 6-10

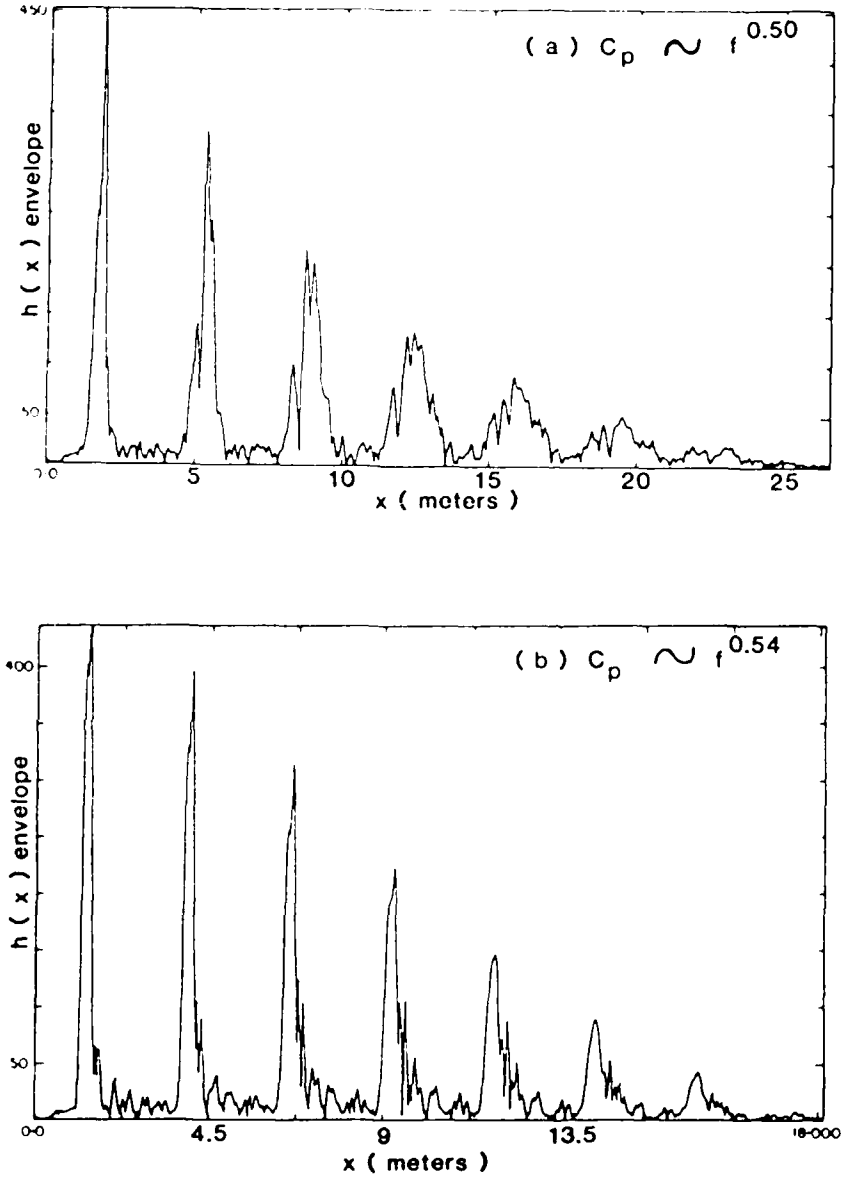


Figure 6-9. Measure Spatial Impulse Response Function of Beam

Data from Ref. 6-10

REFERENCES

6-1. P. H. White, "Cross-correlation in structural systems: dispersion and nondispersion waves", J. Acoust. Soc. Am. 45, p. 1118 (1969).

6-2. J. S. Bendat and A. G. Piersol, Engineering Applications of Correlation and Spectral Analysis (Wiley, New York, 1980), p. 129.

6-3. A. V. Oppenheim and R. W. Schafer, Digital Signal Processing (Prentice-Hall, New Jersey, 1975).

6-4. A. A. Syed, J. D. Brown, M. J. Oliver, S. A. Hills, "The cepstrum: a viable method for the removal of ground reflections", J. Sound and Vib. 71, p. 299-313 (1980).

6-5. K. W. Bushell, "Measurement and prediction of jet noise in flight", in Aeroacoustics: Jet Noise, Combustion, and Core Engine Noise, ed. I. R. Schwartz, vol. 43 in Progress in Astronautics and Aeronautics, p. 137-158.

6-6. G. C. Carter, A. H. Nuttal, and P. G. Cable, "The smoothed coherence transform", Proc. IEEE 61, p. 1497-1498 (1973).

6-7. J. E. Barger, "Noise path diagnostics in dispersive structural systems using crosscorrelation analysis", Noise Control Engr. 6, p. 122-129 (1976).

6-8. M. Abramovitz and I. Stegun, Handbook of Mathematical Functions (National Bureau of Standards, Washington, 1964).

6-9. A. K. Booer, J. Chambers and I. M. Mason, "Fast numerical algorithm for the recompression of dispersed time signals", Electronic Letters 13, p. 453-455 (1977).

6-10. P. R. Brazier-Smith, D. Butler, and J. R. Halstead, "The determination of propagation path lengths of dispersive flexural waves through structures", J. Sound and Vib. 75, p. 453-457 (1981).

CHAPTER 7. IDENTIFICATION OF SOURCES

By

D. H. KEEFE

CHAPTER 7. IDENTIFICATION OF SOURCES

TABLE OF CONTENTS

		Page
LIST OF FIGURES		7-iii
LIST OF SYMBOLS		7-iv
7.1 INPUT/OUTPUT MODEL		7-1
7.1.1 Single Source		7-1
7.1.1.1	Effect of Travel Time Errors on the Coherence Method	7-3
7.1.2 Multiple, Incoherent Sources		7-5
7.1.2.1	Presence of Noise	7-6
7.1.3 Multiple, Coherent Sources		7-7
7.1.3.1	Periodic Sources	7-7
7.1.3.2	Technique of Partially Coherent Source Identification	7-9
7.2 TRANSMISSION MODEL		7-24
7.2.1 Acoustic Intensity Method		7-24
7.2.1.1	Technique	7-24
7.2.1.2	Errors	7-28
7.2.1.3	Tests to Validate Method	7-31
7.2.1.4	Example	7-34

7.2.2 Dominant Source Location..... 7-37

7.2.2.1 Sound-Ranging Technique..... 7-37
7.2.2.2 Cross-Correlation Procedure..... 7-38
7.2.2.3 Cross-spectral Density Technique... 7-46

REFERENCES..... 7-49

LIST OF FIGURES

		Page
7-1.	Spectrum of Helicopter Cabin Noise.....	7-8
7-2.	Structual Vibration Set Up.....	7-20
7-3.	Measured Multiple Coherence Fuction.....	7-22
7-4.	Acoustic Intensity Probe.....	7-26
7-5.	Acoustic Intensity Instrumentation.....	7-29
7-6.	Diesel Engine Surface Views.....	7-35
7-7.	Measured Acoustic Intensity of Engine.....	7-36
7-8.	Source/Receiver Geometry.....	7-41
7-9.	Source Location.....	7-41
7-10.	Measured Time History.....	7-42
7-11.	Measured Cross Correlation.....	7-42
7-12.	Measured Cross Correlations for Different Hydrophone Pairs.....	7-45
7-13.	Measured Cross Correlation vs Angle.....	7-45

LIST OF SYMBOLS

c	Sound speed
d	Distance
G(f)	Spectral density function
H(f)	Transfer function
h	*Amplitude factor*
I	Acoustic intensity
j	$\sqrt{-1}$, complex number
j,k	Indexes
k	Wavenumber
L	Separation distance
L(f)	Transfer function between coherent portions of signals
n	Input variable index
n(t),N(f)	Noise signal, transform in output signal
m(t),M(t)	Noise signal, transform in input signal
p	Acoustic pressure
q	Number of inputs in multiple input systems
$R(\tau)$	Correlation function
r	Distance
s(t)	Signal time history
T	Time period
u	Acoustic particle velocity
W	Sound power
X	Generalized input variable
Y	Generalized input variable
θ	Angle, radians
ϕ	Phase, radians
$\gamma^2(f)$	Coherence function
ρ	Density
τ	Time delay

CHAPTER 7. IDENTIFICATION OF SOURCES

The problem of source identification is complementary to the techniques used in assessing system characteristics described in Chapter 5, and transmission paths in Chapter 6. A unifying technique in all three areas makes use of coherence methods. Widespread employment of this technique has occurred primarily as a result of the development of modern digital Fast Fourier Transform (FFT) processors that allow for the quick and accurate calculation of power and cross-spectral densities from sampled time histories of the measured data.

Identification of the single source is described in Section 7.1.1 using the coherence method. Effects of travel time errors on the coherence method are described in Section 7.1.2. The extension to multiple incoherent sources is given in Section 7.1.3. A step-by-step explanation of the partial coherence method for identification of partially coherent sources is given in Section 7.1.4.

Section 7.2 describes methods of source identification which depend on knowledge of the propagation characteristics of the medium. Section 7.2.1 describes the acoustic intensity method using the cross-spectral density technique. Section 7.2.2 describes source location using the cross-correlation and cross-spectral density techniques.

7.1 INPUT/OUTPUT MODEL

7.1.1 Single Source

An important factor in the use of coherence techniques involves the identification of the input variable. This identification is often based upon a prior understanding of the dynamical behavior of the physical system. A meaningful identification of an input variable associated with a particular source is also strongly dependent on being able to measure that

variable with an appropriate transducer. Angular displacements and moments present measurement problems as sources of vibratory energy in mechanical systems. Measurements of velocity fluctuations in turbulent flow as an aeroacoustic source are difficult to obtain. The outputs of response variables are generally more readily quantified in terms of vibration at a particular location(s) or an average of vibration levels over a region on the structure, or pressure levels at a particular location(s) or average over a region in space.

We consider the single input, single output system illustrated in Figure 5-1. If the dynamical behavior of the system is specified by means of having previously identified the transmission path transfer function magnitude $|H(f)|$, then the source excitation spectral density $G_{xx}(f)$ can be calculated through the measured response spectral density $G_{yy}(f)$ by means of the formula:

$$G_{xx}(f) = G_{yy}(f)/|H(f)|^2 \qquad (7\text{-}1)$$

The effect of the transducer response on this measurement is analogous to the discussion in Section 5.1.3, and the effect of output noise is characterized in Section 5.1.4. The random and bias errors associated with measurement of the output auto-spectral density are discussed in Section 5.2.1.1.

Source identification procedures also make use of the coherence function $\gamma_{xy}^2(f)$ between the input and output signals. This is defined as,

$$\gamma_{xy}^2(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)\, G_{yy}(f)} \qquad (7\text{-}2)$$

This coherence function equals unity in the absence of extraneous noise at either the input or output of the system. That part of

the output autospectral noise density which is coherent with the input is,

$$G_{yy|x}(f) = \gamma_{xy}^{2}(f)\ G_{yy}(f) \qquad (7\text{-}3)$$

The coherence method of source identification provides a check on the modelling of the system. If the signal at the output is primarily due to other independent inputs which have not been included in the model, then the coherence function will be small relative to unity. The effect of noise on the coherence function is discussed in Section 5.1.4. In particular, Equation (5-32) gives the coherence function as a function of transducer response characteristics and input and output noise levels.

7.1.1.1 Effect of Travel Time Errors on the Coherence Method

One source of error in a measurement of this sort is one's relative ignorance about the shape of the transfer function. Particularly near small values of $|H(f)|$, a small error in the transfer function magnitude produces a large error in the estimated source power output.

Additional errors occur in estimating the cross-correlation function between two stationary, random signals, and thereby affect the coherence functions and transfer functions. An important error is due to propagation travel time differences between the input and output signals. See Section 3.3.3 for details on eliminating this travel time error by means of computer data processing. Specialized dual channel digital spectrum analyzers do not typically have the capacity to eliminate this error. Suppose that the output signal is delayed by time τ relative to the input signal and that the FFT sampling period is T (see Section 3.1.1.3). The true value of the cross-correlation function is $R_{xy}(t)$. The delay between the channels

introduces a biased estimate $\hat{R}_{xy}(t)$ of the cross-correlation function as follows [7-1]:

$$\hat{R}_{xy}(t) = (1 - \tau/T)\ R_{xy}(t) \qquad (7\text{-}4)$$

The cross-spectral density function estimate $\hat{G}_{xy}(f)$, being the transform to the cross-correlation function, is biased by the factor $(1 - \tau/T)\ e^{-jf\tau}$ as long as the delay time is independent of the frequency f. The estimated transfer function $\hat{H}(f)$ is also biased by this same factor since it is computed in terms of the cross-spectral density (see Equation 5-25a). The estimated value of the coherence function is affected by the following bias error [7-2]:

$$\hat{\gamma}_{xy}^{\ 2} = (1 - \tau/T)^2\ \gamma_{xy}^{\ 2} \qquad (7\text{-}5)$$

Scattering and reverberation at the output location can produce errors in the spectral estimates since they both involve modifications in the travel time between the input and output transducer locations in the case of scattering or involve multiple delay times in the case of reverberation. Significant scattering from any object can significantly affect the phase of the spectral estimate. Reverberation refers to the process of multiple reflections of a signal within an enclosure. Each of the reflections has its own delay time. In order that the bias error be negligible the longest delay time must be much smaller than the single record sampling period T. A practical criterion is that the ratio of the reverberation time T_{60} discussed in Section 5.4.1 be much less than the sampling period T.

7.1.2 Multiple, Incoherent Sources

Most source identification problems involve the resolution of the combined effect of two or more sources. One often carries out source identification studies as an initial step in achieving acoustic or vibration noise control. After the dominant sources are identified one is in a much better position to design suitable changes in the transmission path or source characteristics. Coherence techniques can be used to identify the contribution from each of the various sources. This section describes the use of coherence techniques under the condition that the various sources are statistically independent noise sources. Related material on systems with multiple, incoherent sources is found in Section 5.2.2.

Two noise sources are statistically independent or equivalently uncorrelated when the coherence function or cross-correlation function of the two inputs is zero. Consider the system shown in Figure 5-12 consisting of N input sources x_i (i=1,...,N) and an output signal y. The coherence function $\gamma^2_{iy}(f)$ between the i-th input and the ou put is defined by,

$$\gamma^2_{iy}(f) = \frac{|G_{iy}(f)|^2}{G_{ii}(f)\ G_{yy}(f)} \qquad (7\text{-}6)$$

where $G_{ii}(f)$ and $G_{yy}(f)$ are the autospectral density functions of the i-th input and the output, respectively, and $G_{iy}(f)$ is the cross-spectral density function between the i-th input and the output. The coherence functions in the absence of noise satisfy the following sum rule:

$$1 = \sum_{i=1}^{N} \gamma^2_{iy}(f) \qquad (7\text{-}7)$$

The output power is the sum of the power from each of the input sources as filtered by each of the transmission paths. The contribution $G_{yy:i}(f)$ to the output autospectral density from the i-th path is

$$G_{yy:i}(f) = \gamma_{iy}^2(f)\ G_{yy}(f) \qquad (7\text{-}8)$$

The sum of $G_{yy:i}(f)$ over all paths is seen from Equations (7-7) and (7-8) to be equal to the total power $G_{yy}(f)$.

7.1.2.1 Presence of Noise

Now suppose that there is noise present at the output measurement point of the system. This 'noise' may be noise present in the transducer or digital signal processing as discussed in Chapters 2 and 3. However, it may also be that the multiple input model which is employed does not have a sufficient number of sources. In order to assess whether the multiple input/output model correctly models the system behavior, the multiple coherence function $\gamma_{y:x}^2(f)$ can be used. It is defined as the ratio of the output energy due to all the inputs to the total measured output energy. Defining the output noise autospectral density function as $G_{nn}(f)$, then the multiple coherence function is,

$$\gamma_{y:x}^2(f) = \frac{G_{yy}(f) - G_{nn}(f)}{G_{yy}(f)} \qquad (7\text{-}9)$$

This function is discussed in Section 5.3.1 within the context of systems with multiple inputs which are partially coherent. However, the multiple coherence function is useful as well when the inputs are incoherent as is the case here.

When all the inputs are incoherent, the multiple coherence function is related to the input/output coherence functions by

$$\gamma_{y:x}^2(f) = \sum_{i=1}^{N} \gamma_{iy}^2(f) \qquad (7\text{-}10)$$

The multiple coherence function in the absence of output noise is by definition equal to unity, signifying that all of the output signal is derived from the multiple inputs. In this case the sum of the coherence functions between each input and the output is also unity, as expressed in Equation (7-7). Equation (7-10) demonstrates that this sum is equal to the multiple coherence function and is a measure of the presence of output noise, or alternatively, additional inputs which have not been correctly modelled.

The coherence function method is limited by travel time errors, scattering and reverberation whether it is applied to single or multiple source systems. These limitations are discussed in Section 7.1.1.1.

7.1.3 Multiple, Coherent Sources

7.1.3.1 Periodic Sources

There exist many systems such that there are multiple inputs to the system which are coherent. If the inputs are of a broadband random or quasi-random character then the partial coherence method may be used. It is often the case that the spectral content of the signal is composed of sharply defined peaks such as is the case with periodic or quasi-periodic signals. Prior knowledge of the inputs which compose the system can assist in the task of source identification.

Figure 7-1 shows the narrow-band rms pressure level spectrum within the cabin of a helicopter. The spectrum consists of harmonic components due to radiated machinery noise superposed

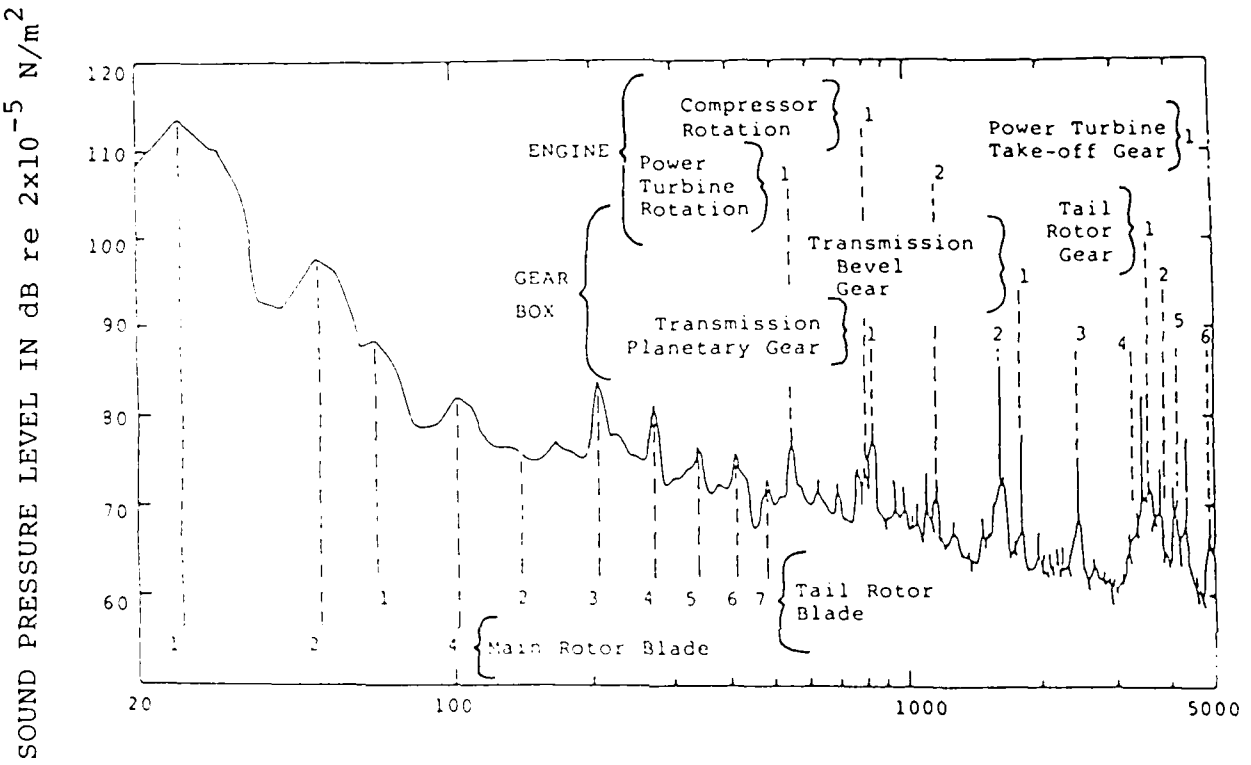


Figure 7-1. Spectrum of Helicopter Cabin Noise
Stripped Interior-Speed 140 Knots

Data from Ref. 7-12

upon a broadband acoustic signal due primarily to turbulent boundary layer noise. The correspondence of lines in the noise spectrum with fundamental and harmonic frequencies of the different rotating machinery serves to identify these components as major noise sources. The primary source is the main gearbox. Tail rotor noise is significant in the frequency range from 100 to 500 Hz but is not a major component of the A-weighted noise. Although the lines in the narrow band spectrum make the major contribution to the A-weighted cabin noise level, the broadband sources are also important. These sources are associated not only with fluctuating pressures in the boundary layer but also broadband components of rotor noise and combustion noise in the turbines.

Periodic signals of separate origin such as those identified in the Figure 7-1 are always coherent with one another, independent of their repetition rates. These periodic signals are much easier to correlate with the sources given the knowledge of the various machinery rotation rates. On the other hand, if one wishes to identify the most important component in the broadband noise, the techniques of partial coherence would be an appropriate tool.

7.1.3.2 Technique of Partially Coherent Source Identification

Section 5.3 introduces the techniques applicable to systems with multiple, partially coherent inputs. The present section applies these techniques to the task of source identification.

Consider the system of Figure 5-12 which shows a system with n inputs (X_1,..., X_i,..., X_n) and a single output Y. A noise source N is present at the output which is incoherent with all the inputs. The system transfer function corresponding to the i-th input is $H_i(f)$. What is desired as regards source identification is to determined the extent to which each source input

contributes to the total output [7-2]. The nine steps in this source identification are described below.

1. The initial step is the collection of experimental data. The autospectral densities of the output and all the inputs are measured. The cross-spectral density functions between each of the inputs and the output is measured. Finally, the cross-spectral density function between each pair of inputs is measured. Methods for estimating these spectral density functions and the errors involved are discussed in Chapters 3 and 5. The effect of travel time and reverberation errors on the cross-spectral density function is discussed in Section 7.1.1.1. All other steps in the partial coherence method of source identification consist of applications of the measured data obtained in this initial step.

2. The second step in source identification is to compute the coherence function γ_{iy}^2 between the i-th input and the output as defined by Equation (7-6). The measured data necessary to compute these coherence functions are the source autospectral densities $G_{ii}(f)$, the output autospectral density $G_{yy}(f)$, and the input/output cross-spectral densities $G_{iy}(f)$. If any of these coherence functions are zero, then the associated inputs may be eliminated from the model.

3. The third step is to compute the coherence function γ_{ij}^2 between all pairs of inputs X_i and X_j; this coherence function is,

$$\gamma_{ij}^2(f) = \frac{|G_{ij}(f)|^2}{G_{ii}(f)\, G_{jj}(f)} \qquad (7\text{-}11)$$

Calculation of these coherence functions depends upon the measured cross-spectral density $G_{ij}(f)$ between all pairs of

inputs and the input autospectral densities. If the coherence function γ_{ij}^2 is identically zero for each pair of inputs, then the sources are incoherent with one another and the methods of Section 7.1.2 should be used. In practice, two sources are regarded as incoherent if their coherence function is less than 0.1. If any of the inter-source coherence functions exceeds 0.1, then proceed to step four.

If the coherence between any pair of inputs equals unity then there exists a deterministic relationship between the inputs as discussed in Section 5.2.3. In this case, the number of inputs in the multiple input model may be reduced by one for each pair of coherent inputs. A practical criterion for the inputs to be coherent is that the coherence function between the inputs exceed 0.9 over the frequency bandwidth of interest.

4. Step four consists of ordering the inputs according to the magnitude of the input/output coherence function. Input one is defined to be the input whose coherence function γ_{iy}^2 with the output exceeds that of any other source. Input two is the input with the next highest coherence function, and so forth. At this point in the calculation there are q partially coherent inputs, where q is less than or equal to the original number of inputs. This potential reduction in the number of inputs is due to the elimination of inputs which do not contribute to the output, and the combination of inputs coherent with one another into a single input.

5. Step 5 describes the process of conditioning the output and inputs 2 through n upon input 1. Input j (where j=2,...,q) and the output are each written in terms of a part which is coherent with input 1 and a part which is incoherent as follows:

$$X_j = L_{j|1}X_1 + X_{j|1} \qquad (7\text{-}12a)$$

$$Y = L_{y|1}X_1 + Y_{y|1} \qquad (7\text{-}12b)$$

The parts of input j and the output which are incoherent with input 1 are $X_{j|1}$ and $Y_{y|1}$, respectively. The coherent part of each of these signals with X_1 may be expressed in terms of the transfer functions $L_{j|1}$ for input j and $L_{y|1}$ for the output. These transfer functions are computed in terms of the measured autospectral and cross-spectral densities by,

$$L_{j|1} = G_{1j}/G_{11} \tag{7-13a}$$

$$L_{y|1} = G_{1y}/G_{11} \tag{7-13b}$$

Having computed the inputs and outputs conditioned upon input 1 and the associated transfer functions, the autospectral and cross-spectral densities conditioned upon input 1 are computed as follows:

$$G_{jk|1} = G_{jk} - L_{k|1}G_{j1} \quad , \text{ for } j,k=1,\ldots,n \tag{7-14a}$$

$$G_{jy|1} = G_{jy} - L_{y|1}\,G_{j1} \tag{7-14b}$$

By definition, $L_{1|1}$ is taken to equal unity. Physically, Equation (7-14) gives the various cross-spectral and autospectral densities (obtained by setting j=k in Equation (7-14a)) which would be measured if the portion of all the signals coherent with input 1 could be eliminated, and this physical correspondence is the intuitive justification for the partial coherence method. A more convenient form for the autospectral density functions is,

$$G_{jj|1} = G_{jj} - |L_{j|1}|^2\, G_{11} \tag{7-15a}$$

$$G_{yy|1} = G_{yy} - |L_{y|1}|^2\, G_{11} \tag{7-15b}$$

In summary, Equation (7-12) motivates the introduction of the conditioned inputs. However, this equation is not directly involved in calculations, since it is not expressed in terms of the measured spectral density functions. Equations (7-13,14,15) fulfill this latter function.

6. The next stage in the partial coherence calculation is to condition the output and the remaining inputs as regards their degree of coherence with input 2. The input $X_{j|1}$ conditioned upon input 1 is written as a sum of a term which is coherent with $X_{2|1}$ (implicitly defined by Equation (7-12)) and a term which is incoherent as follows:

$$X_{j|1} = L_{j|2!}\ X_{2|1} + X_{j|2!} \qquad (7\text{-}16)$$

That part of input j (where j=3,...,q) which is incoherent with both inputs 1 and 2 is $X_{j|2!}$. That part of $X_{j|1}$ which is coherent with $X_{2|1}$ is related by the optimum transfer function $L_{j|2!}$.

Similarly, the output conditioned upon input 1 is written as a signal $Y_{y|2!}$ incoherent with input 2 and a signal coherent with input 2 with the result

$$Y_{y|1} = L_{y|2!}\ X_{2|1} + Y_{y|2!} \qquad (7\text{-}17)$$

The transfer function $L_{y|2!}$ is optimal in the least squares sense in connecting the conditioned input $X_{2|1}$ to the conditioned output $Y_{y|1}$.

The transfer functions are computed in terms of the conditioned spectral densities defined in the previous step by,

$$L_{j|2!} = \frac{G_{2j|1}}{G_{22|1}} \tag{7-18a}$$

$$L_{y|2!} = \frac{G_{2y|1}}{G_{22|1}} \tag{7-18b}$$

Note that these transfer functions conditioned upon the first two ordered inputs depend upon the spectral densities conditioned upon the first input. Similarly, the transfer functions conditioned upon the first three inputs will depend upon the spectral densities conditioned upon the first two inputs. These conditioned cross-spectral densities are given by,

$$G_{jk|2!} = G_{jk|1} - L_{k|2!}\, G_{j2|1} \tag{7-19a}$$

$$G_{jy|2!} = G_{jy|1} - L_{y|2!}\, G_{j2|1} \tag{7-19b}$$

The conditioned autospectral densities are,

$$G_{jj|2!} = G_{jj|1} - |L_{j|2!}|^2\, G_{22|1} \tag{7-20a}$$

$$G_{yy|2!} = G_{yy|1} - |L_{y|2!}|^2\, G_{22|1} \tag{7-20b}$$

The desired system relations obtained from this step are Equations (7-18) through (7-20). If there are only two sources in the system then proceed to step 8.

7. The previous two steps have conditioned the input/output signals on the first two inputs. This step describes the general iterative algorithm used for conditioning the input on the first k inputs. This step is repeated for k=3 up to k=q, where q is the total number of ordered inputs obtained in step 4. The output and the inputs conditioned upon the inputs 1 through k are,

$$X_{j|(k-1)!} = L_{j|k!}\, X_{k|(k-1)!} + X_{j|k!} \qquad (7\text{-}21a)$$

$$Y_{y|(k-1)!} = L_{y|k!}\, X_{k|(k-1)!} + Y_{y|k!} \qquad (7\text{-}21b)$$

This equation is equivalent to Equations (7-12) of step 5 and Equations (7-16,17) of step 6.

The optimum transfer functions occurring in the above are calculated by the general formulas,

$$L_{j|k!} = \frac{G_{kj|(k-1)!}}{G_{kk|(k-1)!}} \quad , \quad \text{for } k<j \qquad (7\text{-}22a)$$

$$L_{y|k} = \frac{G_{ky|(k-1)!}}{G_{kk|(k-1)!}} \qquad (7\text{-}22b)$$

If $k>j$ then $L_{j|k!}=0$; and if $k=j$ then $L_{k|k!}=1$. For the cases where k equals 1 or 2, Equation (7-20) is equivalent to Equations (7-13) or (7-18), respectively. Note that Equation (7-22) depends upon a knowledge of the spectral density functions conditioned upon the first (k-1) ordered inputs which are calculated in the previous iteration.

The cross-spectral density functions conditioned upon tne first k ordered inputs are computed from,

$$G_{ij|k!} = G_{ij|(k-1)!} - L_{j|k!}\, G_{ik|(k-1)!} \qquad (7\text{-}23a)$$

$$G_{jy|k!} = G_{jy|(k-1)!} - L_{y|k!}\, G_{jk|(k-1)!} \qquad (7\text{-}23b)$$

These equations are equivalent to Equations (7-14) or (7-19) for tne cases when k equals 1 or 2, respectively. The corresponding formulas for the autospectral densities are,

$$G_{jj|k!} = G_{jj|(k-1)!} - |L_{j|k!}|^2 G_{kk|(k-1)!} \quad (7\text{-}24a)$$

$$G_{yy|k!} = G_{yy|(k-1)!} - |L_{y|k!}|^2 G_{kk|(k-1)!} \quad (7\text{-}24b)$$

These equations reduce to Equations (7-15) or (7-20) for k equal to 1 or 2, respectively.

This completes the k-th iteration of the process of conditioning the input/output signals. As stated earlier, this step is repeated until the iteration is completed for which k=q.

8. The partial coherence functions, output noise autospectrum and multiple coherence function are computed in this step. The partial coherence function $\gamma_{iy|k!}^2$ between the i-th input and the output conditioned upon the first k ordered inputs is,

$$\gamma_{iy|k!}^2 = \frac{|G_{iy|k!}|^2}{G_{ii|k!}\, G_{yy|k!}} \quad (7\text{-}25)$$

This equation is meaningful only for k<i. For purposes of source identification the only partial coherence functions of interest are the functions $\gamma_{iy|(i-1)!}^2$, where i=1,...,q. A technical detail of the factorial notation is that the subscript 0! can always be neglected. In this particular instance the coherence function for i=1 is $\gamma_{1y|0!}^2$, which is just the ordinary coherence function γ_{iy}^2 between input 1 and the output.

During the last iteration of the conditioning process in step 7, the output autospectral density $G_{yy|q!}$ conditioned upon all q of its inputs is computed. This represents the part of the output power which is incoherent with each of the input signals. Thus, it represents the noise signal which contributes to the output signal as illustrated in Figure 5-13. The noise autospectral density G_{nn} is thereby computed from the partial coherence analysis and is,

$$G_{nn} = G_{yy|q!} \tag{7-26}$$

The multiple coherence function $\gamma_{y:x}^2$ is defined as the ratio of the output energy due to all the inputs relative to the total output energy as follows:

$$\gamma_{y:x}^2 = \frac{G_{yy} - G_{nn}}{G_{yy}} \tag{7-27}$$

The multiple coherence is evaluated in Section 5-3 by means of a product of q terms involving the partial coherence functions $\gamma_{iy|(i-1)!}^2$, and this relationship is

$$\gamma_{y:x}^2 = 1 - \{(1-\gamma_{1y}^2)(1-\gamma_{2y|1}^2)\ldots(1-\gamma_{qy|(q-1)!}^2)\} \tag{7-28}$$

This completes step 8.

9. The final step is the source identification phase. That part $G_{yy:x}$ of the output autospectral density which is due to the presence of the input signals is given by,

$$G_{yy:x} = \gamma_{y:x}^2 G_{yy} \tag{7-29}$$

If the multiple coherence function equals unity then all of the output signal is due to the multiple input excitation. A value of the multiple coherence function much less than unity means that there is a large "noise" signal at the output. Upon checking the experimental measurement techniques it may be that there is a negligible amount of instrument noise. This implies errors in the selection of inputs for the system or in the signal processing. If important inputs are left out of the model, then the contributions to the output from these unaccounted for inputs appear as noise at the output and reduce the multiple coherence function.

A second type of error which can reduce the multiple coherence function is any errors in the signal processing. Equation (5-85) of Chapter 5 gives the random error in the estimate of the multiple coherence function, and this error can be reduced by increasing the number of independent sample time histories.

A more fundamental error is bias error due to travel time, scattering and reverberation effects as was discussed in Section 7.1.1.1. Particularly with regard to structures for which the reverberation time is quite long, the family of multiple delay times which compose the reverberant field degrade the utility of the coherence techniques. One symptom of this bias error is a reduction in the multiple coherence function. Although all the inputs are accounted for by the model, reverberation reduces the coherence of a transmitted signal. This problem is particularly acute in systems such as structures with dispersive path propagation, since the de-reverberation techniques described in Chapter 6 are not well developed for dispersive systems. Bendat [7-3] suggests the criterion that reverberation bias is negligible is that the multiple coherence function should exceed 0.5 at all frequencies in the bandwidth under investigation.

Investigation of the partial coherence functions $\gamma_{iy|(i-1)!}^2$ allows one to determine the output power coherent with each of the conditioned inputs. The output power may be written as,

$$G_{yy} = \gamma_{1y}^2 G_{yy} + \gamma_{2y|1}^2 G_{yy|1} + \cdots + \gamma_{qy|(q-1)!}^2 G_{yy|(q-1)!} + G_{nn} \tag{7-30}$$

The first term is the power coherent with input 1, the second term is the power coherent with input 2 after the inputs have been conditioned to remove the parts coherent with input 1, and

so forth. Equation (7-30) is the decomposition of the output according to the ordered input coherent contributions, and allows one to identify at any given frequency which conditioned input is the dominant source as regards output power. The j-th term in Equation (7-30) is the partial coherent output $G_{yy:j|(j-1)!}$ of the j-th input such that the effects of the first (j-1) inputs have been removed, and equals

$$G_{yy:j|(j-1)!} = \gamma^2_{jy|(j-1)!} \; G_{yy|(j-1)!} \qquad (7\text{-}31)$$

This completes the partial coherence source identification process.

EXAMPLE: These partial coherence techniques have been applied to the study of the relative acoustic and structural vibration excitation of an instrumentation truss from a Titan missile [7-4] as shown in Figure 7-2. The system was artificially excited by three band-limited noise sources whose degree of partial coherence could be manipulated. Two were mechanical inputs applied by shakers mounted on two pushrods of the foundation, and the other input was an acoustic excitation delivered via a horn. The vibration input transducers were accelerometers mounted close to each shaker, and a microphone mounted a fixed distance away from the instrumentation truss. The output transducer was an accelerometer mounted at the position of the instrumentation.

Even if the three noise sources are incoherent the "input" signal sensed by either input accelerometer is partially coherent with each of the sources. This is because the excitation produced by shaker 2 is transmitted to accelerometer 1 through the structure, and the acoustic excitation couples to the structural vibration of the truss and thus produces an acceleration at the location of accelerometer 1.

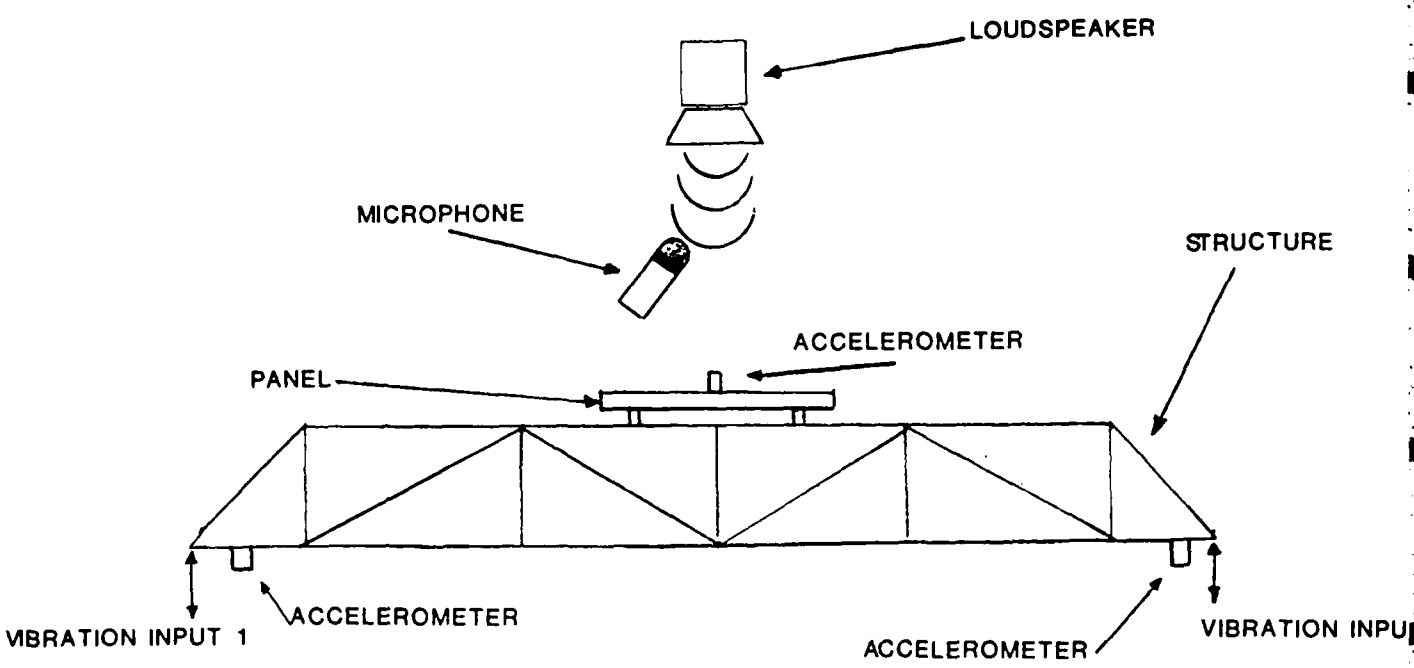


FIGURE 7-2 STRUCTUAL VIBRATION SET UP

Example from Ref. 7-A

Although the instrumentation truss was excited by broadband sources, source identification is important only at those frequencies for which there is a sharp peak in the accelerance of the output. The multiple coherence function is illustrated in Figure 7-3. This function is small over much of the bandwidth, but it has several sharp peaks where its magnitude is on the order of 0.7-0.8. These peaks occur at the frequencies of the accelerance peaks, so that the multiple coherence function is sufficiently large to identify which source is exciting which resonance. Table 7-1 exhibits the partial coherent power analysis at each of the truss resonance frequencies [7-5]. The last column of the table shows that source 3 (the acoustic input) is the dominant source of power injection at the two highest frequencies, 290 and 420 Hz, and makes a negligible contribution at other frequencies. The response at the lower frequencies is a complicated sum of the two input shaker excitations.

The fact that the multiple coherence function differs from unity signals the presence of a large "noise" signal. The main sources for this apparent noise signal are bias errors in the spectral estimates and errors due to reverberation in the structural response.

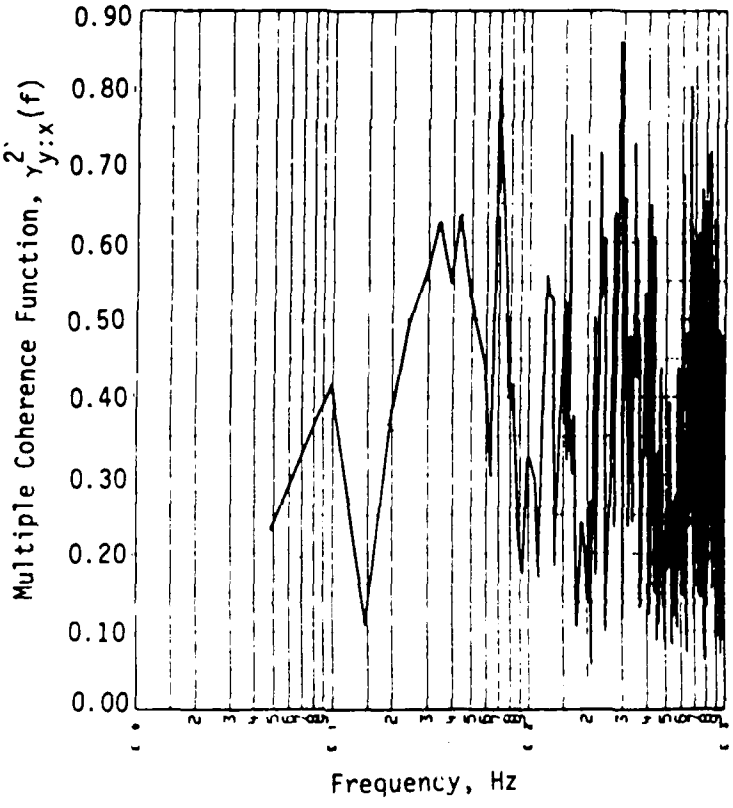


FIGURE 7-3 MEASURED MULTIPLE COHERENCE FUCTION

Data from Ref. 7-4

Table 7-1

Partial Coherent Output Power[7-2]

	Coherence Functions				Outputs, g^2/Hz		Partial Coherent Outputs g^2/Hz and (%)		
Frequency, of peak Hz	$\gamma^2_{y:x}$	γ^2_{1y}	$\gamma^2_{2y\|1}$	$\gamma^2_{3y\|2!}$	Total, G_{yy}	Coherent $G_{yy:x}$	$\gamma^2_{1y}G_{yy}$	$\gamma^2_{2y\|1}G_{yy\|1}$	$\gamma^2_{3y\|2!}G_{yy\|2!}$
68	0.81	0.78	0.09	0.04	0.91	0.74	0.71(96%)	0.02(3%)	0.01(1%)
161	0.73	0.44	0.47	0.08	0.41	0.30	0.18(60%)	0.11(37%)	0.01(3%)
230	0.72	0.15	0.56	0.26	0.14	0.10	0.02(20%)	0.07(70%)	0.01(10%)
290	0.86	0.02	0.01	0.85	0.12	0.10	<0.01(-)	<0.01(-)	0.09(90%)
420	0.64	0.05	0.06	0.60	1.2	0.77	0.06(8%)	0.07(9%)	0.64(83%)

Data from Ref. 7-4

7.2 TRANSMISSION MODEL

7.2.1 Acoustic Intensity Method

7.2.1.1 Technique

The acoustic intensity method involves scanning over a vibrating and radiating surface with a transducer which generates outputs that can be processed to give the acoustic intensity or power. The net acoustic power from the surface is obtained from an average of the intensity evaluated at close distances to the surface and at a sufficient number of locations over the radiating area. By performing this averaging process for the different component surfaces an estimate is obtained of the contributions of each to the total sound power radiated from the body.

This method does not generate information characterizing the directivity of the acoustic power of individual sources and therefore of the total body itself. A ranking of the contributions to the pressure levels measured at a particular distance and orientation with respect to the body is not precisely defined by the intensity measurements. Measured pressure levels depend on the overall directivity of the radiating body, the orientation and distance of the measurement location from the body and the acoustic characteristics of the space adjacent to the microphone location.

It is important to distinguish between the intensity and pressure or velocity variables which are used as inputs and outputs. The intensity is a flow variable describing a net flow of energy through a surface in space. It is obtained as a time average of the product of in-phase components of pressure and velocity on the surface. For different surfaces which have coherent components of velocity an additional factor is involved in assessing the influence of that coherence on the radiated

intensity. This factor is that there is a mutual coupling between the radiation impedances of the two surfaces. That is, the pressure at surface one is influenced by the motion of surface one, but it is also influenced by the motion of surface two as well as any other vibrating surfaces.

The definition of acoustic intensity in terms of acoustic pressure and flow is now introduced. The acoustic pressure p and particle velocity u_r in the r-th direction are related by

$$u_r = - (1/\rho) \int dt \; \partial p/\partial r \qquad (7\text{-}32)$$

where the equilibrium density of air is ρ and the integration is over time. Thus measurement of the pressure gradient $\partial p/\partial r$ in the r-th direction enables a computation of the flow in that direction. The acoustic intensity I_r in the r-th direction is defined to be the time-average of the product of pressure and flow as follows:

$$I_r = \langle p u_r \rangle \qquad (7\text{-}33)$$

where the brackets '<...>' denote a time-average.

The acoustic intensity is measured using two pressure-sensing microphones spaced a distance Δr apart as illustrated in Figure 7-4. The component of the acoustic intensity parallel to the line joining the microphones is measured. Let the pressure time history from the microphone closest to the source be $p_1(t)$ and that from the microphone farthest from the source be $p_2(t)$. The pressure p and pressure gradient $\partial p/\partial r$ at the midpoint between the microphones may be estimated using the finite difference approximation by,

$$p = \tfrac{1}{2} (p_1 + p_2) \qquad (7\text{-}34a)$$

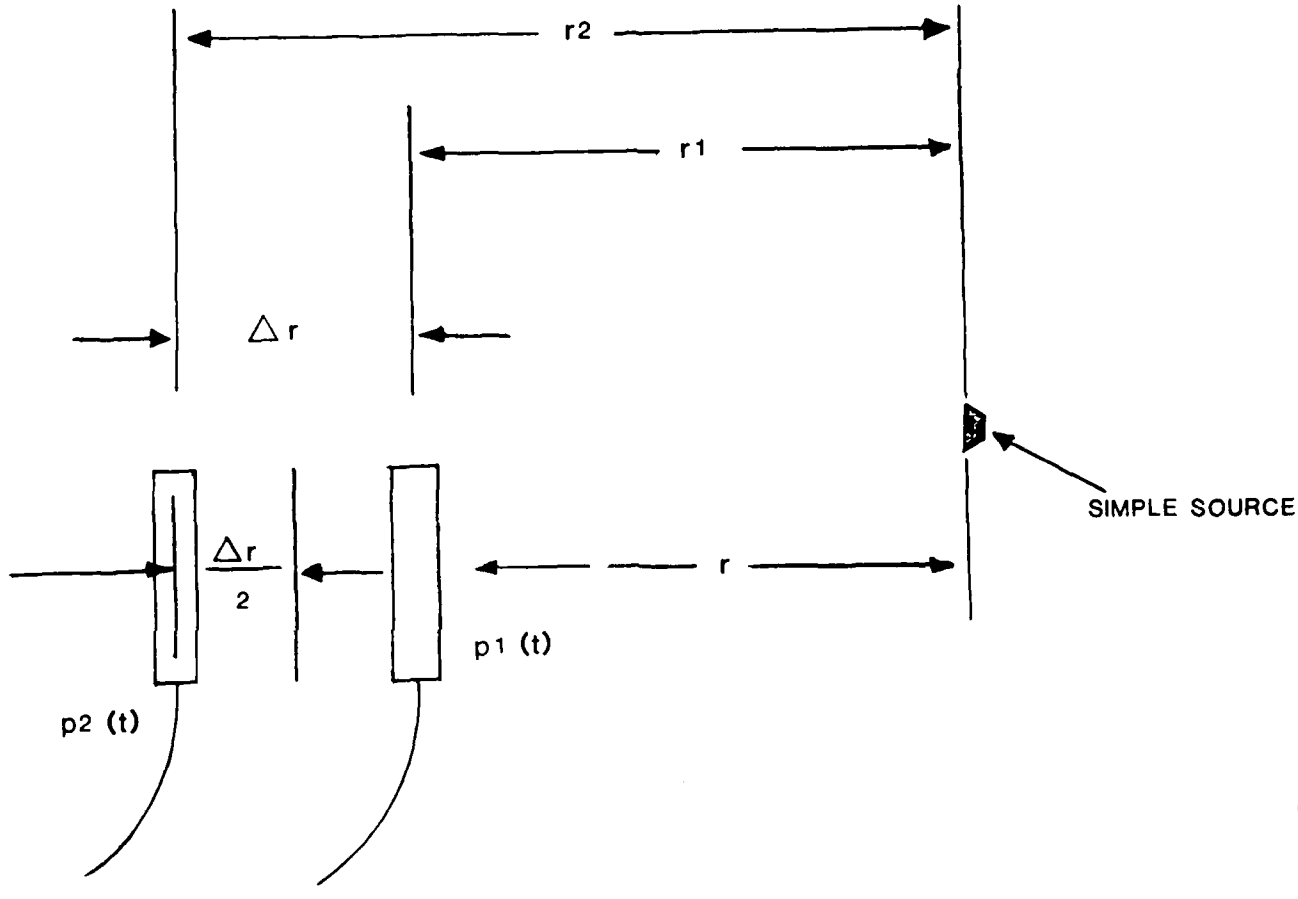


FIGURE 7-4 ACOUSTIC INTENSITY PROBE

$$\partial p/\partial r = (p_2 - p_1)/\Delta r = \Delta p/\Delta r \qquad (7\text{-}34b)$$

so that Equation (7-32) for u_r becomes

$$u_r = - (1/\rho\Delta r) \int (p_2 - p_1)\, dt \qquad (7\text{-}35)$$

The accuracy of the finite difference approximation is discussed later in this section. The acoustic intensity may be written in terms of the two microphone pressures by,

$$I_r = \langle(-1/2\rho\Delta r)\ (p_1+p_2) \int dt\ (p_2-p_1)\rangle \qquad (7\text{-}36)$$

The time domain approach using analog methods is to construct the sum and difference of the pressure signals histories, integrate the pressure difference and carry out the necessary multiplication of the signals. Time-averaging of the resultant product leads to the acoustic intensity.

With the widespread acceptance of digital signal processing techniques, it is more convenient to estimate the acoustic intensity using a frequency domain approach constructed using the sampled time series of the two pressure signals. Fahy [7-6] shows that the acoustic intensity spectrum $I_r(f)$ may be expressed in terms of the imaginary part of cross-spectral density function $G_{12}(f)$ by,

$$I_r(f) = - \mathrm{Im}(G_{12}(f))/2\pi f\rho\Delta r \qquad (7\text{-}37)$$

Ensemble averaging is used in conjunction with Equation (7-37) to estimate the cross-spectral density. As discussed in Section 3.2.1.10, ensemble averaging is equivalent to time averaging in most practical applications. The total acoustic intensity over all frequencies is obtained by integrating (7-37) over all frequencies in the measurement bandwidth.

7.2.1.2 Errors

The main potential obstacle in applying Equation (7-37) to acoustic intensity measurements is to ensure that the phases of the instrumentation are matched between the two channels. Chung [7-7] has developed a technique which bypasses the need for phase calibration between the channels. Figure 7-5 shows the relevant instrumentation needed for a measurement. The i-th channel pressure signal P_i (i=1,2) is sensed by a microphone with a microphone sensitivity transfer function H_{Ti} ,and this signal is amplified by an amplifier with a transfer function H_{Ai}. Thus the total transfer function $H_i(f)$ associated with each channel is

$$H_i(f) = |H_i(f)|\, e^{j\phi_i(f)} = H_{Ai}(f)\, H_{Ti}(f) \qquad (7\text{-}38)$$

It follows that the cross-spectral density $G_{p_1p_2}(f)$ between the actual pressure signals is related to the measured $G_{12}(f)$ by

$$G_{p_1p_2}(f) = G_{12}(f)/[H_2(f)\, H_1^*(f)\,] \qquad (7\text{-}39)$$

Suppose now that the channels are switched and the measured cross-spectral density is denoted $G_{12}^{(s)}(f)$. The cross-spectral density function $G_{p_1p_2}(f)$ is given under these new test conditions by

$$G_{p_1p_2}(f) = G_{12}^{(s)}(f)/[H_1(f)\, H_2^*(f)] \qquad (7\text{-}40)$$

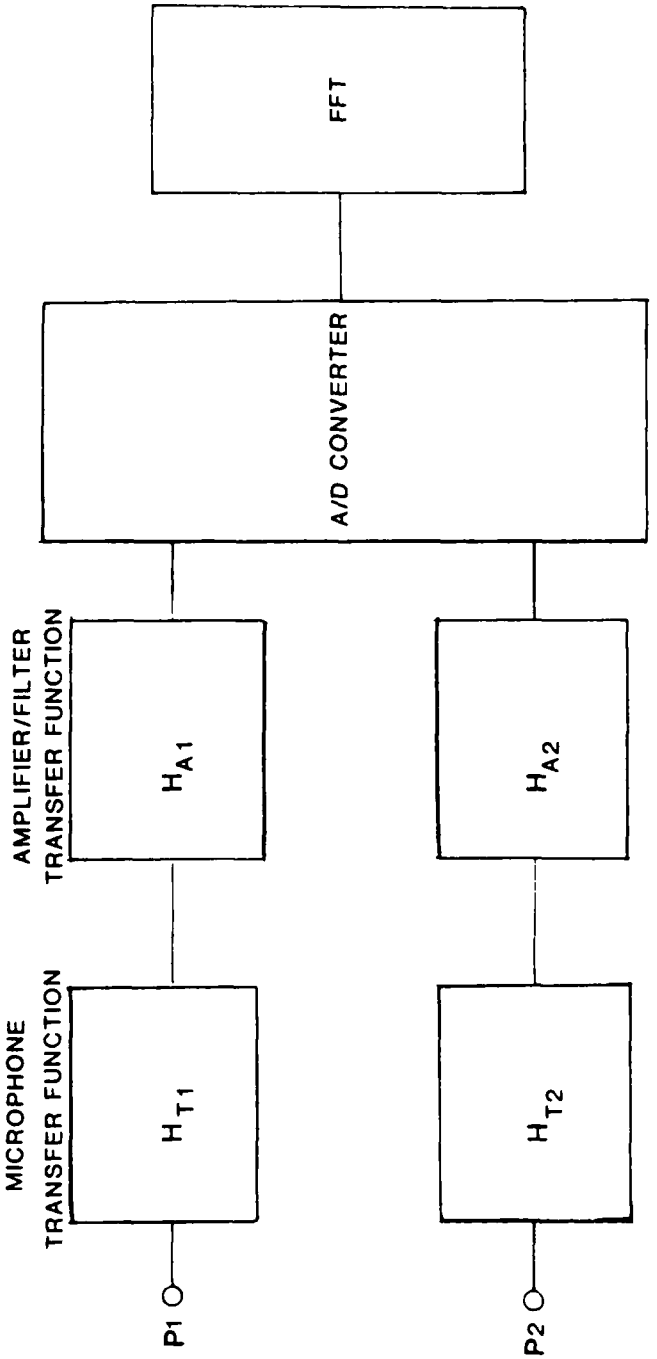


FIGURE 7-5 ACOUSTIC INTENSITY INSTRUMENTATION

Multiplying the above two equations and taking the square root of the resulting expression leads to

$$G_{p_1p_2}(f) = [G_{12}(f)\ G_{12}^{(s)}(f)]^{1/2}/|H_1(f)||H_2(f)| \qquad (7\text{-}41)$$

so that the acoustic intensity may be estimated by

$$I_r(f) = \mathrm{Im}\{[G_{12}(f)\ G_{12}^{(s)}(f)]^{1/2}\}/[2\pi f\rho\Delta r|H_1(f)||H_2(f)|] \qquad (7\text{-}42)$$

Thus the microphone switching technique requires two ensemble averages of a cross-spectral density function, that the sound field remain stationary over both measurements, and it requires taking the square root of a complex quantity--namely, the product of the cross-spectral densities in Equation (7-42). In addition, the magnitude of the instrumentation transfer function must be measured for each channel.

The switching technique requires interchanging all of the instrumentation associated with each channel, the microphone, pre-amplifier, and amplifier, in addition to any other possible filters and recording devices. If one uses a pair of microphones and pre-amplifiers which are phase-matched, then it is only necessary to switch the remaining instrumention.

This instrument switching techniques has the advantage of eliminating errors due to phase calibration uncertainties. Other sources of error remain. The random and bias errors in the estimation of the cross-spectral density functions are discussed in Section 3.3.3.7. The remaining causes of error are those due to limitations in the finite difference approximation (see Equation (7-34)) and microphone interaction effects.

Let I_0 denote the true value of the acoustic intensity at the point midway between the microphones in the direction of the vector pointing from microphone one to microphone two. Let I_e denote the estimate of the acoustic intensity obtained from

Equation (7-36) using the finite difference approximation. If the microphone one is located a distance r_1 from a simple source (an acoustic monopole) and microphone two is a distance r_2 from the source, then the error in decibels between the actual and estimated intensity levels is

$$L_e = 10 \log(I_0/I_e) = 10 \log \left[\frac{\sin(k\Delta r)}{k\Delta r} \frac{r^2}{r_1 r_2}\right] \qquad (7\text{-}43)$$

where $r = \frac{1}{2}(r_1+r_2)$, and the wave number $k=2\pi/\lambda$ for the acoustic wavelength λ. This error is negligible at low frequencies or small microphone spacings, and becomes appreciable for $k\Delta r > 1$. The error when this quantity equals unity is less than one decibel. The errors involved when the source is a dipole or quadropole radiator are discussed by Thompson and Tree [7-8] and a related discussion is given by Elliot [7-9].

Interaction effects between the microphones also cause errors in the acoustic intensity estimate due to scattering and diffraction effects as well as errors due to the finite aperture size of the microphones. These interaction effects are minimized by choosing the smallest possible microphones for the measurements, limited only that they should have sufficient sensitivity to produce an acceptably good signal-to-noise ratios (see Chapter 2). If there are phase errors in the instrumentation, then these phase errors limit the accuracy of the technique at low frequencies (when $k\Delta r<0.1$).

The presence of reverberation or nearby scattering objects can bias the estimate of acoustic intensity. If tests are made in a reverberant room, the intensity probe must be within the near-field of the source.

7.2.1.3 Tests to Validate Method

There are several tests which can be used to validate that the acoustic intensity technique is giving accurate results. The

simplest is a consistency check. The two-microphone intensity probe is used to estimate the acoustic intensity from a simple source. Next the microphones are rotated by 180° and the acoustic intensity is re-measured. The magnitude of the acoustic intensity estimates should remain unchanged, and the direction of energy flow should reverse. A more careful calibration of the directional characteristics is made by measuring the acoustic intensity component as a function of the angle of rotation of the intensity probe. When the vector from microphone one to microphone two points directly parallel to the energy flow in an acoustic plane wave, the acoustic intensity is I_0. When the intensity probe is rotated an angle θ from this axis the acoustic intensity $I(\theta)$ is

$$I(\theta) = I_0 \cos\theta \qquad (7\text{-}44)$$

Measurements of the acoustic intensity as a function of the rotation angle of the probe should agree with the predictions of Equation (7-44).

A more direct validation that the acoustic intensity meter is accurately measuring the desired quantity is to create a one-dimensional acoustic wave excitation. This can be done by using a plane wave excitation in a duct or using a small acoustic source in a free field environment. The acoustic intensity I_0 of such a plane or spherical wave excitation is known to be

$$I_0 = \langle p^2 \rangle / \rho c \qquad (7\text{-}45)$$

where the phase velocity is c. The acoustic intensity should be measured directly by a time-averaged measurement of the squared pressure using Equation (7-45) and compared to the estimate I_e of the acoustic intensity using the cross-spectral method (see

Equation (7-42). Typically, the acoustic intensity meter, including the probe, related electronics and data processing, are working properly if the intensity estimate is within .5 to 1 dB of the estimate based on the measured mean square pressure for the plane wave case.

$$10 \log \frac{I_e}{\langle p^2 \rangle / \rho c} \quad < \quad -10 \text{ dB}$$

The final way to validate the acoustic intensity method is to determine the total sound power of a source using either a source of known power output and compare this to the total sound power computed from the acoustic intensity estimate. The sound power W of a simple source is related to its acoustic intensity measured a distance r from the source by,

$$W = 4\pi r^2 I_0 \tag{7-46}$$

if the source is in free space and $r >> \lambda \frac{c}{f}$.

The acoustic intensity technique measures the acoustic intensity at a point midway between the microphones. Because of mutual interactions effects between different parts of a radiating structure it is sometimes useful to average the acoustic intensity over a small area in space. The advantage is that there is a higher repeatibility in the estimates obtained, at the expense of the fact that the spatial resolution is limited by the size of the scan area. With measurements at a point, the high spatial resolution is obtained at the risk of introducing possibly large errors due to errors in repositioning the acoustic intensity probe. In addition, the smaller the scan area the larger the number of samples which have to be taken. One must decide in the problem definition phase of the study or experiment what the acceptable limits are to the spatial resolution. Given those limits, one can define the largest possible scan area which meets the criteria.

7.2.1.4 Example

The acoustic intensity has been measured at points surrounding a diesel engine/chassis system with the transmission in drive and the engine running at 650 rpm [7-10]. Figure 7-6 shows five views of the engine/chassis system, and Figure 7-7 shows a block diagram of the acoustic intensity (in increments of 25 x 10^{-6} watts/m^2) on each of the five surfaces. The acoustic intensity probe used a pair of ¼" microphones spaced 11 mm apart, and the scan area was composed of a square of length 182 mm on each side. The frequency bandwidth used in the analysis was 500-4000 Hz. The peak intensities are located directly above and in front of the engine. Low acoustic intensity levels are found are the side views at the positions of the two tires. The total acoustic intensity through each surface can be computed by integrating the acoustic intensity over the surface. The largest component of acoustic intensity was through the top, which accounted for 40% of the total radiated power. By contrast, the smallest component was through the back which accounted for only 10% of the acoustic intensity.

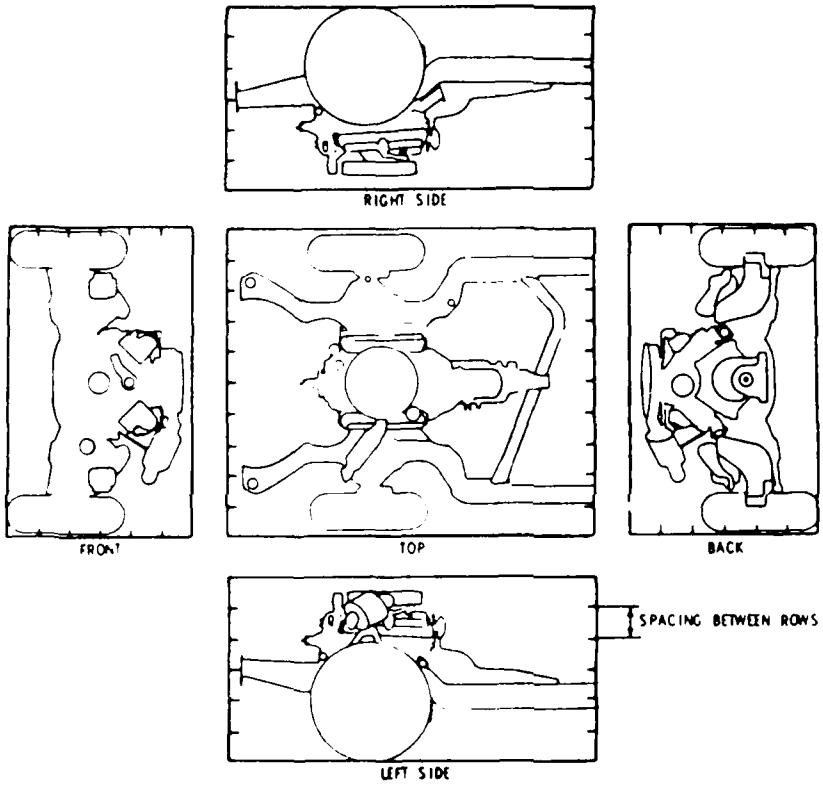


FIGURE 7-6 DIESEL ENGINE SURFACE VIEWS

Data from Ref. 7-10

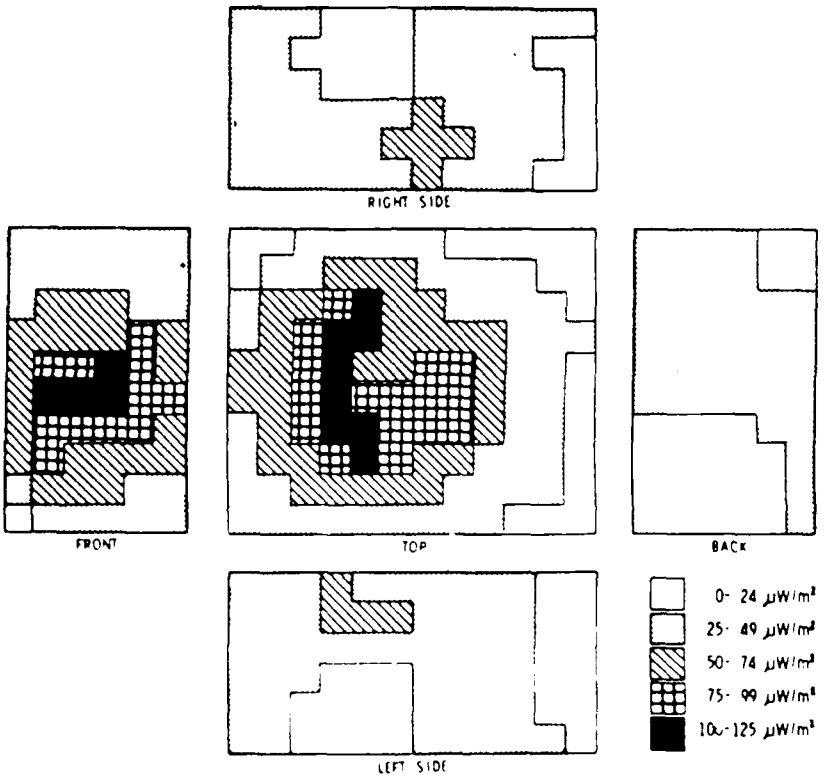


FIGURE 7-7 MEASURED ACOUSTIC INTENSITY OF ENGINE

Data from Ref. 7-10

7.2.2 Dominant Source Location

7.2.2.1 Sound-Ranging Technique

Measured data can be used to identify the location of dominant sources in an acoustic or other non-dispersive environment at any point. Two widely used methods of source location are based upon use of the cross-correlation function and the cross-spectral density function. Either of these functions are then interpreted to obtain information relating to propagation time delays between the source and receiver locations. For correlation functions this information is obtained directly in that the correlation has a peak at a time equal to the delay time between the receiver positions. This delay can be used to infer source location characteristics. For cross-spectral densities the propagation time delays are inferred from the phase characteristics of the cross-spectral density.

The application of these techniques to the location of dominant sources in an acoustic environment involves simultaneous measurements of pressure levels at two and possibly more locations in the acoustic field. The locations are not necessarily at the locations of sources. The pressure levels are analyzed to obtain the time delay between the signals at the two locations. From the time delay, the distance between the microphones, and the speed of sound, the direction of propagation of the sound with respect to the line connecting any pair of microphones can be inferred. A maximum time delay occurs when the line connecting the microphones points towards the source. A third simultaneous pressure level measurement at a different location enables determination of the source distance from the microphone locations.

The above process is simply a sound-ranging procedure. Its usefulness depends on being able to obtain accurate time delay information from the cross-correlation or cross-spectral density

functions. The existence of secondary noise sources can possibly obscure the identification of time delays, particularly if the noise signals are coherent at the pairs of microphone locations. The measurement techniques described in this section of the compendium are limited to nondispersive propagation paths. The effects of dispersion are discussed in Chapter 6.

7.2.2.2 Cross-Correlation Procedure

This section discusses the use of the cross-correlation function as a source identification tool. Methods for estimating the cross-correlation function from received time histories using digital signal processing are discussed in Chapter 3; the cross-correlation is defined Section 5.1.4 as an analog signal; and the use of the cross-correlation function in delay time estimation for nondispersive systems is discussed in Section 6.2.2.

Let the two time history signals be $p_1(t)$ and $p_2(t)$ as follows:

$$p_1(t) = s(t) + m(t) \tag{7-47a}$$

$$p_2(t) = h\, s(t-\tau) + n(t) \tag{7-47b}$$

where $s(t)$ is the source signal received at the first receiver location and $s(t-\tau)$ is the source signal at the second receiver which is delayed by a time τ relative to the first. The source signal at receiver two is scaled by an amplitude factor h relative to receiver one. There are noise signals $m(t)$ and $n(t)$ present at the two locations which are assumed to be incoherent with the source signal, but which may be coherent with one another.

The cross-correlation function $R_{12}(\tau)$ between the two measured time histories is found from Equation (5-17b) or computed from Equation (3-152) to be

$$R_{12}(t) = h\, R_{ss}(t-\tau) + R_{mn}(t) \qquad (7\text{-}48)$$

where the autocorrelation function of the source signal is $R_{ss}(t)$. The autocorrelation function is defined in Equation (5-17a) and may be calculated in terms of sampled time series using Equation (3-39). The cross-correlation function between the noise sources is $R_{mn}(t)$.

A general property of the autocorrelation function is that it attains its maximum value when its argument is equal to zero. Figure 3-11 illustrates this behavior when the source signal is white noise, pink noise, or narrow band filtered noise. This property also applies for sinusoidal signals. This means that the autocorrelation function $R_{ss}(t-\tau)$ is largest when $t=\tau$. If the noise signals are incoherent so that $R_{mn}(t)$ equals zero, then it follows from Equation (7-48) and the above mentioned property of the autocorrelation function that the cross-correlation function $R_{12}(t)$ has its maximum value at a time equal to the delay time between the two signals. This relation is the basis of the use of the cross-correlation function for time delay estimation.

It is also easy to see that the presence of partially coherent noise between the receiver locations can alter the position of the cross-correlation maximum and thus bias the time delay estimation. Further discussion of the problem of noise contamination is contained in Section 7.2.2.3.

The propagation delay time τ between the pair of receivers may be written in terms of the phase velocity c and the path-length d between the receivers as follows:

$$\tau = d/c \qquad (7\text{-}49)$$

This path length difference is illustrated in Figure 7-8, in which the two receivers are separated by L. Knowledge of this path length difference constrains the source to lie on a hyperbolic surface in space. Several of these surfaces are illustrated in Figure 7-9. When the receivers are located far away from the source so that the transmitted wave has a planar wavefront then the asymptotic angle of arrival θ with respect to the line joining the receivers is given by

$$\theta = \cos^{-1} (d/L) \qquad (7\text{-}50)$$

Several values of this angle of arrival are shown in Figure 7-9. The figure also shows that the source location lies at this angular orientation when the source is much farther from either receiver than the spacing L between the receivers.

EXAMPLE: An example of source location by through cross-correlation analysis which shows some of the practical problems involved and their solution is the location of sources underwater using pairs of hydrophones [7-11]. Similar practical problems are encountered in aeroacoustic applications. The source in this case was an underwater explosion 18 meters below the surface and 522 kilometers from the hydrophones. The hydrophone separation varied from 62 to 477 meters.

A typical pair of time histories of the received signals is illustrated in Figure 7-10. The signal is separated into three general regions. In region I of the first receiver signal (denoted hydrophone 2) the initial pressure pulse from the source is denoted by 1A. It is followed by two other peaks denoted 2A and 3A. These two secondary peaks are created by a train of bubble pulses which follows the high intensity pressure wave-front. Region II is a relatively low amplitude portion of the

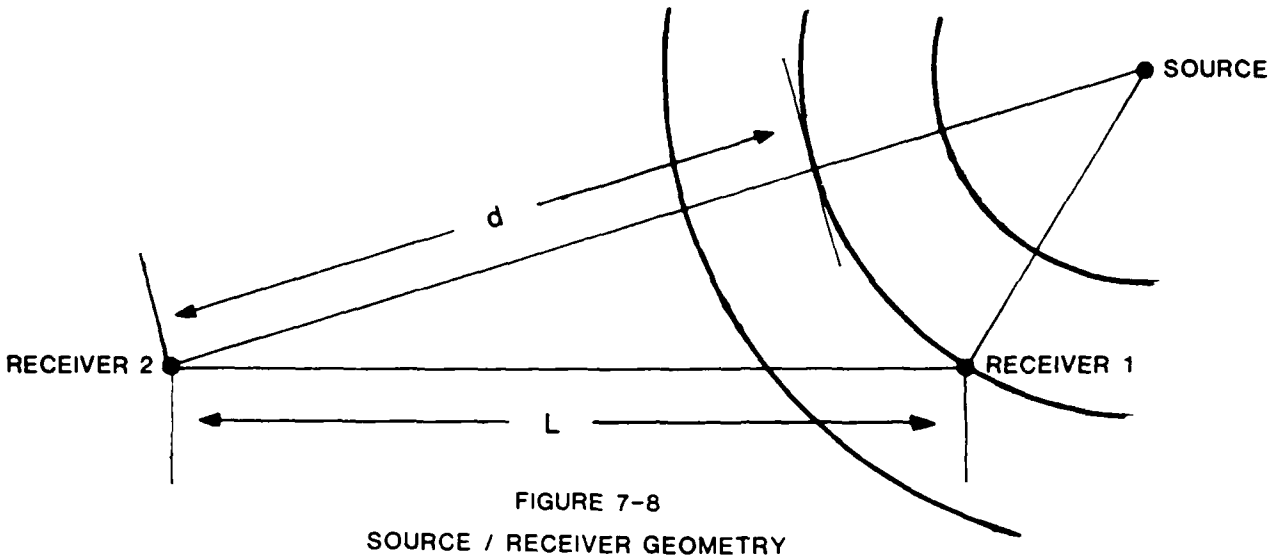


FIGURE 7-8
SOURCE / RECEIVER GEOMETRY

Data from Ref. 7-11

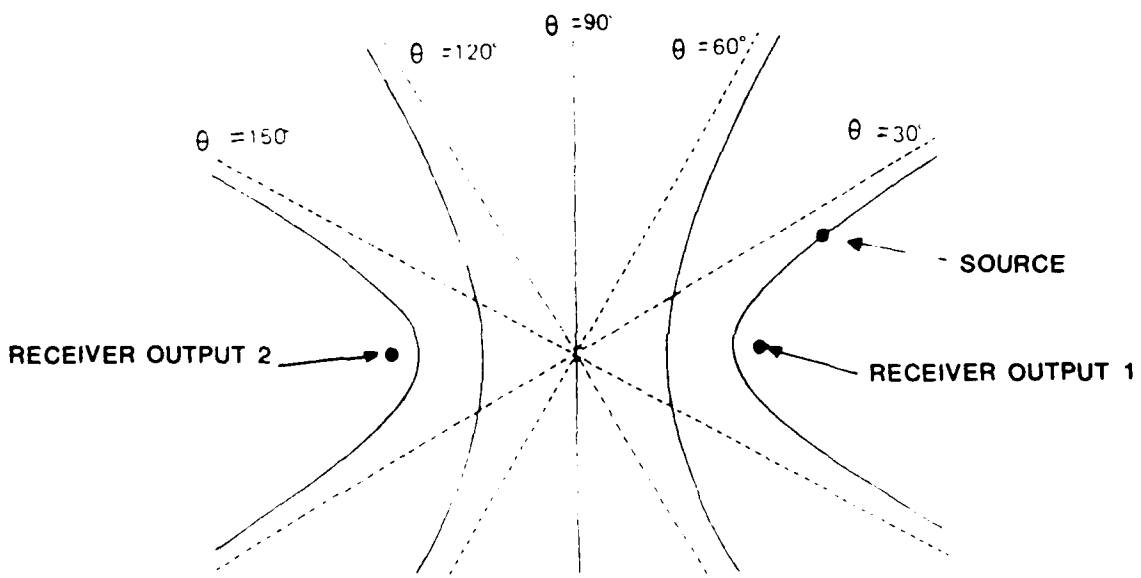


FIGURE 7-9 SOURCE LOCATION

Data from Ref. 7-11

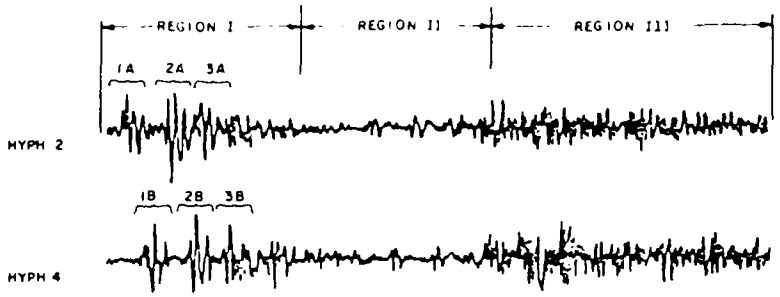


FIGURE 7-10 MEASURED TIME HISTORY

Data from Ref. 7-11

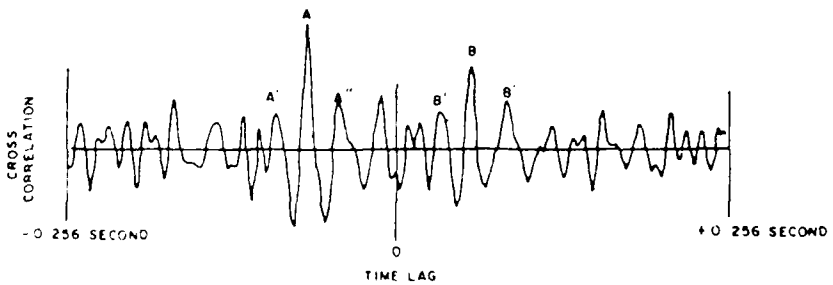


FIGURE 7-11 MEASURED CROSS CORRELATION

Data from Ref. 7-11

Figure 7-12 shows the cross-correlations associated with the six possible combinations of four receivers as a function of angle. The spacings between pairs of receivers are all different, and the receivers are arranged on a line. The 1-4 hydrophone combination in the figure is the same cross-correlation as in Figure 7-11, except that Equation (7-50) has been used to convert the delay time into the angle of incidence. The peak in the cross-correlation associated with the direct signal is clearly seen in each cross-correlation, and the portion of the signal due to the train of bubbles is strongly dependent on the receiver separation.

The average of the six cross-correlations is presented in Figure 7-13. The use of four receivers has made the cross-correlation peak strongly evident--the source is clearly located along a line oriented 26 degrees with respect to the line joining the receivers.

This example shows that the complicating factors of reverberation and complex source characteristics can diminish the usefulness of the cross-correlation source location techniques, unless adequate care is taken in the data collection and analysis phases. The effect of reverberation and multiple reflections can be reduced by suitable time-windowing of the individual time histories. In terms of this example, this was done by using a transient signal and focussing attention upon Region I of the time histories shown in Figure 7-10. The contaminating influence of temporal distortions in the signal, represented in the example by the bubble trains, was removed by using multiple receivers with different separation distances.

time history which includes a weak reflection of the signal from the surface of the water. Region III is the reverberant build-up of the underwater acoustic field due to multiple reflections from the rough bottom. Receiver two (denoted by hydrophone 4) has a similar time history. The initial pressure pulse is denoted by 1B, and the bubble pulses by 2B and 3B. The relative time difference between the hydrophone arrival times is evident in the figure.

Figure 7-11 shows the cross-correlation function of the two time histories. The time difference between the peak marked A is the propagation time difference between the two initial pressure pulses. The peak marked B corresponds to the bubble delay time 2A and 3A for receiver one or 2B and 3B for receiver 2. The secondary peaks at A' and A'' and their delayed counterparts B' and B'' are due to the spectral distribution of the source signal. It is difficult to conclude from Figure 7-11 whether energy is present from multiple angles since there are so many cross-correlation peaks.

The use of multiple pairs of receivers can eliminate this problem. If the spacing between the hydrophones is varied with the source excitation remaining at the same location, the observed delay times τ_1 and τ_2 for receiver spacings of L_1 and L_2 are related by

$$\tau_1/\tau_2 = L_1/L_2 \qquad (7\text{-}51)$$

This equivalence of ratios is apparent from Figure 7-8; an increase in L increases d by the same amount and hence increases the delay time. However, the time delay associated with the bubbles is not affected by the receiver separation. It is only a function of the propagation characteristics of the source; namely that a large amplitude wave motion induces bubble formation.

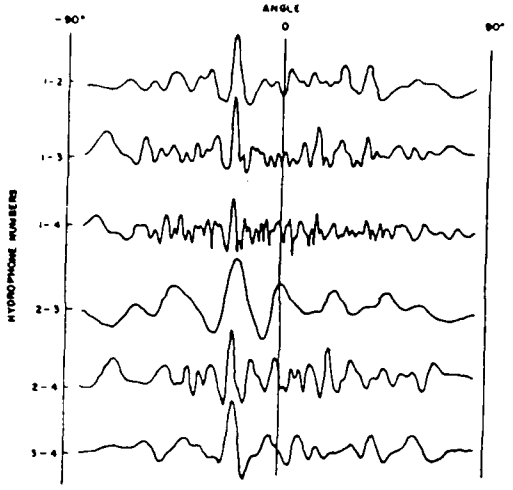


FIGURE 7-12

MEASURED CROSS CORRELATIONS FOR DIFFERENT HYDROPHONE PAIRS

Data from Ref. 7-11

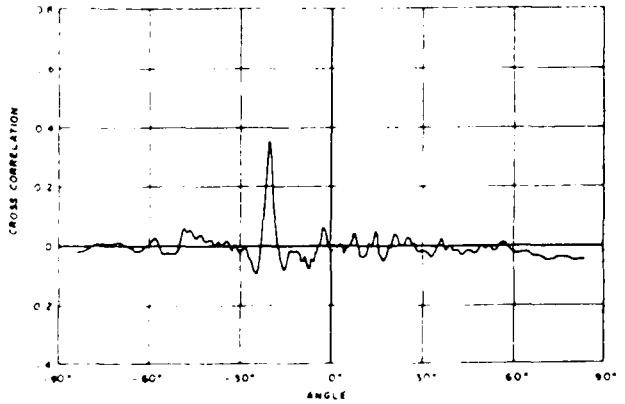


FIGURE 7-13 MEASURED CROSS CORRELATION vs. ANGLE

Data from Ref. 7-11

7.2.2.3 Cross-spectral Density Technique

PROCEDURE: The cross-spectral density function between the time series associated with two receivers may be used to compute source location. Let the two time signals be as given in Equation (7-47) of the previous subsection. Then the cross-spectral density $G_{12}(f)$ between the signals is given by

$$G_{12}(f) = h\, G_{ss}(f)\, e^{-j2\pi f\tau} + G_{mn}(f) \qquad (7\text{-}52)$$

This equation is the Fourier transform of Equation (7-48). The cross-spectral density function between the noise signal M(f) at receiver one at N(f) at receiver two is $G_{mn}(f)$. The amplitude of signal two relative to signal one is h. If the noise present at the two receivers is incoherent then the cross-spectral density is

$$G_{12}(f) = h\, G_{ss}(f)\, e^{-j2\pi f\tau} \qquad (7\text{-}53)$$

This equation shows that the propagation delay time τ may be computed in terms of the phase $\phi_{12}(f)$ of the cross-spectral density as follows:

$$\tau = -\phi_{12}/(2\pi f) \qquad (7\text{-}54a)$$

where

$$G_{12}(f) = |G_{12}(f)|\, e^{j\phi_{12}(f)} \qquad (7\text{-}54b)$$

Once the propagation delay time τ is computed, the source location evaluation process takes place as discussed in Section 7.2.2.1.

Methods for estimating the cross-spectral and autospectral densities are discussed in Section 3.3.3 and 3.2.3, respectively, along with consideration of the random and bias errors which arise in the estimation process.

EFFECT OF REVERBERATION: The estimate of propagation delay time is accurate in the presence of noise at the receiver locations as long as the two noise signals are incoherent so that $G_{mn}(f)$ equals zero. There are many cases where there are extraneous signals at the receiver which are regarded as noise but which are partially or even fully coherent. There are often situations in which the noise signal can be regarded as diffuse. This occurs when there is a reverberent background field at the receiver locations. This problem may be circumvented in the time domain by using an impulsive excitation whose duration is short relative to the reverberation time as was shown in the example discussed in Section 7.2.2.1.

There are also techniques for handling this problem in the frequency domain representation. Assume that the noise field in a volume enclosing both receivers is diffuse. Then the cross-spectral density $G_{mn}(f)$ between the diffuse portion of the signals at the two receivers is given by

$$G_{mn}(f) = G_{nn}(f) \frac{\sin kL}{kl} \qquad (7\text{-}55)$$

where L is the distance between the receivers, the wave number k is the ratio of the radian frequency $\omega=2\pi f$ and the nondispersive phase speed c, and $G_{nn}(f)$ is the spatial average of the autospectral density function of the noise field within the diffuse environment. One can then regard $G_{mn}(f)$ as a known quantity and substitute the relation in Equation (7-55) into Equation (7-52) to obtain an estimate of the delay time.

Unfortunately, if the diffuse field is caused by reverberant build-up of the signal of interest, then this approach can lead to sizeable errors. A better alternative is to use the dereverberation techniques discussed in Chapter 6 for nondispersive systems in order to eliminate the contaminating effect of reverberation.

REFERENCES

7-1. Bendat, J.S. and Piersol, A.G., Random Data: Analysis and Measurement Procedures, Wiley-Interscience, New York, 1971.

7-2. Bendat, J.S. and Piersol, A.G., Engineering Applications of Correlation and Spectral Analysis, Wiley, New York, 1980.

7-3. Bendat, J.S., "Modern analysis procedures for multiple input/output problems", J. Acoust. Soc. Am. 68, 498-503, 1980.

7-4. Barrett, S., "On the use of coherence functions to evaluate sources of dynamic excitation", Shock and Vibration Bull. 49, 43-58, 1979.

7-5. Bendat, J.S. and Piersol, A.G., 1980, p. 232.

7-6. Fahy, F.J., "Measurement of Acoustic Intensity Using the Cross-Spectral Density of Two Microphone Signals", J. Acoust. Soc. Am. 62, p. 1057-1059, 1977.

7-7. Chung, J.Y., J. Acoust. Soc. Am. 64, p. 1613-1616, 1978.

7-8. Thompson, J.K. and Tree, D.R., "Finite Difference Approximation Errors in Acoustic Intensity Measurements", J. Sound and Vib. 75, p. 229-238, 1981.

7-9. Elliot, S.J., "Errors in Acoustic Intensity Measurements", J. Sound and Vib. 78, p. 439-445, 1981.

7-10. J. Pope, R. Hickling, D.A. Feldmaier, and D.A. Blaser, "The Use of Acoustic-Intensity Scans for Sound Power Measurement and for Noise Source Identification in Surface Transportation Vehicles", Proceedings of the International Congress and Exposition, Society of Automotive Engineers, Paper No. 810401, 1981.

7-11. Young, H.J., "Underwater sound arrival estimation by multiple cross-correlation measurements", 1978 IEEE International Conference on Acoustics, Speech & Signal Processing, IEEE 78CH1285-6 ASSP, 1978, p. 659 564.

7-12. Bellavita, P., and Smullin, J.,"Cabin Noise Reduction for the Agusta A-109 Helicopter", Paper No. 61, Fourth European Rotorcraft and Powered Lift Aircraft Forum, September, 1978, Stresa, Italy.

CHAPTER 8. VIBRATION RESPONSE PREDICTION

By

R. G. DEJONG

CHAPTER 8. VIBRATION RESPONSE PREDICTION

TABLE OF CONTENTS

		Page
	LIST OF FIGURES	8-iii
	LIST OF TABLES	8-iv
	LIST OF SYMBOLS	8-v
8.1	INTRODUCTION	8-1
8.2	PROPAGATING PROGRESSIVE WAVE MODELS	8-3
	8.2.1 System Model	8-3
	8.2.2 Excitation and Response Model	8-7
8.3	TRANSMISSION LINE ANALYSIS	8-10
8.4	MODAL ANALYSIS	8-16
	8.4.1 Normal Mode Analysis	8-16
	8.4.2 Evaluating Modal Parameters	8-18
8.5	STATISTICAL ENERGY METHOD	8-23
	8.5.1 System Model	8-23
	8.5.2 Excitation Model	8-25
	8.5.3 Coupling Between Subsystems	8-27
8.6	DISCRETE ELEMENT MODEL	8-32
	8.6.1 Lumped Parameter Models	8-34
	8.6.2 Finite Element Model	8-34
	8.6.3 Excitation and Response Model	8-37

8.7 TRANSFER FUNCTION MODELS..... 8-40

8.7.1 System Model..... 8-40
8.7.2 Mobility Equations..... 8-42
8.7.3 Cascading Elements..... 8-44

8.8 EXTRAPOLATION TECHNIQUES..... 8-48

8.9 COMPARISON OF METHODS..... 8-53

REFERENCES..... 8-57

LIST OF FIGURES

8-1. Propagating Wave Solution..... 8-6

8-2. Impedance of Rod-Transmission Line Analysis..... 8-11

8-3. Comparison of Measured and Predicted Transfer Impedance of Isolation Mount..... 8-13

8-4. Two Mode Curve Fit in Modal Analysis..... 8-20

8-5. Comparison of the SEA Predictions with 1/4 Scale and Full Scale Measurements of the Acceleration Response of the Gear Housing..... 8-30

8-6. Lumped-Parameter Model of the Crank Shaft..... 8-35

8-7. Imput, Output model of Individual Components..... 8-41

8-8. Comparisons of Calculated and Measured Transfer Functions in a Diesel Engine..... 8-46

8-9. Scattered Diagram of Acceleration Levels L on B-58 Airplane Plotted Against Pressure Levels L_p in 600 to 1200 Hz Octave Band..... 8-51

8-10a-c. Test Structure..... 8-54

LIST OF TABLES

		Page
8.1	Summary of Procedures for the Propagating Wave Solution Method	8-9
8.2	Summary of the Procedures for the Transmission Line Analysis Method	8-15
8.3	Summary of the Procedures for the Modal Analysis Method	8-22
8.4	Summary of the Procedures for the Statistical Energy Analysis Method	8-31
8.5	Summary of the Procedures for the Finite Element (Modal Superposition) Method	8-39
8.6	Summary of the Procedures for the Transfer Function Method	8-47
8.7	Summary of the Procedures for the Extrapolation Method	8-52

LIST OF SYMBOLS

[A]	Coupling matrix
A_{ij}	Element in coupling matrix
A,B	Arbitrary constants
[B]	System matrix
$C(\omega)$	Wave speed
[C]	Viscous damping matrix
c	Wave speed
DR	Decay rate (dB/sec)
E	Elastic (Young's) modulus
e	Exponential function
F	Excitation force
{F}	Vector of external nodal loads
$\langle F^2 \rangle$	Mean square amplitude of excitation force
$F(\omega)$	Amplitude of excitation force at ω
f	Frequency in hertz
G	Real part of mobility
G_i	Real part of Y_i
$H_{f/v}$	Response due to unit V or transfer function
I_{12}	Transmission factor
[K]	Stiffness matrix
k	Wave number
L	Length
M	Mass
[M]	Nodal mass matrix
m	Subsystem mass
N	number of equilibrium equations
$n(\omega)$	Modal density
P	Excitation load

Q	Quality factor, measure of sharpness of a modal response
R	Magnitude of reflection coefficient
R	Real part of impedance
S	Area
$T_{v_1 F_2}$	Transmission matrix element
t	Time
$V(\omega)$	Amplitude of system velocity response at ω
V_i	Velocity at point i
$\langle V^2 \rangle$	Mean square amplitude of vibration velocity response
W_{in}	Input power from source
X	Vector of nodal responses
x	Spatial location
Y	Wave amplitude
$[Y]$	Mobility matrix
Y_i	Mobility of subsystem i
Y_{ij}	Entry in mobility matrix
$Y(x)$	Wave amplitude response function
y	Amplitude response
$Z(\omega)$	Characteristic impedance of system
Z_{11}	Drive point impedance
Z_n	Characteristic modal impedance
$\bar{G}$	Average of the real part of point mobility
[8-]	Reference
ρ	Density
Λ	Spatial derivative operator
∂	Partial differential operator
η	Loss factor
ζ	Critical damping ratio
γ	Spatial propagation constant
ϕ	Phase of reflection coefficient

$\psi_n(x)$	Mode shape
ω_n	Characteristic frequency
δ_{mn}	Delta function (=1 for m=n; 0 for m ≠n)
η_d	Damping loss factor
Πin	Input power
Π_{diss}	Dissipated power
ε	Average subsystem modal energy
$\Pi_{1;2}$	Net transmitted power from subsystem 1 to subsystem 2

CHAPTER 8. VIBRATION RESPONSE PREDICTION

8.1 INTRODUCTION

This chapter is concerned with how measured acoustic and vibration data can be used with system models to predict the vibration response of a structure. The process of modeling the system is an important step in moving from simply observing how the system responds to understanding why the system responds the way it does and how to improve it. For engineering applications involving design modifications the more useful modeling techniques are those which use physically realizable parameters rather than merely mathematical ones.

Some caution, however, must be used to avoid too much reliance on system characteristic models to solve all the problems. Theoretical investigations of signal processing methods show seemingly unlimited possibilities while practical applications fall far short of their goals. Part of this discrepancy is the result of signal noise in the measured data used as inputs to the models - the impact of which is only recently being understood in theoretical terms. As a result there continues to be, and rightfully so, as much dependence on "previous experience" as on any new analysis tool.

Two factors contribute to a continuing improvement in the practical usefulness of system modeling. The first factor is simply the continued development of digital processing systems which enable detailed correlation, coherence and frequency response analyses of multi-input, multi-output systems to be implemented in practice. This is covered in detail in Chapter 3. The second factor is the continued development of useful algorithms for setting up and solving the system of equations used in modal analysis, finite element and statistical energy analysis.

It is helpful to distinguish between two different types of system models. Ones that have parameters directly relatable to physical properties of the structure (dimensions, materials, shape, etc.) are called "physical." Ones that provide an efficient and concise mathematical representation of the data are called "mathematical." Physical models include modal analysis, propagating waves, statistical energy analysis, and some discrete element models. Parametric models include transfer function analysis, and some empirical methods. Physical models are based on an analytical solution of the wave equation of the system or on a numerical solution of a discretized approximation of the wave equation. Mathematical models are based on fitting measured data to a set of equations in order to solve for the unknown parameters. The purpose of this chapter is to outline the basic elements of these various prediction techniques, providing references to more detailed descriptions. Section 8.9 presents a comparison of the different techniques.

For the purposes of this chapter, the wave equation for a structure will be written as,

$$\rho \frac{\partial^2 y(x,t)}{\partial t^2} + E(1+j\eta)\ \Lambda\ y(x,t) = p(x,t)\ , \qquad (8\text{-}1)$$

where $y(x,t)$ is the response of the system to the excitation $p(x,t)$ both being functions of the system spatial coordinates x and time t. Also, ρ is the density, E is the elastic modulus and η is the damping loss factor of the system material ($\eta/2=\zeta$, where ζ is the critical damping ratio), Λ is the spatial derivative operator which is dependent on the type of wave motion being studied, and $j=\sqrt{-1}$.

The damping is included in this manner for convenience only. A damping term proportional to the first time derivative

of y could be included to model, for example, viscous damping. However, this unnecessarily complicates the mathematics. Instead, the value of η is adjusted to represent the level of all forms of damping. This is not a problem as long as the loss factor is small ($\eta<0.2$).

8.2 PROPAGATING PROGRESSIVE WAVE MODELS

The propagating wave analysis is used for systems that are sufficiently large or with sufficiently high damping that the reverberant response of the structure is not important. It is most frequently applied to transient or shock applications.

The wave Equation (8-1) is solved as if the system were infinite in extent. Then the source of excitation causes a wave of vibration to propagate through the system. A model is developed to calculate the response of the system to this traveling wave and possibly a number of reflections off boundaries or discontinuities.

8.2.1 System Model

The freely propagating wave is found by setting $p(x,t)=0$ in Equation (8-1) and assuming a sinusoidal motion at frequency $\omega=2\pi f$ of the form:

$$y = Y(x)\ e^{j\omega t} = Y(x)\ \{\cos(\omega t)+j\sin(\omega t)\}.$$

Then Equation (8-1) simplifies to

$$\{-\rho\omega^2 + E(1+j\eta)\ \Lambda\}\ Y(x) = 0. \qquad (8\text{-}2)$$

The solution of equation (8-2) depends on the form of the spatial derivative Λ. For example, with the <u>one dimensional compression wave</u> $\Lambda = \frac{\partial^2}{\partial x^2}$. The general form of the solution is then

$$Y(x) = Ae^{-j\gamma x} + Be^{j\gamma x} \qquad (8\text{-}3)$$

which represents one wave traveling in each direction, with A and B arbitary constants and γ the spatial propagation constant.

Substituting Equation (8-3) into Equation (8-2) gives

$$\gamma^2 = \frac{\omega^2\rho}{E(1+j\eta)} \qquad (8\text{-}4)$$

Further simplification is obtained by noting the wave speed $c=\sqrt{E/\rho}$ and the wave number $k= \omega/c$ (which is a measure of the spatial wave density). Then

$$\gamma \approx k(1-j\eta/2) \quad \text{for } \eta<0.2 \qquad (8\text{-}5)$$

Consider a single compression wave traveling in the +x direction. It must have the form,

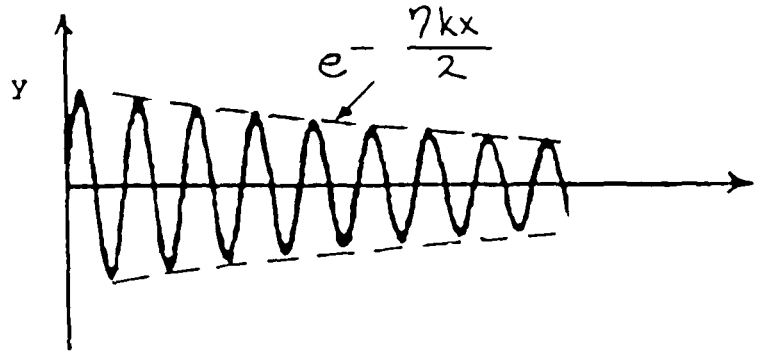
$$y(x,t) = A\, e^{j\omega t}\, e^{-\frac{\eta}{2}kx}\, e^{-jkx} \qquad (8\text{-}6)$$

This is shown in Figure 8-1a for a particular time as a function of x. The term $e^{-\eta kx/2}$ represents the decay of the wave as it propagates in the +x direction.

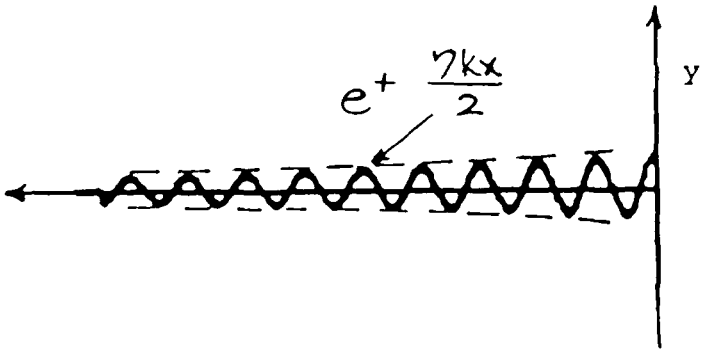
Next, consider in addition a reflected wave of amplitude $B=Re^{j\phi}A$ travelling in the -x direction, where R and ϕ are the magnitude and phase, respectively, of the reflection coefficient. The system response is then given by:

$$y(x,t) = A\, e^{j\omega t}\, [e^{-\frac{\eta}{2}kx}\, e^{-jkx} + R\, e^{j\phi}\, e^{\frac{\eta}{2}kx}\, e^{jkx}] \qquad (8\text{-}7)$$

Figure 8-1b shows the shape of the reflected wave. Note that the amplitude of this wave is growing exponentially in the +x direction and, therefore, must have a position of origin a finite distance away.



a.) Traveling Wave



b). Reflected Wave

Figure 8-1. Propagating Wave Solution

8.2.2 Excitation and Response Model

The propagating wave analysis can model excitations in the form of either imposed motions, or applied forces (or pressures or stresses). In principle, these excitations can be distributed spatially, but most frequently only localized excitations are considered. With localized excitations the freely propagating waves derived in the previous section are used to match the response of the system to the response at the points of excitation, and then are used to give the response of the rest of the system where no excitations are applied.

For the case of imposed motions, a propagating wave in the form of Equation (8-6) is set up with its amplitude fixed at the location of the excitation.

For the case of applied forces, the response is determined by the characteristic impedance $Z(\omega)$ defined by

$$Z(\omega) = \frac{F(\omega)}{V(\omega)} \tag{8-8}$$

where $F(\omega)$ is the amplitude of the excitation force and $V(\omega)$ is the amplitude of the system velocity response at frequency ω. When considering sinusoidal motions of the form $e^{j\omega t}$, the velocity is given by $V = j\omega Y$. Then the wave amplitude Y is given by

$$Y = \frac{F}{j\omega Z} \tag{8-9}$$

at the location of the force excitation F. The value of the impedance depends on the particular system being studied.

For example, with the one dimensional compression wave, the motion y, at position x, is related to a pressure p, by

$$p = -E(1+j\eta)\frac{\partial y}{\partial x} \tag{8-10}$$

With a pressure acting over a surface area S, the excitation force F is given by pS. Combining this with Equation (8-7) gives

$$H_{f_1|v_1} = Z_{11} = \rho c\, S \sqrt{1+j\eta}\left[\frac{1 - Re^{j\phi}}{1 + Re^{j\phi}}\right] \tag{8-11}$$

Table 8-1 summarizes the procedures used in the propagating wave method.

For a more detailed description of the propagating wave solution and examples of its application see Reference [8-1,2].

Table 8-1 Summary of Procedures for the Propagating Wave Solution Method.

A. Inputs Required
   1. Idealized geometric model of system.
   2. Dynamic properties of waves in system model (e.g. characteristic impedance Z, wavespeed c, wave number k, damping loss factor η).

B. Steps of Solution
   1. Identify wave type to be used in solution.
   2. Identify wave equation for chosen wave type.
   3. Solve wave equation for freely propagating system response.
   4. Match response amplitude to the amplitude imposed at the excitation location.
   5. Solve for the system response at other locations.

C. Outputs Obtained
   1. System response (amplitude and phase) to a particular excitation as a function of position.
   2. System transfer function (response per unit excitation).

8.3 TRANSMISSION LINE ANALYSIS

An important extension to the propagation wave analysis of dynamic systems is the transmission line analysis. This method is generally applied to one-dimensional propagation problems where the system being analyzed can be modeled as a series of one-dimensional elements each having uniform properties. The propagating wave solution in the form of Equation (8-11) is applied to each element, to represent an excitation at one end and reflection off the other end. This results in a complete solution for the standing wave (reverberant field) in each element.

Then by combining these solutions with the appropriate boundary conditions at each connection between elements, the characteristic transfer function for the system can be found.

For example, the one-dimensional compression waves in a rod of length L would have a reflection characteristic at a free end of $R = +1$ and $\phi = 2k(1-j\frac{\eta}{2})L$. Substituting these values into Equations (8-11) gives

$$Z_{11} = j\rho c\, S \sqrt{1+j\eta}\, \tan\{k(1-j\tfrac{\eta}{2})L\} \qquad (8\text{-}12)$$

This function is plotted in Figure 8-2.

A model can be developed of a series of rods (of arbitrary physical parameters) connected together by setting the boundary conditions at each junction to be consistent with the type of connection. For rigid connections, the velocities of two elements at a common junction are set equal while the forces generated at the junction are set equal in amplitude but opposite in sign.

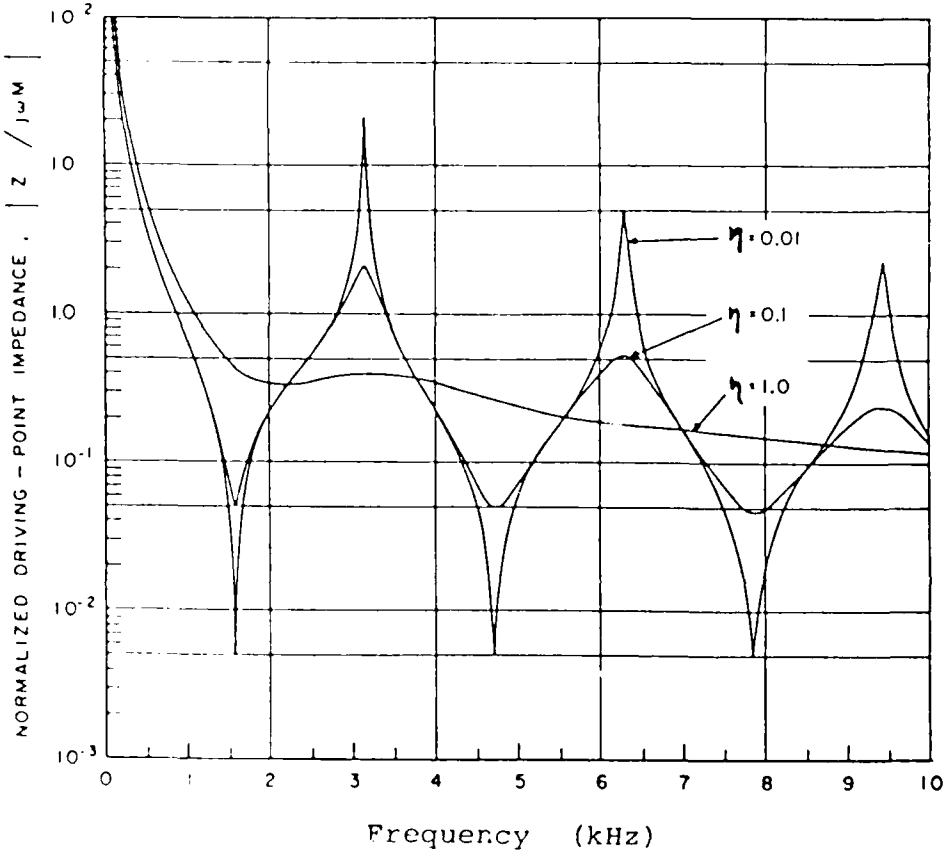


Figure 8-2. Impedance of Rod-Transmission Line Analysis [8-1]

Assembling this type of model is made easier by using the transmission matrix equation

$$\begin{matrix} V_1 \\ F_1 \end{matrix} = \begin{matrix} T_{V_1V_2} & -T_{V_1F_2} \\ T_{F_1V_2} & -T_{F_1F_2} \end{matrix} \begin{matrix} V_2 \\ -F_2 \end{matrix} \qquad (8\text{-}13)$$

where the elements of transfer functions matrix T are related to the impedances [8-1]. Then if point 2 of element "a" is rigidly attached to point 1 of element "b", the conditions

$$\begin{matrix} V_2 \\ -F_2 \end{matrix}_a = \begin{matrix} V_1 \\ F_1 \end{matrix}_b \qquad (8\text{-}14)$$

result in a new transmission matrix

$$\begin{matrix} V_1 \\ F_1 \end{matrix}_a = T_a \; T_b \begin{matrix} V_2 \\ -F_2 \end{matrix}_b \qquad (8\text{-}15)$$

Figure 8-3 shows an example of the applications of this type of a compound model to a rubber isolation mount. A comparison of the measured and predicted vibration transfer function of the mount shows that while the general nature of this frequency response function is reproduced in the model, some of the details are missing in the predictions. This is because the model only includes one type of wave propagation (one-dimensional, compression waves) while there are really several wave types present. These additional wave types could be included in the model but the computations become considerably more complex.

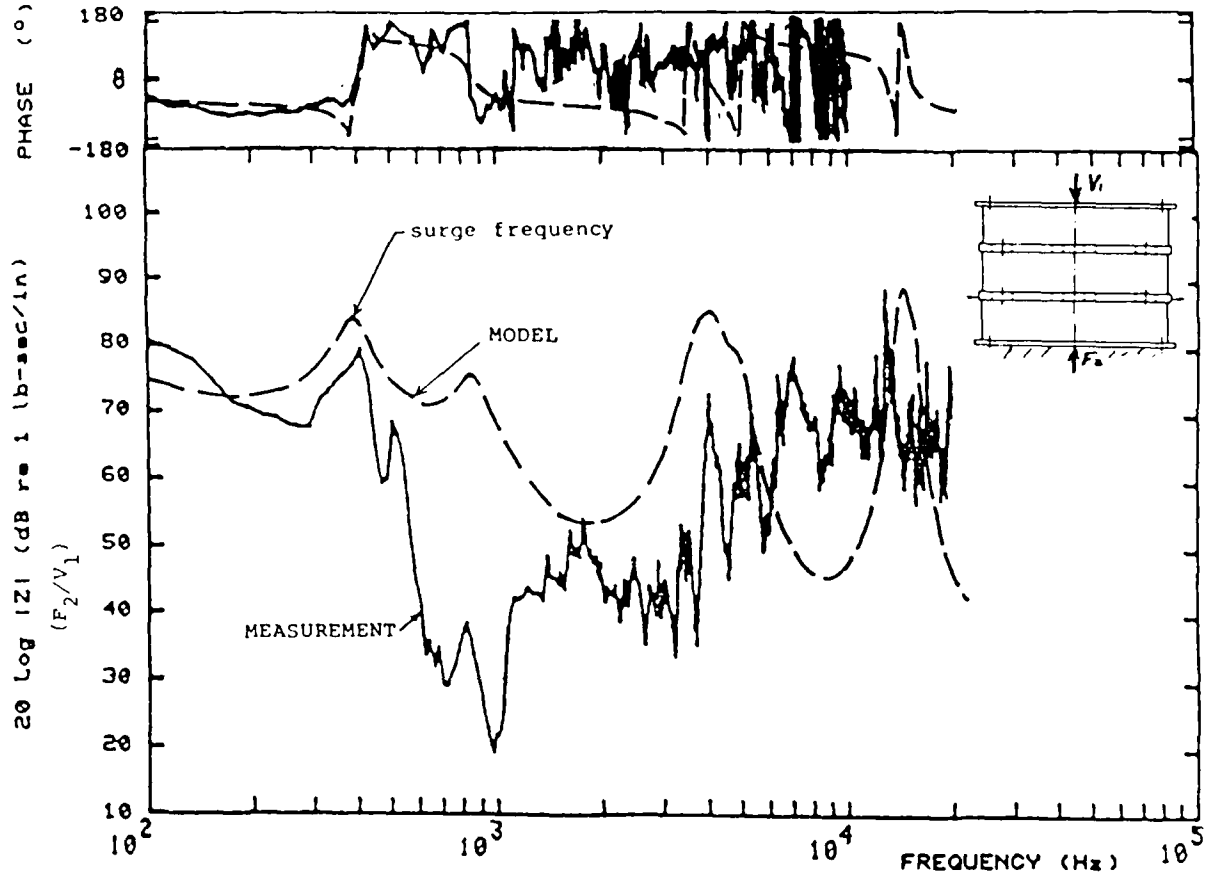


Figure 8-3. Comparison of Measured and Predicted Transfer Impedance of Isolation Mount

Table 8-2 summarizes the procedures used in the transmission line analysis method.

More information on this transmission line analysis can be found in [8-1,2,3].

Table 8.2 Summary of Procedures for the Transmission Line Analysis Method

A. Inputs Required
   1. Geometric model of system including a set of interconnected idealized subsystems.
   2. Dynamic properties of waves in each subsystem (e.g. characteristic impedance Z, wavespeed c, wave number k, damping loss factor η).
   3. Idealized model of the excitation as localized forces or imposed motions at the subsystem junctions.

B. Steps of Solution
   1. Identify wave type to be used in each subsystem.
   2. Solve each subsystem wave equation to obtain the elements in the transmission matrix.
   3. Set up complete system transmission matrix based on interconnections of subsystems.
   4. Solve matrix equation for system transfer functions.
   5. Solve for the system response at subsystem junctions by multiplying the system transfer function by the excitations.

C. Outputs Obtained
   1. System response (amplitude and phase) at subsystem junction points to a particular excitation at subsystem junction points.
   2. System transfer functions (response at juntions per unit excitation).

8.4. MODAL ANALYSIS

8.4.1 Normal Mode Analysis

An alternate solution to the propagating wave solution of the wave equation is the normal mode solution. Temporarily taking the form of Equation (8-2) without damping or excitation gives

$$(-\rho\omega^2 + E\,\Lambda\,)\;Y(x) = 0 \qquad (8\text{-}16)$$

A solution is formed by expanding Y(x) in a set of normal modes (mode shapes)

$$Y(x) = \sum_{n=1}^{\infty} Y_n\;\Psi_n(x) \qquad (8\text{-}17)$$

where Y_n is the amplitude associated with each mode. With the appropriate boundary conditions applied to Equation (8-16), a set of $\Psi_n(x)$ are found for a particular set of eigenvalues ω_n (resonant frequencies).

An additional constraint of orthogonality is placed on the modes so that the following condition is met:

$$\frac{1}{M}\int \Psi_m(x)\;\rho\;\Psi_n(x)\;dx = \delta_{m,n} \qquad (8\text{-}18)$$

where $\delta_{m,n} = 1$ if $m = n$

and $\delta_{m,n} = 0$ if $m \neq n$

The excitation F(x) is also expanded in the set of mode shapes:

$$F(x) = \sum_{n=1}^{\infty} F_n \Psi_n(x) \tag{8-19}$$

Substituting these expansions into the wave Equation (8-1) gives

$$\sum_{n=1}^{\infty} \{-\rho\omega^2 + \rho\omega_n^2(1+j\eta)\} Y_n \Psi_n(x) = \sum_{n=1}^{\infty} F_n \Psi_n(x) \tag{8-20}$$

Since the mode shapes are orthogonal, Equation (8-20) must hold for each value of n independently.

Therefore

$$\{\omega_n^2(1+j\eta) - \omega^2\} Y_n = \frac{1}{\rho} F_n \tag{8-21}$$

The characteristic modal impedance is given by

$$Z_n = \frac{F_n}{j\omega Y_n} = \frac{\rho}{j\omega} \{\omega_n^2(1+j\eta) - \omega^2\} \tag{8-22}$$

The modal excitations due to applied forces are given by

$$F_n = \int F(x) \Psi_n(x) \, dx \tag{8-23}$$

When the system is excited by an imposed motion y_o, the modal excitation for the response in the coordinate system $w(x) = y(x) - y_o$ is given by [8-4]

$$F_n = -\rho \frac{d^2 y_o}{dt^2} \int \Psi_n(x)\, dx \qquad (8\text{-}24)$$

8.4.2 Evaluating Modal Parameters

The modal parameters can be obtained in one of several methods. For simple systems, closed form analytical solutions can be found using known mathematical functions (such as sine and cosine) for the mode shapes. These solutions can be used as an alternative method for deriving transmission line models as described in the previous sections.

For complex systems, the discrete element models can be used to evaluate the approximate mode shapes and resonant frequencies. These methods are described in Section 8-6.

In this section, the experimental technique known as modal analysis is described. This method uses measured values of the frequency response functions of the system at a finite set of points to evaluate the modal parameters.

The complete response of a system at position x_r is given by

$$Y(x_r) = \sum_{n=1}^{\infty} \frac{j\omega\, F_n\, \Psi_n(x_r)}{\rho\{\omega_n^2(1+j\eta) - \omega^2\}} \qquad (8\text{-}25)$$

For a point force excitation of amplitude F_o at position x_s, $F_n = F_o\, \Psi_n(x_s)\, \rho/M$ where M is the mass of the system. Then the transfer function between the response at x_r and the excitation at x_s is given by

$$H_{y_r f_s} = \frac{Y(x_r)}{F_o} = \frac{j\omega}{M} \sum_{n=1}^{\infty} \frac{\Psi_n(x_r)\ \Psi_n(x_s)}{\omega_n^2(1+j\eta) - \omega^2} \qquad (8\text{-}26)$$

Two methods can be used to experimentally determine the values of $\Psi_n(x)$ for a structure. One is known as the sine-dwell method which excites the structure at multiple points with pure tone force excitation at a frequency near a known resonant frequency. Then by adjusting the frequency of the excitation and monitoring the relative phases between the forces and several response points, the exact resonant frequency can be found and the response is assumed to be dominated by a single mode. Then the mode shape can be measured directly by a scan of the vibration response of the structure.

While the sine-dwell method has been the choice of many people in the past, it is gradually being replaced by a second method because of the high labor intensive effort required for the sine-dwell method.

The second method is most frequently referred to as "modal analysis" and involves curve fitting the frequency response functions (frf) measured at a number of points on the structure. By measuring the drive point frf $H_{y_1 f_1}$ at position x_1, the values of the $\Psi_n(x_1)$ can be obtained along with the associated values of ω_n and η, as long as they can be accurately distinguished in the measurement. This is done by using a curve fitting routine (such as least squares) to match the modal response equation to the measured data in a frequency range around each resonant frequency. The simplest of these routines fit a single mode at a time, while more complex routines are available to fit multiple modal responses simultaneously. The multi-mode curve fits do a better job where modes overlap in frequency and therefore are difficult to handle separately. An example of this is shown in Figures 8-4.

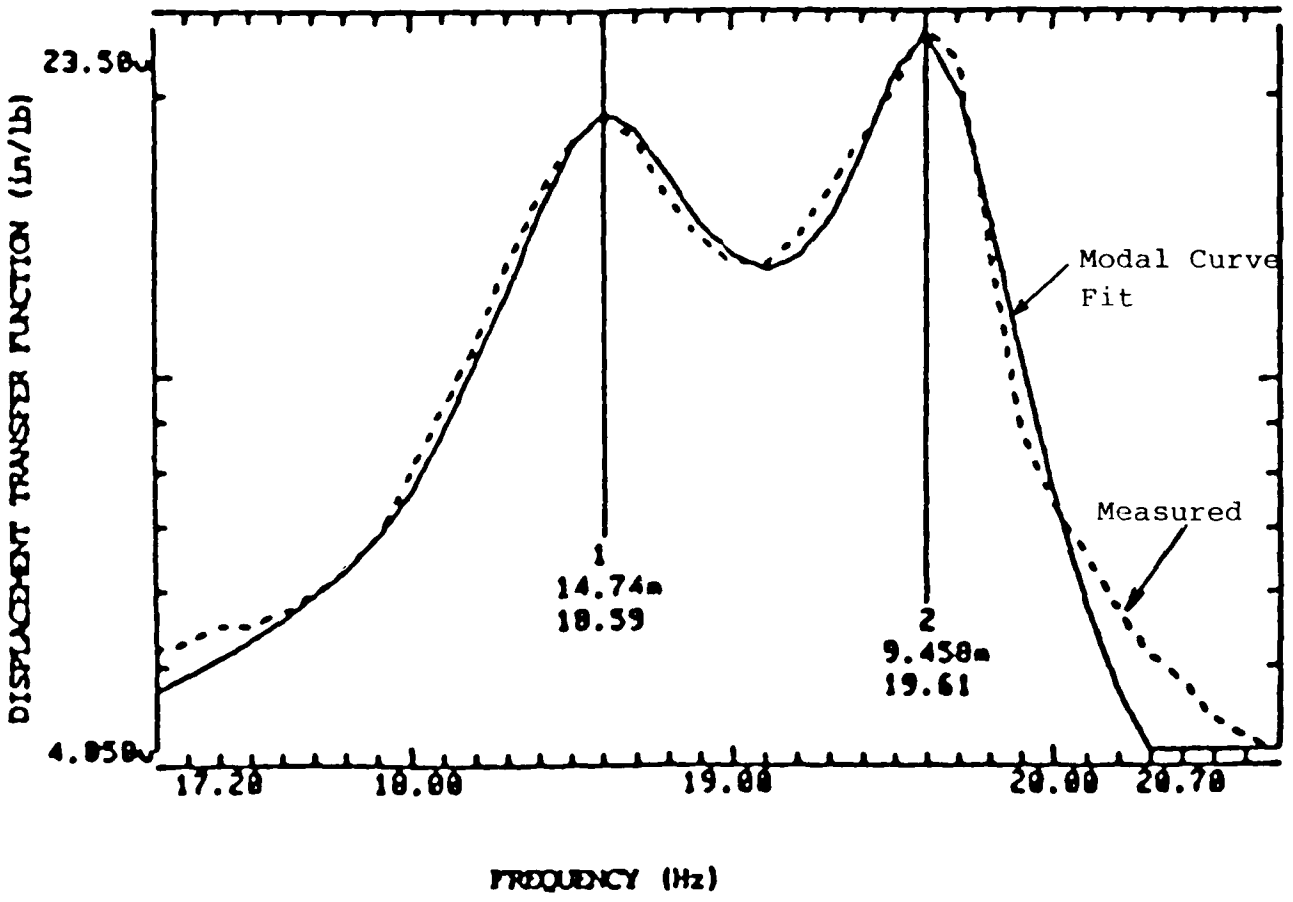


Figure 8-4. Two Mode Curve Fit in Modal Analysis.

Data from Ref. 8-6

With the values of $\Psi_n(x_1)$ obtained experimentally, the values of $\Psi_n(x_i)$ at other positions x_i can be obtained by measuring the transfer frf $H_{y_i f_1}$. It is important that the proper boundary conditions be applied to the structure when it is tested because these affect the modal parameters. A check for consistency in the measured data can be made by measuring some reciprocal frf (with the excitation and response points reversed) and comparing results to verify that $H_{y_1 f_i} = H_{y_i f_1}$.

Table 8-3 summarizes the procedures used in the method. More information on modal analysis can be obtained from References [8-5,6,7].

Table 8-3 Summary of Procedures for the Modal Analysis Method

A. Inputs Required
   1. Geometric description of system including a set of interconnected node points.
   2. Access to physical system for instrumentation and measurement of frequency response functions at the node points.
   3. Idealized model of the excitation as a set of forces or motions at the node points.

B. Steps of Solution
   1. Identify the frequency range and dynamic range of the measurements to give good data.
   2. Measure a complete set of frequency response functions that include the drive point response of one node and the transfer response at all other nodes.
   3. Evaluate the modal parameters using an appropriate numerical curve fitting routine.
   4. Check for consistency by repeating measurements with loading at other node points.
   5. Compute the system response at the node points by multiplying the modal transfer functions by the excitation.

C. Outputs Obtained
   1. System response (amplitude and phase) at node points to a particular excitation at node points.
   2. System transfer functions (response of node points to unit excitation at same or other node points).

8.5 STATISTICAL ENERGY METHOD

The method of statistical energy analysis has been developed for use in the frequency range where structural responses have a high density of resonant modes [8-8]. This method uses the form of Equation (8-20) to model the vibration response of each structural mode and evaluate the coupling of different modes by assuming a statistical distribution of the resonant frequencies.

8.5.1 System Model

SEA models a dynamic system by dividing it into subsystems of resonant modes which can be considered as independent energy storage units. These subsystem models are connected together analytically by expressions which represent the physical coupling between modes which allows power transmission from the more energetic modes to the less energetic ones. The coupling of modal responses of the individual subsystems is analyzed on a statistical basis. The important dynamic parameters of each subsystem are the modal density (in radian frequency) $n(\omega)$, the impedance Z at each connection point, the internal damping loss factor η , and the mean square velocity $\langle v^2 \rangle$ of the vibration response averaged over time and space.

The statistical nature of this type of analysis comes from the assumption that the system modal responses are randomly distributed in frequency and space. This assumption is generally more valid for higher order modes than for the lower order modes and therefore limits the application of the analysis at low frequencies. However, with this assumption two types of averaging can be used to greatly simplify the evaluation of the mean square response of interconnected subsystems at higher frequencies.

One type of averaging is over frequency. If the response of a subsystem is to be averaged over a frequency band (such as a 1/3 octave band) which contains a sufficiently large number of

subsystem modal frequencies, then only the average dynamic parameters of the subsystem must be known. In this case the exact frequency of each resonant mode does not have to be known, only the average modal density in each frequency band.

The second type of averaging is over ensembles. If a subsystem response is to be averaged over a sufficiently large number of similar but not identical samples or the subsystem is connected to others by a sufficiently large number of independent junctions, then the same type of analysis can be used as in frequency averaging. The number of modes or ensembles necessary to meet the "sufficiently large" requirement depends on more detailed information about the subsystems and on the desired confidence interval of the response prediction.

The definition of a subsystem in SEA is a group of resonant modes that share energy on an equal basis because they have similarities in their dynamic characteristics. These groups of modes can usually be identified by the different classes of modes in a structure (such as longitudinal, bending, shear, etc.).

The subsystems of a complex structure can be identified first by dividing the structure into substructures which are interconnected by clearly definable junctions. Then, for each substructure, the different classes of modes are identified as an individual subsystem. Each subsystem is described dynamically by three parameters: the modal density $n(\omega)$, the wave number $k(\omega)$, and the damping loss factor η_d .

The modal density is the number of resonant modes in a narrow frequency band divided by the bandwidth in radians/sec. Formulas for evaluating the modal density for many types of subsystems are given in Reference [8-2]. The wave number is the spatial density of the modes measured in radians/unit length, and is related to the wave speed $c(\omega)$ by

$$k(\omega) = \frac{\omega}{c(\omega)} \qquad (8\text{-}27)$$

The damping loss factor is a dimensionless measure of the rate of energy dissipation in the subsystem based on an energy decay of the form $e^{-\omega\eta t}$. The value of η is related to other measures of damping, such as the critical damping ratio ζ , the Q-factor, and the decay rate DR (dB/sec), by

$$\eta = \frac{1}{Q} = 2\zeta = \frac{DR}{4.3\omega} \qquad (8\text{-}28)$$

8.5.2 Excitation Model

When an isolated subsystem is excited by a steady-state driving force, the vibration will reach an equilibrium where the input power from the excitation is equal to the dissipated power due to the subsystem damping (assuming for now that the subsystem is not connected to others). With an excitation force of mean square amplitude $\langle f^2 \rangle$, the average input power Π_{in} is given by

$$\Pi_{in} = \langle f^2 \rangle \bar{G} \qquad (8\text{-}29)$$

where $\bar{G}$ is the average value of the real part of the subsystem point force mobility (inverse of impedance). Except at locations on the structure which are free edges, the value of $\bar{G}$ is given by

$$\bar{G} = \frac{\pi}{2} \frac{n(\omega)}{M} \qquad (8\text{-}30)$$

At free edges the value of $\bar{G}$ is approximated by

$$\bar{G}_{\substack{\text{free}\\ \text{edge}}} \simeq \pi \frac{n(\omega)}{M} \tag{8-31}$$

The dissipated power Π_{diss} in a subsystem is given by

$$\Pi_{diss} = \omega \eta_d n(\omega)\ \varepsilon \tag{8-32}$$

where ε is the average subsystem modal energy defined by

$$\varepsilon = \frac{M\langle a^2 \rangle}{\omega^2 n(\omega)} \tag{8-33}$$

where M is the subsystem mass and $\langle a^2 \rangle$ is the mean square acceleration of the subsystem (averaged over space and time).

The condition of equilibrium gives

$$\Pi_{in} = \Pi_{diss} \tag{8-34}$$

Then the ratio of the average subsystem acceleration to the force can be expressed as

$$\frac{\langle a^2 \rangle}{\langle f^2 \rangle} = \frac{\omega \bar{G}}{\eta_d M} \tag{8-35}$$

This ratio is valid for locations away from the driving force.

8.5.3 Coupling Between Subsystems

When a subsystem is connected to another there is an additional loss of power due to the vibration transmission across the connection. The net transmitted power from subsystem 1 to subsystem 2 is denoted by $\Pi_{1;2}$ and is given by

$$\Pi_{1;2} = \omega \eta_{1;2} n(\omega)_1 \quad [\varepsilon_1 - \varepsilon_2] \tag{8-36}$$

where $\eta_{1;2}$ is called the coupling loss factor. The power flow between two subsystems is in the direction toward the subsystem with the lower average modal energy.

The coupling loss factor at a connection is given by

$$\eta_{1;2} = \frac{2 I_{12}}{\pi \omega n(\omega)_1} \frac{G_1 G_2}{|Y_1 + Y_2|^2} \tag{8-37}$$

where the mobilities are those which characterize the motion of each system at the connection (called "junction" mobilities) and I_{12} is a transmission factor, and $G_i = \mathrm{Re}(Y_i)$. If the connection is a point, then I = 1.

When there is more than one type of power flow at a junction due to multiple degrees of freedom, the total coupling loss

factor can be obtained by summing the individual $\eta_{1;2}$ in the form of Equation (8-37) for each degree of freedom.

In the general case of N interconnected subsystems, the average response of each subsystem can be obtained by solving the set of N equilibrium equations:

$$\Pi_{i;diss} + \sum_{\substack{j=1 \\ (j\neq i)}}^{N} \Pi_{i;j} = \Pi_{i;in} \tag{8-38}$$

By substituting in the expressions for the power flow, these N equations can be expressed in the matrix form

$$[A] \{\varepsilon\} = \frac{1}{\omega} \{\Pi_{in}\} \tag{8-39}$$

where

$$A_{ij} = -n(\omega)_i \eta_{i;j} \quad i\neq j \tag{8-40}$$

and

$$A_{ii} = n(\omega)_i \eta_{i;diss} + \sum_{\substack{j=1 \\ (j\neq i)}}^{N} n(\omega)_i \eta_{i;j} \tag{8-41}$$

The [A] matrix is symmetric since

$$n(\omega)_i \eta_{i;j} = n(\omega)_j \eta_{j;i} \tag{8-42}$$

Solving for the subsystem modal energies gives

$$\{\varepsilon\} = [A]^{-1} \frac{1}{\omega} \{\Pi_{in}\} \tag{8-43}$$

The average acceleration level of each subsystem is then given by

$$\langle a_i^2 \rangle = \frac{\omega^2 n(\omega)_i \varepsilon_i}{M_i} \tag{8-44}$$

This matrix equation can be solved for each frequency band of interest.

An example of the application of SEA to the vibration response of a gear box is shown in Figure (8-5). Table 8-4 summarizes the procedures used in the Statistical Energy Analysis method. More information on SEA can be found in References [8-8,9,10].

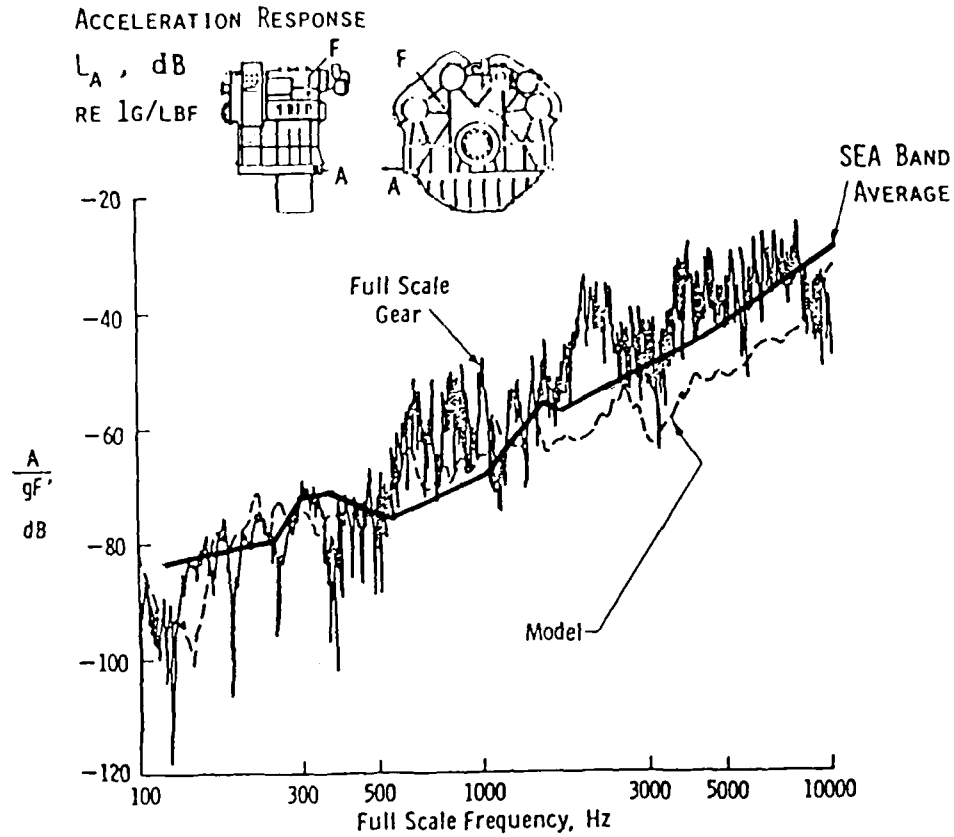


Figure 8-5. Comparison of the SEA Predictions with ¼ Scale and Full Scale Measurements of the Acceleration Response of the Gear Housing.

Table 8-4 Summary of Procedures for the Statistical Energy Analysis Method

A. Inputs Required
   1. Geometric description of system including a set of idealized groups of modes in each subsystem.
   2. Statistical properties of the modes in each subsystem (e.g. modal density $n(w)$, impedance Z, wave number k, damping loss factor η).
   3. Statistical model of excitation giving input power to each group of modes.

B. Steps of Solution
   1. Calculate the statistical parameters of each subsystem in the frequency range of interest.
   2. Calculate the coupling factors between mode groups of interconnected subsystems.
   3. Set up matrix equation for power balance between subsystems.
   4. Solve matrix equation for subsystem modal energies based on input power from excitations.
   5. Calculate statistical response of subsystems and power flow between subsystems.

C. Outputs Obtained
   1. System response (statistical amplitude) averaged over region of each subsystem and averaged over frequency.
   2. Power flow between subsystems.

8.6 DISCRETE ELEMENT MODEL

There are many different types of discrete element models being used to solve acoustic and vibration problems. Probably the most widely used method of this kind is the finite element model. The discrete element models are distinguished by their approximate solution to the wave equation over a finite distance or element size in a dynamic system. A model of the complete system is then built up by assembling a sufficient number of these pieces in the proper manner.

The spatial domain of the system is discretized into a number of nodal positions where the mass of the system is assumed concentrated. These nodal positions are then interconnected by stiffness (and sometimes damping) elements that represent the forces generated by the relative motions of the nodal positions. When these forces are balanced with the inertial forces and external forces on the system, the following matrix equation is obtained

$$[\ M\ ]\ \left\{ \frac{d^2x}{dt^2} \right\} + [\ C\ ]\ \left\{ \frac{dx}{dt} \right\} + [\ K\ ]\ \{\ x\ \} = \{\ F\ \} \qquad (8\text{-}45)$$

where {x} is the vector of nodal motions, [M] is the diagonal matrix of nodal masses, [C] is the viscous damping matrix, [K] is the stiffness matrix (having complex terms for material damping if desired), and {F} is the vector of external forces applied to the nodal positions.

Assuming sinusoidal motion at frequency $\omega = 2\pi f$ of the form

$x(t) = X(\omega)\ e^{jwt}$ gives

$$[\ -\omega^2\ [M] + j\omega\ [C] + [K]\ ]\ \{X\} = \{F\} \qquad (8\text{-}46)$$

This system of equations can be solved for X if all the othere terms are known. The matrix

$[B] = [\ -\omega^2\ [\ M\ ] + j\omega\ [\ C\ ] + [\ K\ ]\ ]$ is called the system matrix.

A few general observations of discrete element models can be made before considering the details of specific methods. Since the system wave equation is reduced to a finite number of equations, only a finite set of degrees of freedom (resonant modes, or eigenvalues) are modeled. Therefore, the model is valid only up to a certain frequency. In order to extend the model to higher frequencies, more degrees of freedom (and therefore more elements) must be included in the model. Since the solution requires the manipulation of a matrix of the dimension of the number of degrees of freedom, the computation time required grows at approximately a rate proportional to the square of the number of degrees of freedom in the model.

However, since the discrete element models divide up very complex structures into small pieces that can be modeled accurately, they can provide very accurate results within the valid frequency range.

8.6.1 Lumped Parameter Models

Lumped parameter models divide a structure up into ideal mass, spring, and dashpot elements. Both translational and rotational motions can be included. The basic equations for these three types of elements (F=ma, F=cv, F=kx) are combined with appropriate force and motion compatibility. An example of a lumped parameter model with the associated frequency response function is shown in Figure 8-6. More information on lumped parameter models can be found in References [8-11,12].

8.6.2 Finite Element Model

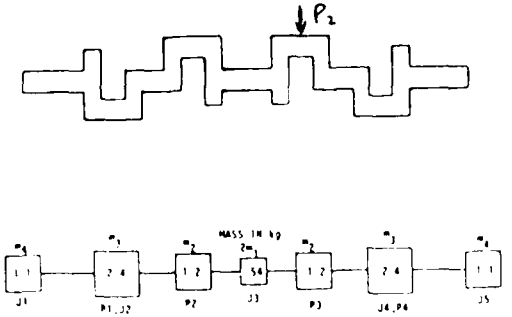
By far the most popular and extensively developed discrete element model is the finite element model (FEM). The nodal positions in a structure are interconnected by elastic elements which give the FEM the power of an accurate solution with modeling ease. Many different computer codes have been developed to provide elastic elements of different types and complexity of shapes with the necessary equations for generating the mass, stiffness (and sometimes damping) matrices built in.

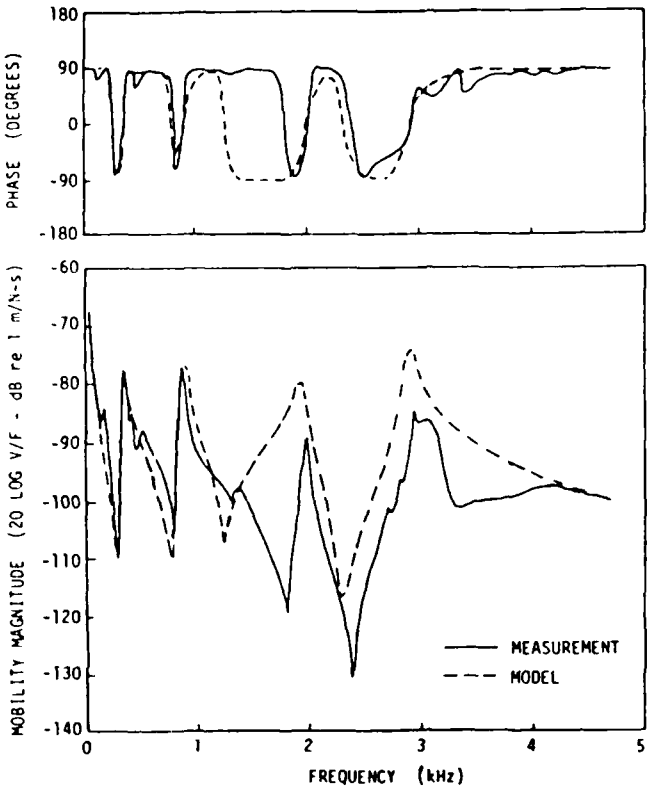
For example, a compressional rod element of length L and with a node at each end would have a mass and stiffness matrix given by [8-13].

$$[\,M\,] = \begin{bmatrix} M/2 & 0 \\ 0 & M/2 \end{bmatrix} \qquad (8\text{-}47)$$

$$[\,K\,] = \frac{ES}{L} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

By assembling a number of these elements with the proper geometric relationships, the dynamics of the system being modeled can be approximated. The accuracy of the FEM depends in part on the





DRIVE-POINT MOBILITY P2-P2

Figure 8-6. Lumped-Parameter Model of the Crank Shaft.

choice of the element types to be used. The FEM computer codes have a number of different element types from which to choose, and are programmed to calculate the appropriate matrix elements based on the geometric and material parameter specifications supplied by the user. Although each node point has in general six degrees of freedom (three translational and three rotational), most element types only use a subset of these degrees of freedom. Therefore, it is important to understand what types of motions each element type is trying to model in order to construct an accurate FEM.

Since a solution with damping terms would require the use of complex arithmetic and therefore much slower solution times, large models usually do not attempt to represent the details of the damping in the model, but include it after the fact in an approximate manner.

8.6.3 Excitation and Response Model

Two methods of solution are possible for the set of equations generated by discrete element models in order to predict the vibration response of the system. The first is called the "direct" method because it depends directly on the system matrix in terms of the modal degrees of freedom. For a steady state force excitation the system matrix can be inverted for each frequency of interest to obtain a solution for the modal displacements in the form:

$$\{X\} = [\ B\ ]^{-1}\ \{\ F\ \} \qquad (8\text{-}48)$$

Alternatively, for transient excitation, the direct method uses a time integration method to approximate the time derivatives in Equation (8-45) at a series of time steps, and solves for the modal displacements at each time step.

The advantages of the direct method is that it allows for a very general form of the damping and excitation functions in both space and time. The disadvantage is that a large number of computations are required to solve the equations over a wide range of frequencies or time steps.

A second method, called "modal superposition", is more frequently used. This method transforms the system equation into a modal formulation where the resonant frequencies are eigenvalues and the mode shapes are the eigenfunctions of the equation:

$$\text{Det}\ [\ B\ ] = 0 \qquad (8\text{-}49)$$

Usually the system equations are solved without damping in order to avoid complex arithmetic. The damping is then added to each mode in a manner similar to the modal solution methods described in Section 8.4. When the damping is very small ($\eta < 0.02$) the mode shapes can be assumed to be unchanged by the presence of the damping.

The response of each mode to the excitations is then summed up to give the total response at each node point. For an arbitrary excitation function in time, all of the modes must be included in the summation. However, for a steady state excitation at a single frequency, the system response can be evaluated with good accuracy by summing up only those modal responses with resonant frequencies near the excitation frequency.

The advantage of the modal summation method for steady state excitation is that the number of computations required for an accurate solution is greatly reduced from the other methods. The disadvantage is that it is limited in the types of damping and excitation functions it can handle.

Table 8-5 summarizes the procedures used in the Finite Element method.

More information on finite element analysis can be found in References [8-13,14,15].

Table 8-5 Summary of Procedures for the Finite Element (Modal Superposition) Method

A. Inputs Required
   1. Geometric description of system including a set of idealized elements interconnected at node points
   2. Static and dynamic properties of each element type (e.g. modal mass and stiffness matrix)
   3. Idealized model of the excitation either as a set of forces or motions at the node points or as a set of modal forces

B. Steps of Solution
   1. Calculate mass and stiffness factors at each node point in each element as a function of frequency
   2. Assemble the full matrix equation in the form of an eigenvalve solution for the resonant frequencies
   3. Solve the matrix equation for the resonant frequencies and mode shapes of interest
   4. Calculate the response of each node to the modal excitations
   4. Calculate the response of each node point by summing the contributions of the different modes

C. Outputs Obtained
   1. System response (amplitude and phase) at node points to a particular excitation at node points
   2. System transfer function (response of node points to unit excitation at same or other node point)

8.7 TRANSFER FUNCTION MODELS

Transfer function models fall into the general category of mathematical modeling. In these cases the measured frequency response functions of a system are modeled analytically by pole-zero representations using linear prediction methods. The pole-zero models can then be used to identify system parameters. In some cases pole-zero representations of input frequency response functions can be converted into lumped element equivalents which relate to physical system parameters; in which case the model becomes physical [8-11]. In most cases, however, the pole-zero model is simply a convenient analytic form, particularly when the dynamics of the combined system are to be predicted from measurements made on each part of the system independently. The use of measured data to define system parameters in this case has received much attention.

The frequency range over which these models are valid is determined by the number of degrees of freedom incorporated in them. The upper frequency limit is usually set by the maximum number of resonant frequencies that the model can predict which is of the order of the number of elements times the number of degrees of freedom of the basic element.

8.7.1 System Model

The method of modeling mechanical elements by pole-zero models is taken directly from the theory of electrical network synthesis of passive systems [8-16]. A direct analogy between electrical systems and one dimensional mechanical systems can be made using modern system dynamic theory [8-17]. However, it has been shown [8-18] that this theory can be generalized to systems of more than one dimension, having more than one input and output.

The concept of a mechanical element is shown schematically in Figure 8-7. In its simplest form, it is a collection of

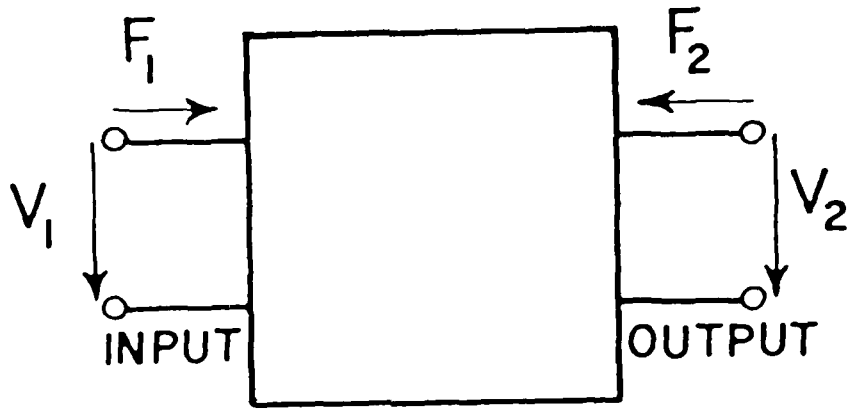


Figure 8-7. Input, Output Model of Individual Components.

matter with two points of interest, an input and output terminal, which move only in one dimension (e.g., translation along a line). The dynamic state of one of the terminals can be uniquely described by the two variables: force and velocity. These two variables are most convenient to use here since the power flow at the terminal point is measured by their product. The direction of positive velocity at each terminal can be arbitrarily chosen. Then a positive force is assigned to the direction which produces velocity at that terminal with all other forces equal to zero.

This model can be generalized to more than one dimension by treating the force and velocity variables as vectors of up to six dimensions (three translational and three rotational). Also, elements with more than one input or output terminal can be described by expanding the force and velocity variables to higher dimensional vectors which are sets of sub-vectors for each terminal. In this way, a general method can be developed using matrix equations of arbitrary dimensions, and the relevant equations to the proper dimensional order [8-18].

8.7.2 Mobility Equations

Referring to Figure (8-7), the force and velocity at the input and output terminals of a mechanical element can be related by the matrix equation,

$$\begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \qquad (8\text{-}50)$$

The quantity Y_{ij} is called the drive point mobility if $i = j$ and the transfer mobility if $i \neq j$. The quantities V, F, and Y are assumed to be complex amplitudes at one frequency, f, having a time dependence of $e^{j\omega t}$, where $\omega = 2\pi f$. A completely equivalent relationship can be written using the inverse of the mobility matrix

which is the impedance, Z,

$$\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix} \cdot \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \qquad (8\text{-}51)$$

The impedance form is usually the most convenient way to derive the equations of motion for an element. But, experimentally, it is easier to measure the mobility directly by taking the ratio of one velocity to one force with the other force equal to zero. For example,

$$Y_{11} = \left. \frac{V_1}{F_1} \right|_{F_2 = 0} \qquad (8\text{-}52)$$

The same type of measurement can be done conceptually for the impedance,

$$Z_{11} = \left. \frac{F_1}{V_1} \right|_{V_2 = 0} \qquad (8\text{-}53)$$

However, in most mechanical systems it is much easier to eliminate the force at a terminal than it is to eliminate the velocity. Therefore, the mobility equations are more convenient to use here.

The complete set of relationships for finding the individual functions in the mobility matrix is then,

$$Y_{11} = \left. \frac{V_1}{F_1} \right|_{F_2 = 0} \qquad Y_{12} = \left. \frac{V_1}{F_2} \right|_{F_1 = 0} \qquad (8\text{-}54)$$

$$Y_{21} = \left.\frac{V_2}{F_1}\right|_{F_2 = 0} \qquad Y_{22} = \left.\frac{V_2}{F_2}\right|_{F_1 = 0}$$

One simplification can be made using the reciprocity principle, which states that for linear, passive systems, $Y_{12} = Y_{21}$.

8.7.3 Cascading Elements

This section develops the equations for calculating the mobility functions for a system of two mechanical elements joined end to end, or cascaded, as shown in Figure 8-6. This makes it possible to determine the characteristics of a large, complicated system from the characteristics of the elements which make up the system and which might be easier to model analytically.

In considering the cascading of the two elements, a and b, it is easier to rewrite the mobility Equation (8-50) for each element in an input-output form,

$$\begin{bmatrix} V_1 \\ F_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} V_2 \\ -F_2 \end{bmatrix} \tag{8-55}$$

where,

$$\begin{aligned} A &= Y_{11}Y_{12}^{-1} \\ D &= Y_{12}^{-1}Y_{22} \\ C &= Y_{12}^{-1} \\ B &= AC^{-1}\,D - C^{-1} \end{aligned} \tag{8-56}$$

Then, if terminal 2 of element a is connected to terminal 3 of element b, $V_2 = V_3$ and $F_2 = -F_3$.

It is then possible to relate terminals 1 and 4 by,

$$\begin{vmatrix} V_1 \\ F_1 \end{vmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}_a \cdot \begin{bmatrix} A & B \\ C & D \end{bmatrix}_b \cdot \begin{bmatrix} V_4 \\ -F_4 \end{bmatrix} \qquad (8\text{-}57)$$

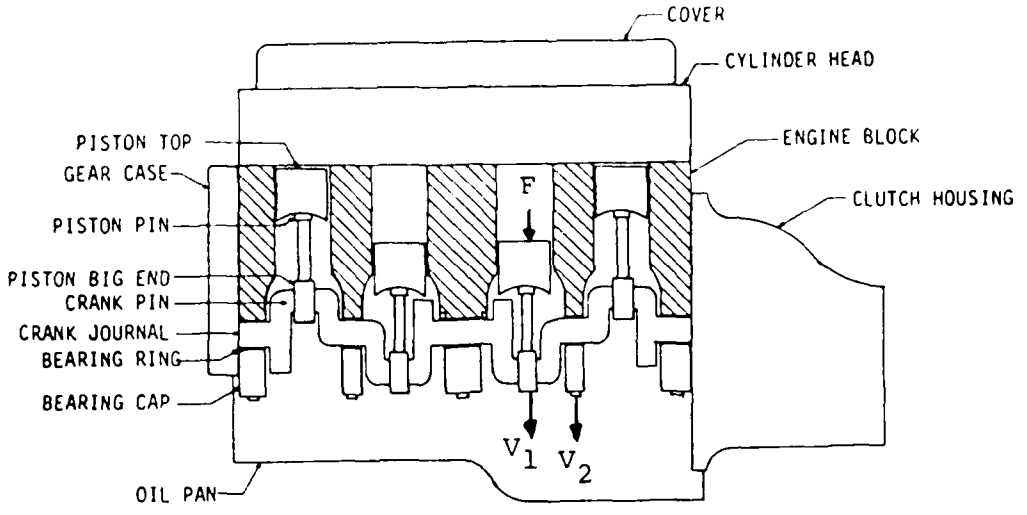
or

$$\begin{bmatrix} V_1 \\ F_1 \end{bmatrix} = \begin{bmatrix} A^aA^b + B^aC^b \\ C^aA^b + D^aC^b \end{bmatrix} \cdot \begin{bmatrix} A^aB^b + B^aD^b \\ C^aB^b + D^aD^b \end{bmatrix} \cdot \begin{bmatrix} V_4 \\ -F_4 \end{bmatrix} \qquad (8\text{-}58)$$

This procedure can be repeated for each successive element connected to either end of the previous series of elements.

An example of this procedure applied to the connection of two internal components in an engine is shown in Figure 8-8.

Table 8-6 summarizes the procedures used in the transfer function method.



CROSS-SECTION VIEW OF A FOUR-CYLINDER DIESEL ENGINE

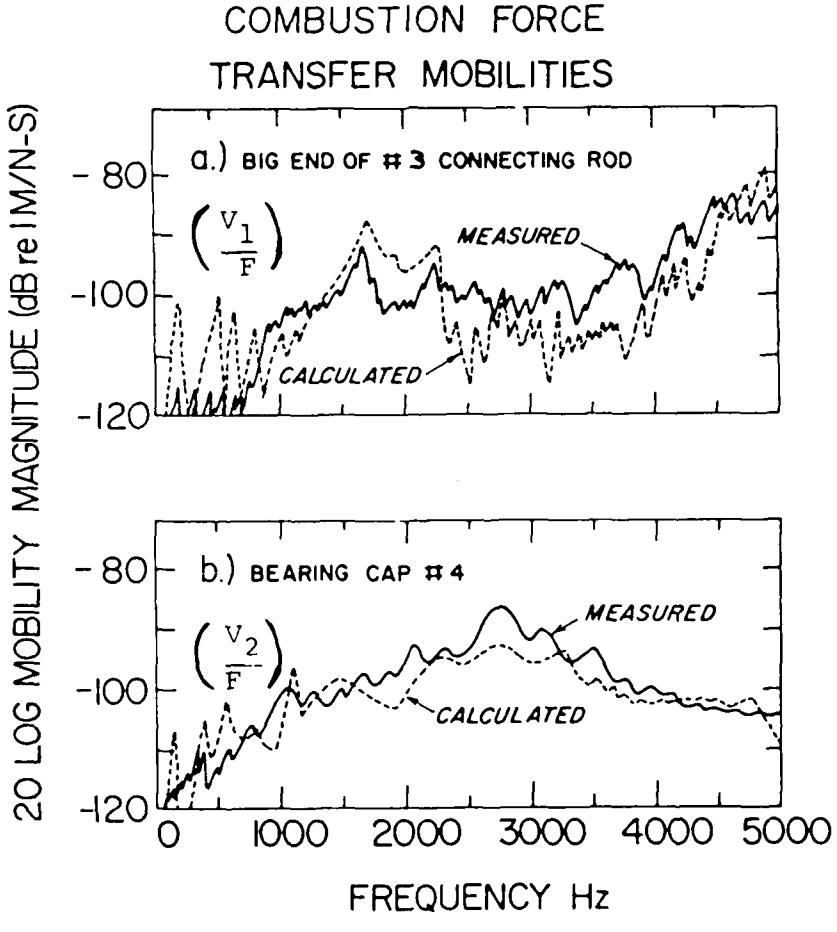


Figure 8-8. Comparisons of Calculated and Measured Transfer Functions in a Diesel Engine.

Table 8-6 Summary of Procedures for the Transfer Function Method

A. Inputs Required
   1. Geometric description of system as a set of substructures interconnected at idealized junctions
   2. Access to physical substructures for instrumentation and measurement of frequency response functions (mobilities) at attachment points
   3. Idealized model of the excitation as a set of forces or motions at the substructures attachment points

B. Steps of Solution
   1. Identify the frequency range and dynamic range of the measurements to give good data
   2. Measure the complete set of mobilities for each substructure
   3. Evaluate the substructure parameters (poles & zeroes) using an appropriate numerical curve fitting routine
   4. Assemble the substructure models into a complete transmission model of the entire system
   5. Compute the system response at the substructure attachment points based on the system excitations by performing the necessary matrix computations

C. Outputs Obtained
   1. System response (amplitude and Phase) at substructure attachment points to the particular excitaiton
   2. System transfer functions between attachment points

8.8 EXTRAPOLATION TECHNIQUES

In many cases where minor design changes are made in a structure and an estimation of the change in response level is desired, it is not necessary to construct a complete model of the system to obtain an answer of acceptable accuracy. Rather, it is often possible to use simple scaling laws and extrapolation techniques for these problems. These techniques are sometimes derived from exact analysis of very simple systems such as a flat plate. Many of the concepts in statistical energy analysis fall in this category. Other extrapolation techniques have been empirically derived from detailed measurement programs [8-19,20,21].

These techniques are an important part of any prediction method no matter how complicated in that they give as a minimum a check on the order of magnitude of the prediction which comes out of the model. At other times the extrapolation rules may be the only thing available in the time constraints of the problem.

The basic equation from SEA that is used for extrapolations in many cases is

$$\langle a^2 \rangle = \frac{\omega W_{in}}{\eta M} \qquad (8\text{-}59)$$

where $\langle a^2 \rangle$ is the space-time averaged acceleration of the structure, $\omega = 2\pi f$ is the frequency, W_{in} is the input power from the source of excitation, η is the damping loss factor and M is the mass of the structure.

If the input power remains constant while the other system parameters are changed then the following scaling laws can be used

$$\langle a^2 \rangle_{new} = \frac{M_{old}}{M_{new}} \langle a^2 \rangle_{old} \quad (W_{in} \text{ constant }) \qquad (8\text{-}60)$$

$$\langle a^2 \rangle_{new} = \frac{\eta_{old}}{\eta_{new}} \langle a^2 \rangle_{old} \quad (W_{in} \text{ constant }) \qquad (8\text{-}61)$$

$$\langle a^2 \rangle_{new} = \frac{\omega_{new}}{\omega_{old}} \langle a^2 \rangle_{old} \; ; \; (W_{in} \text{ constant }) \qquad (8\text{-}62)$$

The complicating factor is that changes in M and ω often cause change in W_{in}. Two forms of W_{in} can be used to evaluate these changes, depending on the type of excitation. For an imposed motion excitation of a given vibration velocity $\langle v^2 \rangle$, then

$$W_{in} = R \langle v^2 \rangle \qquad (8\text{-}63)$$

where R is the real part of (Z) .

For an imposed force excitation $\langle f^2 \rangle$, then

$$W_{in} = G \langle f^2 \rangle \qquad (8\text{-}64)$$

where G is the real part of (1/Z).

For example, consider the one dimensional compression waves in a rod of cross sectional area S, with a force $\langle f^2 \rangle$ applied to one end. The input power is given by

$$W_{in} = G \langle f^2 \rangle = \frac{\langle f^2 \rangle}{\rho S c} \qquad (8\text{-}65)$$

In order to predict the change in vibration level due to a change in the cross sectional area of the rod, the following scaling law can be used

$$\langle a^2 \rangle_{new} = \frac{M_{old}}{M_{new}} \cdot \frac{W_{new}}{W_{old}} \langle a^2 \rangle_{old} \frac{S_{old}^2}{S_{new}^2} \qquad (8\text{-}66)$$

An example of the scaling law for the vibration response of an aircraft structure caused by the fluctuating pressure levels is shown in Figure (8-9), [8-22]. Table 8-7 summarizes the procedures used in the extrapolation method.

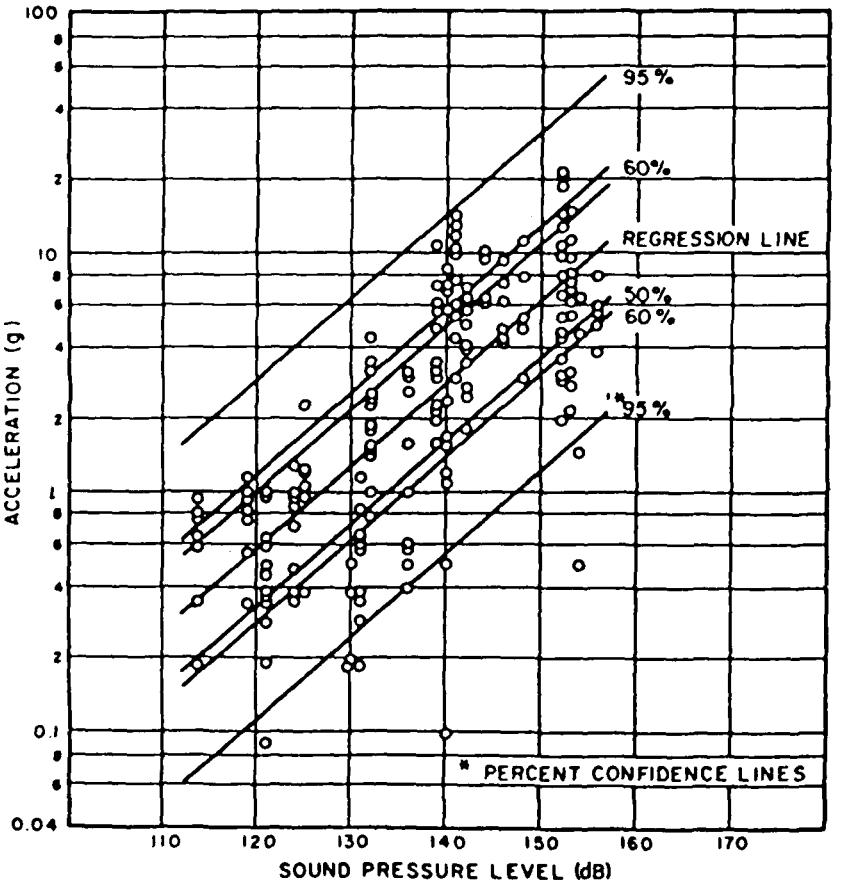


Figure 8-9. Scatter Diagram of Acceleration Levels L_a on B-58 Airplane Plotted Against Pressure Levels L_p in 600 to 1200 Hz Octave Band [8-22].

Table 8-7 Summary of Procedures for the Extrapolation Method

A. Inputs Required
   1. General geometric description of system in terms of relationship to previously studied systems
   2. General description of system excitations in terms of relationship to previously studied systems
   3. Information relating to the response of similar excitations

B. Steps of Solution
   1. Identify the type of response information needed
   2. Obtain desired response information from previously studied similar systems
   3. Identify scaling laws that describe significant differences between previous and current systems
   4. Evaluate scale factors that qualify the changes in respone due to changes in the system or excitation characteristics
   5. Compute the system response relative to the previous systems based on these scale factors

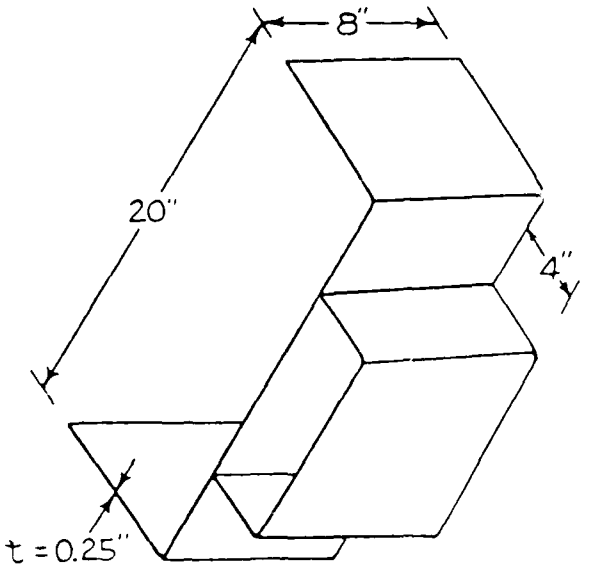
C. Outputs Obtained
   1. System response (approximate amplitude) based on information obtained from similar systems

8.9 COMPARISON OF METHODS

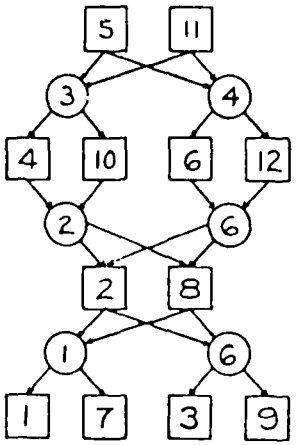
A typical comparison of three of the methods described in this chapter is shown in figure (8-10). This figure shows a comparison of the vibration response predictions for a test structure as obtained by modal measurements, FEM, and SEA. Figure (8-10a) shows the details of the test structure and the various models. Figure (8-10b) shows the narrow band (0-1400 Hz) results of the modal measurements and FEM along with the band average predictions of SEA. In general there is good agreement between the FEM and the modal measurements in this frequency range. The band average SEA results tend to overpredict the average measured results.

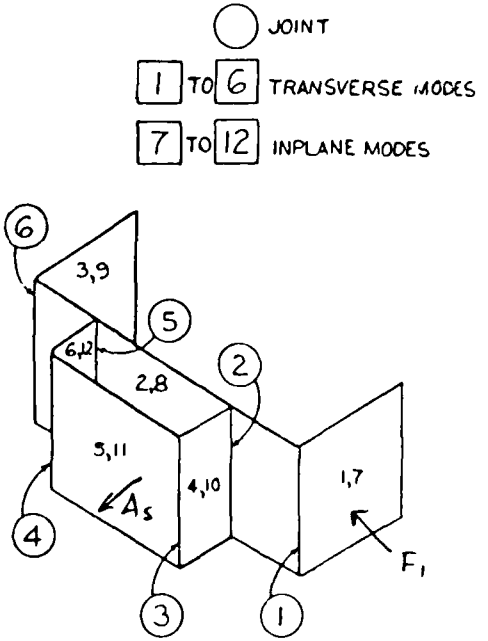
Figure (8-10c) shows the band average results of the three methods in a higher frequency range. Above about 6000 Hz the FEM prediction drops off due to the lack of modes. In general there is good agreement between the SEA and the band average measured levels in this frequency range.

Table 8-8 compares the relative merits of the different methods in relationship to the following list of criteria: inclusion of damping, level of expertise required, effort required, and frequency range of applicability. This table can be used to identify the best method to be used in a particular case depending on the requirements given.

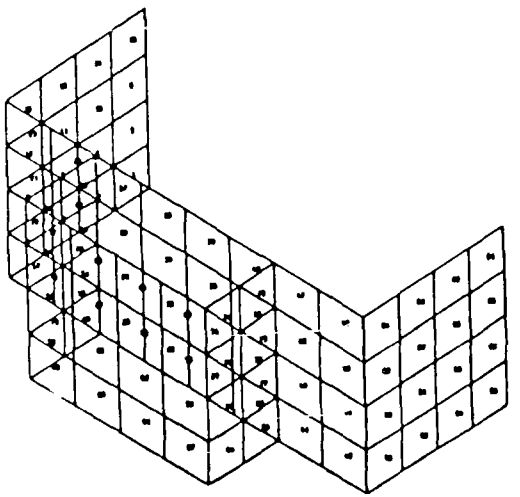


Test Specimen





Sea Model



Finite Element Model

Figure 8-10a. Test Structure

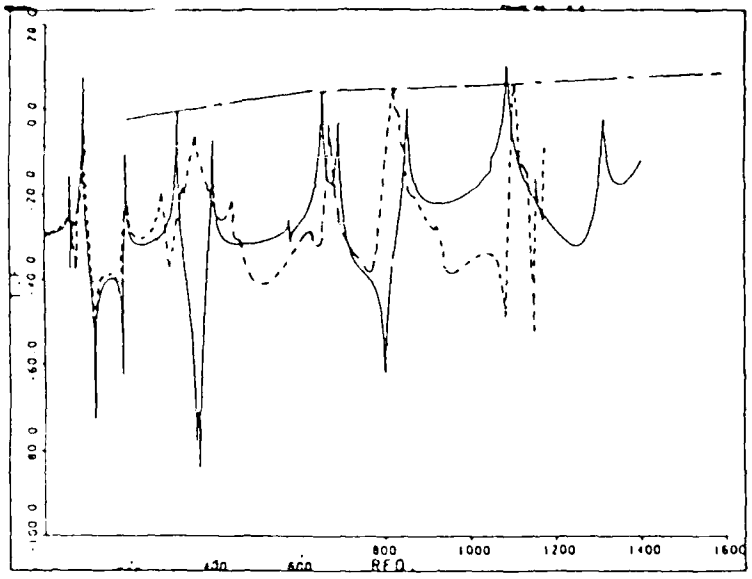


Figure 8-10b

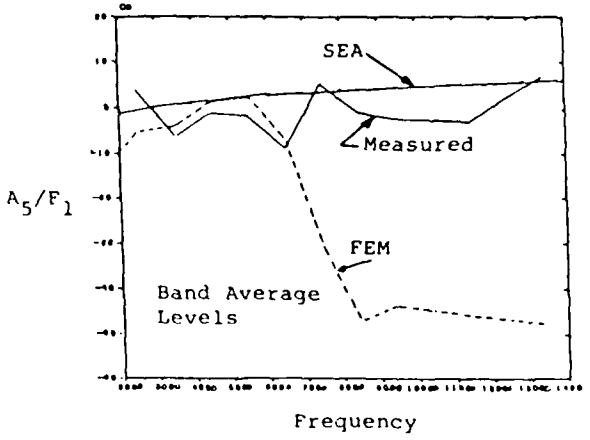


Figure 8-10c

Table 8-8 Comparison of Vibration Prediction Methods

Criteria: Method:	Propagating Wave	Modal Analysis	Finite Element	Statistical Energy Analysis	Lumped Parameter	Transmission Line	Transfer Function	Extrapolation Techniques
Model Type	physical	mathematical	physical	physical	physical	physical	mathematical	physical
Parameter Solution	theoretical	measurement	numerical	theoretical or measurement	numerical	theoretical	measurement	theoretical or measurement
Damping Included	sometimes	yes	sometimes	yes	yes	sometimes	yes	yes
Level of Expertise: -Set Up & Interpretation	high	med-high	med-high	high	high	high	high	med-high
-Implementation	high	low-med	low-med	med	low-med	med	med	med-high
Human Effort/Cost	med-high	high	med-high	med	med	med	med	low
Instrumentation and Computer Req't/Cost	low	high	high	low-med	low	med	high	low
Types of System Complexity Handled	low	low-high	low-high	med-high	low	low-med	low-high	med-high
Frequency Range (Relative to 100th Mode)	med-high	low	low-med	med-high	low	all	all	all
Limitations	no modes	requires test object	finite set of modes	only statistics of response	finite set of modes	idealized set of modes	requires test objects	requires previous data or experience

REFERENCES

8-1. Snowdon, J.C., Vibration and Shock in Damped Mechanical Systems, John Wiley & Sons, Inc., New York, NY, 1968.

8-2. Cremer, L., and Heckl, M., Structure-Borne Sound, Springer-Verlag Berlin Heidelberg, New York, NY, 1973.

8-3. Skudrzyk, E.J., Simple and Complex Vibratory Systems, Pennsylvania State U.P., University Park, 1968.

8-4. Timoshenko, S., Young, D.H., and Weaver, W. Jr., Vibration Problems in Engineering, John Wiley & Sons, New York, NY, 1974.

8-5. Formenti, D., and Welaratna, S., "Structural Dynamics Modification - An Extension to Modal Analysis", SAE Paper No. 811043, 1981.

8-6. Gimmestad, D.W., (Ed.), "An Improved Ground Vibration Test Method", Wright-Aeronautical Laboratories, Ohio, September 1980.

8-7. "Proceedings of 1st International Modal Analysis Conference", Orlando, FL., November 8-10, 1982, (Sponsored by: Union College, Schenectady, NY.)

8-8. Lyon, R.H., Statistical Energy Analysis of Dynamical Systems: Theory and Applications, The M.I.T. Press, Cambridge, MA, 1975.

8-9. Lyon, R.H. and Eichler, E., "Random Vibration of Connected Structures", J. Acoust. Soc. Am., Vol. 36, No. 7, pp. 1344-1354, July 1964.

8-10. Lyon, R.H. and Scharton, T.D., "Vibrational Energy Transmission in a Three Element Structure", J. Acoust. Soc. Am., Vol. 38, No. 2, pp. 253-261, February 1964.

8-11. DeJong, R.G., "Vibrational Energy Transfer in a Diesel Engine", SC.D. Thesis, Massachusettes Institute of Technology, August 1976.

8-12. Neubert, V.H., "Series Solutions for Structural Mobility", J. Acoust. Soc. Am., Vol. 38, No. 5, pp. 867-876, Nov., 1965.

8-13. Bathe, Klaus-Jurgen, Wilson, E. L., Numerical Methods in Finite Element Analysis, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1976.

8-14. Zienkiewicz, O.C., The Finite Element Method in Engineering Science, McGraw-Hill Book Company, New York, 1971.

8-15. Schaeffer, H.G., MSC/NASTRAN Primer: Static and Normal Modes Analysis, Wallace Press Inc., Mont Vernon, N.H., 1979.

8-16. Guillemin, E.A., Synthesis of Passive Networks, John Wiley & Sons, Inc., 1957.

8-17. Shearer, J.L., Murphy, A.T. and Richardson, H.H., Introduction to System Dynamics, Addison-Wesley Publishing Company, 1967.

8-18. Sykes, A.O., "Application of Admittance and Impedance Concepts in the Synthesis of Vibrating Systems", Synthesis of Vibrating Systems, ASME, pp. 22-37, 1971.

8-19. Mahaffey, P.T. and Smith, K.W., "A Method of Predicting Environmental Vibration Levels in Jet-Powered Vehicles", Shock and Vibration Bulletin, No. 28, Part 4, p. 1, August 1960.

8-20. "A Guide to Predicting the Vibrations of Fighter Aircraft in the Preliminary Design Stages", AFFDL-TR-71-63, 1973.

8-21. "Vibration Response of Ballistic Re-entry Vehicles", AFFDL-TR-72-140, 1973.

8-22. Lyon, R.H., Random Noise and Vibration in Space Vehicles, Shock and Vibration Information Center, U.S. Government Printing Office, Washington, DC, 1967.

8-23. L.K.H.Lu, et.al, "Comparison of Statistical Energy Analysis and Finite Element Analysis Vibration Prediction with Experimental Results", The Shock and Vibration Bulletin No. 53, Part 4, p. 145, May 1983.