Berkeley Physics Course

Volume 1

Mechanics

(In SI Units)

Charles Kittel
Walter D Knight
Malvin A Ruderman
Carl A Helmholz
Burton J Moyer

2e

VIBRATION AND ACOUSTICS

*Measurement
and Signal Analysis*

iitmadras
CELEBRATING
TECHNOLOGY

VIBRATION AND ACOUSTICS

Measurement and Signal Analysis

C. Sujatha

Indian Institute of Technology Madras
Chennai

To

*My parents Ponnammal and (late) Viswanathan who are
responsible for making me the person I am today*

and

*My husband Chandra Mohan who has unflinchingly
supported me in realizing my dreams*

Preface

This book aims at describing all topics concerning vibration and acoustic measurements, as well as, signal analysis techniques in a simple and concise manner. It is the culmination of more than two decades of teaching and research in these areas. I have had the opportunity to teach interrelated courses like theory of vibration, diagnostic maintenance, signal processing for mechanical engineers, modal analysis of mechanical systems and acoustics and noise control. The book evolved with my teaching and interactions with practising engineers from the industry and hopefully will be useful in analyzing real-life vibration and acoustics problems.

This book will serve as a reference book for postgraduate students, research scholars, practising engineers, researchers, academicians and professionals in the field of vibration and acoustics. It offers a survey of experimental techniques in vibration and acoustic measurements, the likes of which are not found in most books on similar themes. It has a comprehensive coverage of a wide range of vibration transducers and exciters, as well as, other transducers and equipment that go hand in hand with vibration measurement, signal analysis fundamentals, processing of random signals, common vibration monitoring techniques for condition monitoring, modal analysis, acoustic transducers, common acoustic measurements and case studies in vibration and acoustic measurements. The book contains adequate theoretical mathematics for a sufficient understanding of the frame work of the topics discussed. The basic theory of vibration and acoustics is also outlined, without which the book would be incomplete. Mechanical engineers often feel diffident when faced with the task of making vibration or acoustic measurements since they generally do not have an understanding of the working of transducers, signal conditioners and real-time analyzers and treat them as black boxes. The treatment of this theory in this book is sufficient for people with little or no exposure to measurements or electrical engineering to get started on measurements. The emphasis in this book is more on vibration than on acoustics.

Organization

Chapter 1 is a general introduction to the subject of vibration/acoustic measurements and signals. Chapter 2 describes the salient theory of vibration of lumped mass systems; continuous systems have not been discussed. Chapters 3 and 4 offer an elaborate discussion of a wide range of pickups, exciters and excitation techniques with guidelines as to their choice. The coverage of vibration and acoustic transducers is extensive, with photographs, wherever possible, applications and typical specifications of transducers of each kind. It may be noted that equivalent circuits of transducers are not considered. Chapter 5 deals with other transducers closely associated with vibration measurement and signal conditioning amplifiers,

as well as non-contact measurement techniques for vibration monitoring. Chapter 6 explains the fundamentals of signal processing such as signal types and operation, data acquisition, Fourier transforms and properties, sampling and aliasing, leakage, etc. Chapter 7 discusses the fundamentals of probability theory, modelling of random signals and the procedure for obtaining the response of a single degree of freedom system to random excitation. Chapter 8 deals with the commonly used vibration monitoring and analyses techniques. Many of the latest signal processing techniques for fault diagnosis and classification are presented. Chapter 9 examines vibration severity criteria and standards for machinery, whole-body and hand-arm vibration. Chapter 10 throws light on the basics of experimental modal analysis.

Chapter 11 explains the fundamental principles of acoustics and terminology often encountered. The theory behind sound propagation, transmission and behaviour in enclosed spaces is discussed. Chapter 12 elucidates various acoustic transducers such as microphones, loudspeakers and calibrators with photographs, typical specifications and criteria for their selection. Chapter 13 spells out various techniques for the measurement of common acoustic quantities like sound pressure level, sound power, intensity, reverberation time, absorption coefficient, transmission loss, etc. It also discusses noise regulation standards. Chapter 14 deals with case studies from my experience in industrial consultancy, as well as, students and research projects. The case studies discussed are simple meant for someone who does not have any exposure to vibration measurements.

The reader may refer to other books or monographs for more detailed information on these topics. Select references have been added at the end of each chapter, the list covering sources of general information. Detailed papers in specific topics have been omitted for the reason that pertinent literature (other than books) is very vast even in narrow areas and hence giving a superficial sample can mislead the reader. No numerical problems have been solved in this book since the range of topics covered is very wide and each chapter is a subject by itself, requiring discussion of a wide variety of problems. SI units have been used in this book though in a couple of case studies, FPS units are used.

All human efforts do not always yield perfect results. Readers can write to me at *sujchand@hotmail.com* if any errors and significant omissions are noticed. I will appreciate constructive criticism, comments and suggestions for improvement of this book.

C. Sujatha

Acknowledgements

There are many people I would like to thank for helping me in this endeavour. I am grateful to Prof. M.S. Ananth, Director, IIT Madras, for facilitating this project by offering special sabbatical leave under the Golden Jubilee Book Writing Scheme, which provided the much needed impetus for this book and without which it might not have taken off. I feel fortunate to have been able to work at IIT Madras, which has provided me academic freedom and enabled me to write this book. I am indebted to all my teachers, especially Prof. V. Ramamurti, my Ph.D. supervisor who introduced me to this fascinating subject of vibration and acoustic measurements. I would like to express my gratitude to my colleagues of Machine Design Section for all the fruitful discussions which have enhanced my knowledge and to the technical staff whose support made experimental work a pleasure. I am thankful to all my students and research scholars who have helped me understand concepts better and the industries which provided me insights into real-life vibration/acoustics problems through their consultancy projects.

It is my pleasure to acknowledge the individuals who have contributed to the evolution of this book. I thank Poongodi for having typed portions of the draft copy of the book and Asha who made most of my drawings, tirelessly doing them over and over again. I owe my gratitude to research scholars Ganesh Babu and Praveen Krishna, who helped generate some of the drawings for the book and Janarthanan who helped with some of the photographs. I would like to thank my husband Chandra Mohan for having painstakingly gone through the manuscript and made corrections in language and flow. I am extremely grateful to Prof. P. Chandramouli and Prof. S. Swarnamani for having reviewed the book and provided critical inputs. I would like to acknowledge the patience and sacrifices made by my sons Adithya and Aravind in this undertaking. Words cannot do justice to the support I have received all through my career from my in-laws (late) Savithri, Venkataramani, Jayasree and Siva. I appreciate the constant encouragement given to me by my brother Suresh, friend Beena and uncles Natarajan and Rajagopalan for completing this work. Finally, I would like to thank all those who have contributed in some way or the other and whose names do not feature here.

C. SUJATHA

Contents

Chapter

1

Introduction

1.1 NEED FOR VIBRATION AND ACOUSTIC MEASUREMENTS

Vibrations can create undesirable behaviour in machinery and structures. Vibration tests are commonly carried out either in the laboratory or in the field. Tests typically done are those during the developmental phase of a product to find out what is happening or qualification tests after design. Vibration and acoustic testing find applications ranging from machinery diagnostics to monitoring of structural integrity in space shuttles. A machine or structure or vehicle may:

- be uncomfortable to ride in
- be too difficult to control
- make too much noise
- not maintain tolerances
- wear too fast
- fatigue prematurely
- break down unexpectedly

All these possibilities necessitate the measurement of vibration and noise.

Today many sophisticated vibration measuring instruments are available at highly affordable prices and give very good performance. However it is required that they are handled by an engineer who has a thorough understanding of the principles of operation of the transducers available and their limitations, the appropriate signal-conditioning amplifiers, data acquisition systems commercially available and their capabilities, as well as recording techniques for storing data and processing techniques for retrieving the required information. The engineer should also know how to adapt a field test to laboratory conditions, what boundary conditions to use, where to locate the transducers and how to orient them. He or she should also be in a position to judge whether a given set of specifications for a vibration test is reasonable or not. The engineer often also has to sit in judgement over a report generated by someone else. This book is written primarily with the aim of equipping the vibration/acoustic engineer with the knowledge required.

Typically vibration measurements are made for the following reasons:

- Determination of the response of a vibrating machine/vehicle/structure
- Determination of vibration levels in road vehicles, off-road vehicles and tracked vehicles
- Characterization of suspensions
- Finding the response of a human being in a vibrating environment, such as in a vehicle, in a building or while operating a machine

- Finding out the natural frequencies, damping ratios and mode shapes associated with various modes of vibration (modal analysis)
- Determination of the forces causing a particular vibration
- Evaluation of vibration isolation
- Calibration of a transducer
- Evaluation of the damping of materials
- Condition monitoring of a machine to find out if it is operating within safe levels; if not the cause of vibration leading to the response has to be determined, or mitigation measures have to be taken
- Estimating vibration levels on massive structures like machine foundations, bridges, highways and buildings
- Measurement of ground motion during an earthquake and response of structures to earthquake excitation
- Comparison of vibration levels in similar consumer items of different makes, e.g. air conditioner, refrigerator, etc.
- Dynamic qualification tests for different requirements
- Estimation of torsional vibration response or system properties
- Passive and active vibration control
- Estimation of optimal parameters for fault detection

Audio acoustic measurements would typically fall under the following categories.

- Estimation of sound pressure level emanating from a machine and its directivity index
- Frequency analysis of sound signals
- Estimation of sound power levels of identical machines/products to compare their performances
- Estimation of sound intensity levels to identify the source of noise in a machine
- Determination of reverberation time of an enclosure
- Evaluation of sound absorption coefficients of acoustic materials
- Studies on transmission of acoustic signals
- Noise reduction techniques in machines and enclosed spaces
- Measurement of environmental noise, such as traffic noise, noise near an airport or in a commercial area
- Measurement of vehicle-pass-by noise
- Calibration of microphones
- Subjective rating of noisiness
- Audiometry to ascertain hearing loss
- Measurement of noise to ascertain machinery condition
- Measurement of noise in workshops and factories
- Measurement of construction noise
- Measurement and analysis of transient signals, such as sonic booms, explosions, gun shots, etc
- Study of the effects of the above on human beings

1.1.1 Requirements of a Transducer

For the kind of measurements mentioned above, we use vibration/acoustic transducers that perform the conversion of energy from one variable into another, typically from mechanical energy into electrical

energy or vice versa. Examples are piezoelectric accelerometers, loudspeakers and electrodynamic exciters. An accelerometer, for example, generates an electrical signal proportional to the acceleration input. Sensors come in a seemingly countless number of sizes and technical specifications and are based on myriad mechanical and electrical principles. Many of the conventional industrial applications for sensors require very good reliability and ruggedness with strong mechanical packaging and extensive shock, vibration and temperature ratings, necessitating the use of analogue sensors. In large-scale tests, the cost of the test is largely dictated by the reliability of the sensors and measurement system, since loss of data or erroneous data could render the tests useless. There has been, of late, a tremendous growth of new microelectric, micromechanical and wireless technologies, widening the scope of applications. Much of the new technology implemented into sensor design is specific to the digital domain. However, harsh environmental conditions for some applications still prevent the use of digital sensors.

A transducer is expected to have a good dynamic response, convenient instrumentation, giving a sufficiently high analogue output signal with high signal-to-noise ratio (SNR), low cost, stability and reliability, minimum errors in measurement, insensitivity to temperature changes, vibration and other environmental variations. The important features of a transducer to be considered are as follows:

- Frequency range: appropriate range depending on the expected frequencies in the signal to be measured
- Sensitivity: output electrical quantity per unit input mechanical quantity; the larger the sensitivity, the better is the measurement, but it comes at an increased cost. If $x(t)$ is the input physical quantity to be measured and $y(t)$ is the output electrical quantity (voltage), then the sensitivity of transducer is $c = \dfrac{y(t)}{x(t)}$
- Dynamic range: range of values of the physical quantity that can be measured by the transducer
- Ruggedness: the ability to withstand environmental conditions and overloads, with safety stops for overload protection
- Linearity: the ability to reproduce input/output behaviour, symmetrically and linearly, such that the output is a true reproduction of the input when taken as a function of time
- Stability: the ability to produce an output that is relatively unaffected by temperature, vibration and other environmental conditions
- Repeatability: the ability to reproduce the same output signal when the measurement is repeated under similar environmental conditions
- Convenient instrumentation, sufficiently high analogue output with a large SNR for some applications digital output is preferred

1.2 SIGNALS AND THEIR REPRESENTATION

1.2.1 Deterministic and Non-deterministic Signals

Signals may be broadly classified as deterministic and non-deterministic (random) as shown in Fig. 1.1. Deterministic signals are those which can be described in the time domain by a mathematical expression. Hence it is possible to compute the value of the signal at any instant of time. Non-deterministic signals are random in nature and have to be described in statistical or probabilistic terms rather than in the form of explicit equations. No physical data are, in practice, truly deterministic since there is always a possibility

that uncertainties exist in the process producing the data. No data are truly random either, since explicit mathematical descriptions might be possible if the mechanisms or processes producing the data were fully understood. In practice, if it is possible to repeat an experiment producing specific data with identical repeatable results, then the data may be said to be deterministic. Deterministic signals may be further classified as periodic and aperiodic (non-periodic). Periodic signals are those that repeat with a definite time period T, whereas non-periodic or aperiodic signals are those with period equal to infinity. Periodic signals may be further classified into simple harmonic and complex periodic signals.
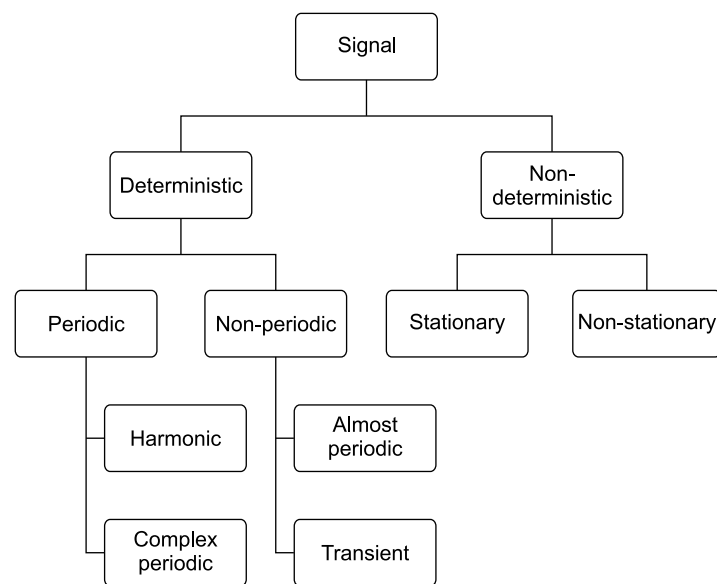
Figure 1.1 Classification of signals

The potential across a condenser as it discharges through a resistor and the temperature of water as heat is applied to a bath are examples of deterministic signals. Examples of non-deterministic signals are the motion of a power transmission line during a storm and the vibration of a building during an earthquake. More examples are given in the following sections.

1.2.2 Harmonic Signal

Vibration essentially deals with the oscillatory motion of bodies, the simplest form of which can be expressed as harmonic motion. Examples of simple harmonic motion (SHM) are:
- motion of a rotor rotating at a constant speed
- electrical power supply voltage
- motion of a swing
- motion of a satellite in orbit around a planet
- vibration of a tuning fork

- motion of the balance wheel of a watch
- motion of a cam follower arrangement
- vibration response of a rotating machine with an unbalanced mass
- voltage output of an electrical alternator

SHM is the simplest form of periodic motion and can be expressed by the equation

$$x(t) = X \sin (\omega t + \phi)$$

$$= X \sin (2\pi f t + \phi) \tag{1.1}$$

$$= X \sin \left(\frac{2\pi t}{T} + \phi \right)$$

Here X is the amplitude of motion measured as zero to peak value, ω is the angular frequency in radians per second, f is the frequency in Hertz and T is the period of vibration in seconds or time taken for the completion of one cycle; '$+\phi$' denotes the phase angle (lead) with respect to the reference point in time and '$-\phi$' would denote a phase lag. It is the phase angle or time lead/lag with respect to the reference time of measurement which makes the equation complex. SHM may be expressed in complex form as

$$x(t) = X e^{i\omega t} = X \cos \omega t + i X \sin \omega t = X_R + i X_I \tag{1.2}$$

Here X_R denotes the real part and X_I the imaginary part. In mathematical terms, $x(t)$ may be thought of as being generated by a vector of amplitude X rotating at constant speed ω as shown in Fig. 1.2.
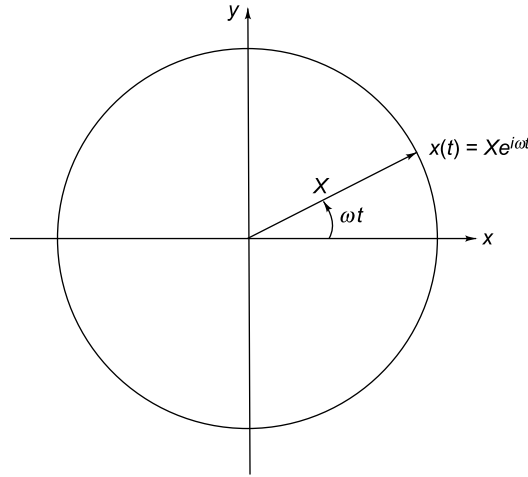
FIGURE 1.2   Harmonic motion depicted as a rotating vector

Knowing displacement $x(t)$, one can easily obtain the velocity and acceleration of motion by differentiation of Eqs (1.1) or (1.2). Figure 1.3 shows a time domain representation of displacement, velocity and acceleration, as well as the phase relationship between the three quantities.
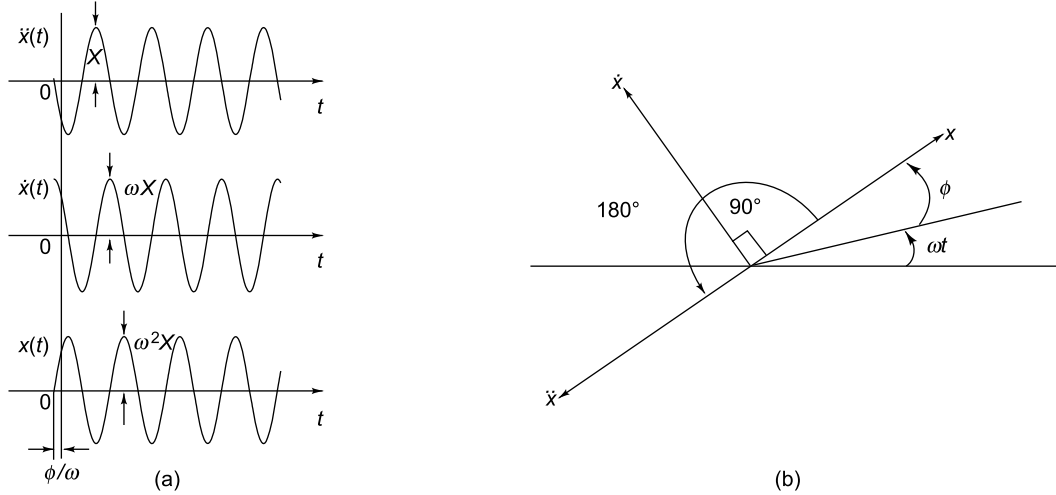
FIGURE 1.3 Displacement, velocity and acceleration in SHM: (a) time history, (b) phase representation

Differentiating Eq. (1.1) gives

$$\dot{x}(t) = \omega X \cos(\omega t + \phi) = \omega X \sin\left(\omega t + \phi + \frac{\pi}{2}\right) \tag{1.3}$$

$$\ddot{x}(t) = -\omega^2 X \sin(\omega t + \phi) = \omega^2 X \sin(\omega t + \phi + \pi) \tag{1.4}$$

Differentiation of Eq. (1.2) gives

$$\dot{x}(t) = \omega X e^{i\omega t} \tag{1.5}$$

$$\ddot{x}(t) = -\omega^2 X e^{i\omega t} \tag{1.6}$$

1.2.3 Complex Periodic Signals

Any motion which obeys the following rule is said to be periodic:

$$x(t) = x(t + T) \text{ for all } t \tag{1.7}$$

In such motion, generally a large number of frequency components are superimposed to produce $x(t)$, as against the case of SHM where we have only a single frequency component. Vibrations from a stringed instrument, such as a violin, guitar or veena consist of a fundamental frequency f and its higher harmonics $2f$, $3f$, etc. The frequency spectra of such signals appear structured, since they show a fundamental and repetitive harmonics with equal frequency spacing. The sawtooth voltage input given to the x-plate of an oscilloscope is another example of this type of signal. In practice, physical phenomena producing complex periodic signals are more common than those generating simple harmonic signals. Examples are the voltage produced by an electrical alternator and vibration response of a multi-cylinder reciprocating engine. The former produces barely discernable higher harmonic frequencies, while the latter displays a considerable number of higher frequencies. Figure 1.4 shows a periodic signal with period T.
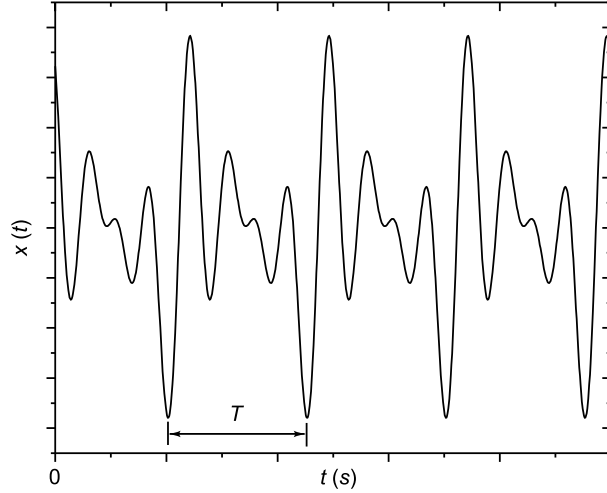
FIGURE 1.4 Complex periodic signal with period T

1.2.3.1 Fourier series Quite often it becomes difficult to extract information regarding frequency content of a signal from the time record and the signal has to be transformed to the frequency domain. A continuous-time periodic signal $x(t)$, may be represented in the frequency domain by an infinite Fourier series as shown in Eq. (1.8). The series involves summation over positive and negative frequencies and it also contains a constant or zero frequency or DC (from direct current in electrical engineering) term $a_0/2$, which is nothing but the average value of $x(t)$. In practice, positive frequencies alone are considered since negative frequencies do not have any physical meaning.

$$x(t) = \frac{a_0}{2} + a_1 \cos \omega_1 t + a_2 \cos \omega_2 t + \cdots + a_n \cos \omega_n t + b_1 \sin \omega_1 t + b_2 \sin \omega_2 t + \cdots + b_n \sin \omega_n t$$

$$= \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos\left(2\pi n \frac{t}{T}\right) + b_n \sin\left(2\pi n \frac{t}{T}\right) \tag{1.8}$$

Here $\omega_1 = 2\pi/T$ is the fundamental frequency; $\omega_n = n\omega_1$ is the nth harmonic or nth integral multiple of the fundamental frequency and T is the period corresponding to the fundamental frequency. The coefficients a_n, b_n are the Fourier or spectral coefficients for the function $x(t)$ or in other words they represent the spectral amplitudes corresponding to frequencies ω_n and can be computed from the relations

$$a_n = \frac{2}{T} \int_0^T x(t) \cos\left(2\pi n \frac{t}{T}\right) dt$$

$$b_n = \frac{2}{T} \int_0^T x(t) \sin\left(2\pi n \frac{t}{T}\right) dt \tag{1.9}$$

The Fourier series can also be expressed in exponential form as

$$x(t) = \sum_{n=-\infty}^{\infty} c_n e^{in\omega_1 t} \tag{1.10}$$

This equation is called the synthesis equation since it gives the mathematical description of the addition or superposition of a number of harmonic signals of different amplitudes, frequencies and phases to get a composite signal. The analysis equation giving the spectral amplitudes is

$$c_n = \frac{1}{T}\int_0^T x(t)e^{-in\omega_1 t}\,dt \tag{1.11}$$

The Fourier coefficients are often displayed in modulus and phase form

$$c_n = (a_n^2 + b_n^2)^{1/2} \qquad \text{and} \qquad \phi_n = \tan^{-1}\left(\frac{-b_n}{a_n}\right) \tag{1.12}$$

It should be borne in mind that no information is gained or lost in transforming a signal from the time domain to the frequency domain and vice versa. Transformation merely enables looking at the signal from a different perspective. Figure 1.5 shows the magnitude and phase spectra of the complex periodic wave depicted in Fig. 1.4.
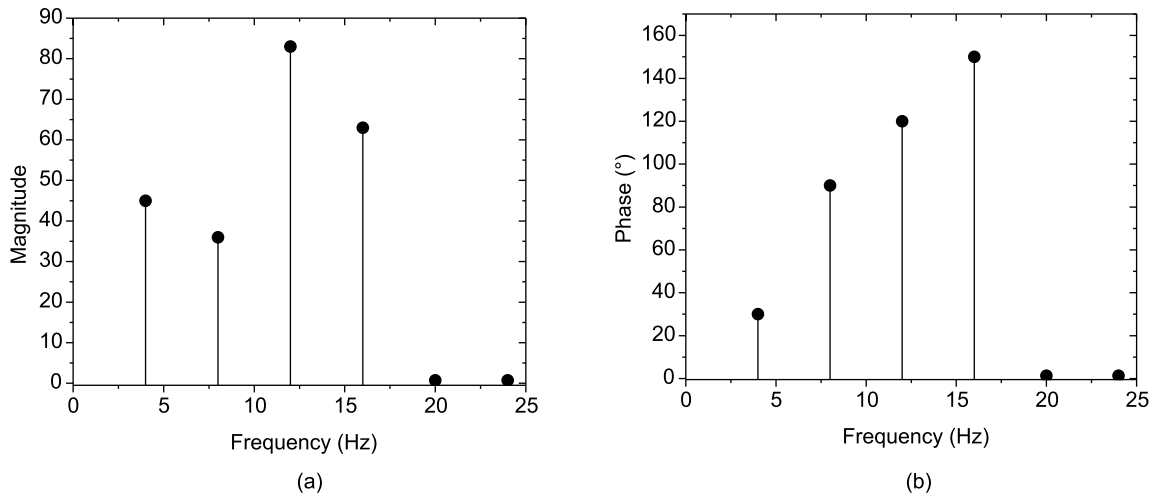
FIGURE 1.5   Spectrum of a complex periodic signal

A periodic signal $x(t)$ may also be represented as

$$x(t) = \sum_{k=-\infty}^{\infty} x_p(t - kT) \tag{1.13}$$

where $x_p(t)$ is the periodic or repetitive block of the signal and T the period of the signal:

$$x_p(t) = \begin{cases} x(t) & t_1 \leq t < t_1 + T \\ 0 & \text{elsewhere} \end{cases} \tag{1.14}$$

1.2.3.2 *Period of a composite signal*   The sum of N periodic continuous-time signals is not necessarily periodic. It is periodic with period T_0 if and only if the following condition is satisfied:

$$\frac{T_0}{T_{0i}} = n_i, 1 \le i \le N \tag{1.15}$$

where T_0 is period of composite signal, T_{0i} = period of the ith signal in the sum and n_i = an integer.

The equivalent necessary and sufficient condition for a sum of periodic continuous-time signals to be periodic is that all the ratios of the period of the first signal in the sum to that of any other signal in the sum be rational, i.e. ratios of integers:

$$\frac{T_{01}}{T_{0i}} = \frac{n_i}{n_1}, \ 2 \le i \le N \tag{1.16}$$

Consider the signal as follows:

$$x(t) = X_1 \cos(2t + \theta_1) + X_2 \cos(5t + \theta_2) + X_3 \cos(7t + \theta_3) \tag{1.17}$$

It is periodic because the ratios of the periods $\dfrac{5}{2}$ and $\dfrac{7}{2}$ are ratios of integers or rational numbers.

1.2.4   Almost Periodic Signals

If the composite signal obtained from a superposition of a number of harmonic signals does not satisfy Eq. (1.16), it is not periodic. Consider a slight variation of the earlier example:

$$x(t) = X_1 \cos(2t + \theta_1) + X_2 \cos(5t + \theta_2) + X_3 \cos(\sqrt{30}t + \theta_3) \tag{1.18}$$

This signal is not periodic and may be termed 'almost-periodic' since $2\pi/\sqrt{30}$ is an irrational number. The time record, as well as the spectrum of this signal, gives the appearance of a periodic signal. When the effects of two or more unrelated periodic phenomena are mixed, the resultant data are almost periodic. A typical example is the vibration produced in an aeroplane with multiple engine propellers, the engines being out of synchronization.

1.2.5   Transient Non-periodic Signals

Signals that fall under this classification are those which are of short duration, but can be described by some suitable functions. Many physical phenomena generate such signals. Typical practical examples of such signals include:

- free damped vibration of a system after the exciting force is removed
- drop in temperature of the water in a kettle after the flame has been turned off
- impact produced by a forging hammer
- force felt by a package when it is dropped on the floor
- stress in an end-loaded cable after the cable has been snapped

Such signals cannot be expressed in terms of the Fourier series as was done for a periodic signal. However, if they are considered to be periodic with period T tending to infinity, the discrete spectrum closes up and gives rise to a continuous spectrum described by the Fourier integral or Fourier transform. The forward transform or analysis equation is given by

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t}\,dt \qquad (1.19)$$

The inverse Fourier transform or synthesis equation is given by

$$x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} x(t)e^{i\omega t}\,d\omega \qquad (1.20)$$

If expressed in terms of frequency f rather than in terms of circular frequency ω, the two equations become identical:

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft}\,dt \qquad (1.21)$$

$$x(t) = \int_{-\infty}^{\infty} X(t)e^{i2\pi ft}\,df \qquad (1.22)$$

Figure 1.6 shows time records of two transient non-periodic signals. Figure 1.7 shows another practical example: the free vibration response of a damped system (exponentially decaying sinusoidal signal) and its spectrum.
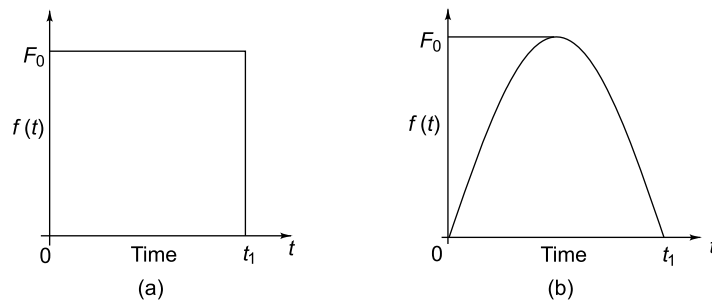
FIGURE 1.6   Transient non-periodic signals: (a) rectangular pulse, (b) half sine pulse
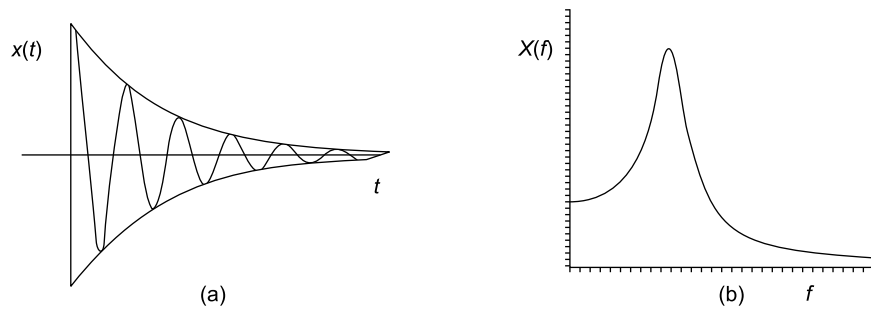
FIGURE 1.7   Typical free vibration response: (a) time record, (b) spectrum

1.2.6   Non-deterministic or Random Signals

While deterministic signals are those which can be described by an explicit mathematical expression, non-deterministic signals or random signals are those that cannot be expressed in the form of an equation.

For such signals, a complete description is possible only with the knowledge of the signal over time $t = -\infty$ to $+\infty$. The physical phenomena giving rise to such non-deterministic data have one thing in common, that is, the unpredictability of their instantaneous values at any future instant of time. If $x(t_1)$ is known, it does not enable one to find out the value $x(t = t_2)$. A large amount of data is therefore necessary to establish reliability. A number of examples for random signals have been given further:

- noise of a jet engine
- height of waves in a choppy sea
- road undulations
- pressure gusts encountered by an airplane in flight
- ground motion during an earthquake
- electrical output of a noise generator
- acoustic pressures generated by air rushing through a pipe
- microroughness of a machined part

Figure 1.8 shows two typical random records.
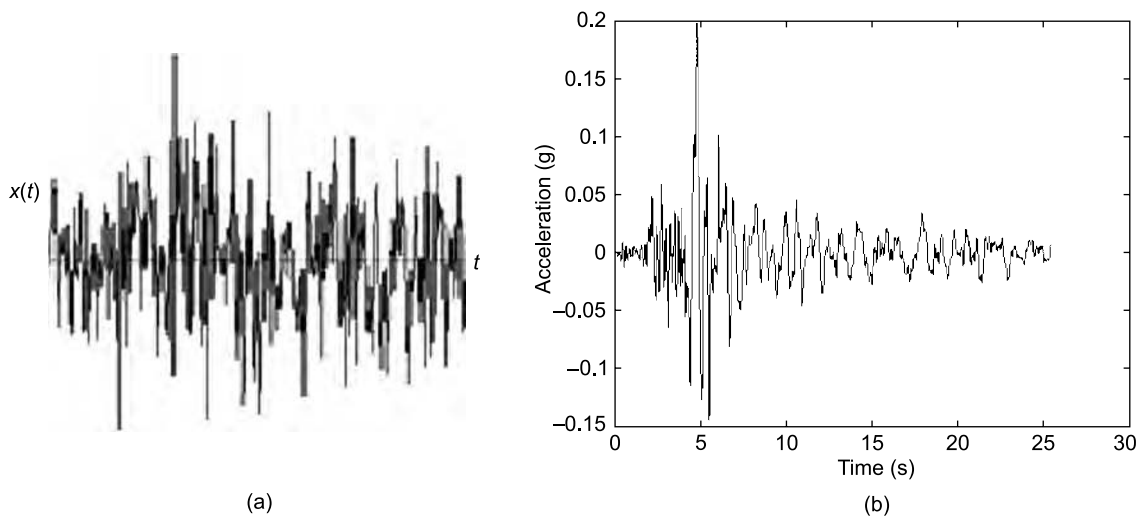
FIGURE **1.8**   Random signals: (a) output of noise generator, (b) accelerogram

1.2.6.1 *Descriptors for random signals*
In spite of irregularities, random functions do exhibit some degree of statistical regularity. Certain averaging procedures can therefore be applied to establish gross characteristics useful in engineering design. Random functions may be represented in terms of:

- mean, mean square and other higher moments and statistical values
- probability density and distribution functions
- autocorrelation functions
- autospectral density functions

Chapter 7 gives a detailed explanation of characterization of random signals.

1.2.7   Signals and Systems

We often study signals to get an idea about system behaviour or performance. When studying a system, say a simple one of the type shown in Fig. 1.9, we have to deal with three quantities:

  (i) input to the system

 (ii) impulse response function (IRF) or frequency response function (FRF) which is a function of system parameters

(iii) output of the system
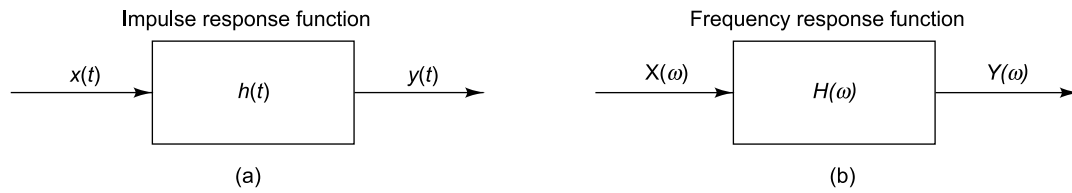
FIGURE 1.9   System model: (a) time domain, (b) frequency domain

    The three operations of common interest are (i) given the input to the system and system parameters, to determine the output; (ii) given the input and output, to characterize the system or find the system parameters; and (iii) knowing the output and the system parameters, to find the input that gives rise to the output. Table 1.1 provides examples of typical systems along with their inputs and outputs.

TABLE 1.1   Examples of signals and systems

System	Signals		System parameters
	Input	*Output*	
Electric circuit	$E(t)$	Current through the circuit $i(t)$, or voltage across resistor $V_R(t)$	Resistor, inductance, capacitance
Road vehicle	Road undulation $y(t)$	Displacement $x(t)$ or acceleration $\ddot{x}(t)$ at driver's location	M_{sprung}, $M_{unsprung}$, K_{tyre}, C_{tyre}, $K_{suspension}$, $C_{suspension}$
Human upper body	Seat vibration $x(t)$	Vibration of spine	Mass, stiffness and damping of bones, ligaments and muscles
Building	Earthquake ground undulation $y(t)$	Displacement of building $x(t)$	M, K and C of various floors of building
Transmission line	Wind force $f(t)$	Displacement of line $x(t)$	M, K and C of line

    When dealing with signals and systems, it is very important to know what approximations are to be made regarding the signals as well as the system. This is largely dictated by the objective of the exercise. If one is interested in finding the output of the electric circuit (shown in Fig. 1.10) that is subjected to 220 V AC power at 50 Hz frequency, then one may model the voltage input as shown in Eq. (1.23) and find the steady-state response of the circuit:

$$v(t) = \sqrt{2}(220)\sin(2\pi \times 50 \times t)\,V, \quad -\infty < t < \infty \tag{1.23}$$
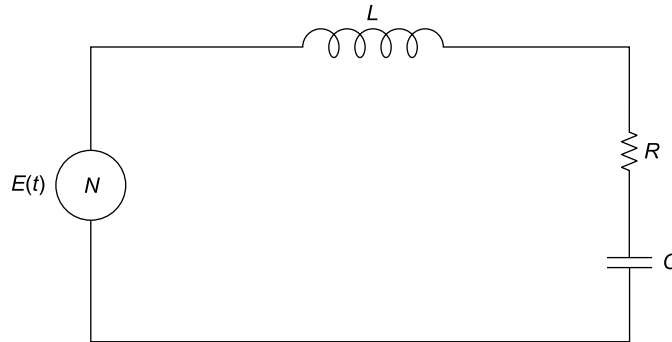
FIGURE 1.10   Electrical system

The fact is that this voltage may not be truly deterministic, keeping in view the frequency and voltage fluctuations, as well as unexpected breaks in supply. But modelling the supply voltage as a harmonic signal would be appropriate for finding steady-state response (voltage across one of the system components, say). On the other hand, if we are interested in finding the fluctuations in voltage across the component of interest, then the input of concern would be the fluctuations in the input, which would then no longer be deterministic. Wherever possible and within allowable error limits, we simplify and try to model periodic signals as simple harmonic, nearly periodic signals as periodic and random signals as stationary and ergodic to reduce computational effort.

Similarly in the modelling of a system, for reasonably small inputs, it may be assumed that the system is operating in its linear range and it is then possible to apply the principle of superposition. But for very large inputs, for which a linear relationship between the inputs and outputs no longer exists, the assumption of a linear system would be wrong. We always try to make simplifications in modelling systems, a stable, linear time-invariant (LTI) model being a very convenient assumption that makes computations fairly simple. We generally start with the simplest model of a system, namely single-input single-output (SISO) model and once sufficient confidence has been built, we can graduate to more complex models such as single-input multiple-output (SIMO) models or multi-input multi-output (MIMO) models.

Chapter

2

Theory of Vibration

2.1 INTRODUCTION

We shall primarily be concerned with the measurement and analysis of vibration and acoustic signals in this book; hence for a better understanding, some fundamental theories about vibration are discussed in this chapter. The study of vibration is concerned with the oscillatory motions of bodies. Most engineering structures and machines experience vibrations which may be inherent and unwanted and occur because of the dynamic effects of manufacturing tolerances, clearances, rolling and rubbing contact between machine parts, out-of-balance forces in rotating and reciprocating members or due to resonance of structural parts as in the case of vibrations of a vehicle induced by ground undulations. On the other hand, some vibrations perform a useful job and are purposely introduced, as in the case of a mechanical sieve. We also generate vibration intentionally in concrete compactors, ultrasonic cleaning baths, rock drills, etc. Vibration shakers/exciters are designed to impart a controlled level of vibration energy to machines/components/structures to examine their response and to ascertain if they can withstand the operational vibration environments. Irrespective of whether the vibrations are inherently or deliberately produced, an accurate description of the vibration is required through measurement and analysis.

The parameters of concern in the study of vibration are amplitudes, which may be expressed as displacement, velocity or acceleration and excitation forces. In the metric system, the practical units for these quantities would be mm or μm (one thousandth of a mm) for displacement, mm/s for velocity, m/s^2 or g (one g being equal to 9810 mm/s^2) for acceleration and newton for force.

The frequencies of interest in any vibration study are the natural frequencies of the system and the excitation frequencies, which can be related to the running speeds and their harmonics. The natural frequencies depend on the system masses and stiffness. Table 2.1 shows some typical values for the natural frequencies of some practical systems. The values given are indicative only; the exact values would depend on the system geometry and material properties. The exciting or forcing frequencies of mechanical systems are usually in the range 1–160,000 rpm (1/60–2667 Hz) as shown in Table 2.2.

Any system, which possesses mass and elasticity is capable of vibrating. Most practical structures around us have multiple degrees-of-freedom (DOF) in the sense that a large number of independent coordinates are required to completely describe the motion of these systems. However, even very complicated multi-degree-of-freedom (MDOF) systems can be reduced to the simplest model, namely that of a single-degree-of-freedom (SDOF) system. The study of vibration encompasses free and forced vibration. Free vibration is due to forces inherent in the system, while forced vibration is due to externally

impressed forces. If the frequency of excitation of a structure coincides with any of its natural frequencies, resonance occurs. Failure of major structures like bridges, aeroplane wings and buildings is due to resonance.

TABLE 2.1   Typical natural frequencies

System	Natural frequencies (Hz)
Offshore oil rig	1 or 2
Bridge	1–10
Car—bounce	1
Car—pitch	2.5
Military tank—bounce	0.5–1.5
Motorcycle weave	0.2–4
Motorcycle wobble	4–9
Ship	0.5–10
Multi-storeyed building	1–10
Railway car—bounce	1
Railway bogie—bounce	5
Railway wheelset—bounce	200
Human trunk	2.5–5
Human vertebrae	100s
Power train of sports utility vehicle (SUV)	50–300
SUV compartment cavity	100s
Hard disk drive	100–1000s
Turbine blade low pressure stage	10–1000s
Turbine blade high pressure stage	100–1000s
Inner/outer races of rolling element bearings	Few 1000s

TABLE 2.2   Typical running speeds

System	Normal running speed (rpm)
Kiln	1
Ball mill	15–20
Low speed marine diesel engine	60–480
Compressors	500–3000
ID/FD fan	600–1200
Induction motor (4% slip)	1440 for 50 Hz supply frequency
DC motor	100–6000
Steam turbine	3000 for 50 Hz supply frequency
Car (including race car) engine	600–15,000
Hard disc drives	5400–15,000
Kitchen blender motor	0–20,000
High speed spindles of lathes	60,000 and above
Turbochargers	150,000
Dental drill	Up to 600,000

2.2 SINGLE-DEGREE-OF-FREEDOM SYSTEM

The motion of such a system can be described by a single coordinate, which could correspond to translation or rotation. Reduction of an MDOF system to one with an SDOF is done by lumping all masses to an equivalent one and by finding the equivalent stiffness and damping of the system.

2.2.1 Translatory System

The SDOF vibratory model consists of a non-elastic mass, which is suspended from a linear massless spring (Fig. 2.1a). Figure 2.1(b) shows the free body diagram of the mass of a simple undamped system, which is assumed to move in the vertical direction only and which is described by a single coordinate x.

When the system is set into motion, it vibrates at a frequency f_n called the natural frequency, which is a function of its mass and stiffness parameters. If the deformation of the spring corresponding to the static equilibrium position is assumed to be Δ, the spring force $k\Delta$ is equal to the gravitational force $W = mg$ acting on

FIGURE 2.1 Undamped single-degree-of-freedom system: (a) translatory SDOF system, (b) free body diagram of mass m

the mass. Therefore, by measuring the dynamic displacement x from the static equilibrium position, the spring force and the gravitational force nullify each other and need not be considered in the equations of motion as long as the spring is linear. If the system is disturbed by moving the mass downwards by a distance x and released, the spring k deforms in proportion to the force acting on it according to Hooke's Law. The mass m experiences acceleration proportional to the force acting on it. While drawing the free body diagram, a fictitious inertia force equal to $m\ddot{x}$, which is in a direction opposite to that of the acceleration, is introduced on the mass, to reduce the problem to one of statics. Since there are only two forces keeping the mass in equilibrium, their sum must be equal to zero. Hence applying Newton's second law of motion to the mass, we obtain

$$m\ddot{x} + kx = 0 \tag{2.1}$$

This equation is a homogeneous second-order linear differential equation. Its solution can be derived mathematically and is found to be harmonic. It can be expressed in the form

$$x(t) = X \exp(i\omega t) \tag{2.2}$$

The system has only one mode of vibration, which is defined by the motion along the x coordinate and which is obtained by substituting Eq. (2.2) in (2.1) to give

$$(-m\omega^2 + k)X = 0 \tag{2.3}$$

or

$$\omega = \pm \left(\frac{k}{m} \right)^{1/2} = \pm \omega_n \tag{2.4a}$$

Here ω_n is the natural frequency in rad/s. The natural frequency in Hz is

$$f = \pm \frac{1}{2\pi}\left(\frac{k}{m}\right)^{1/2} = \pm f_n \tag{2.4b}$$

The period of oscillation in seconds is

$$T = \frac{1}{f_n} = \frac{2\pi}{\omega_n} \tag{2.4c}$$

It is to be noted that the natural frequency and the period depend only on the mass and stiffness of the system, which are system properties. The general solution to Eq. (2.1) is of the form

$$x(t) = A \exp(i\omega_n t) + B \exp(i\omega_n t) \tag{2.5}$$

The constants A and B can be obtained by substituting the known values of initial conditions of the problem, i.e. the known values of the initial displacement and initial velocity.

2.2.2   Torsional System

Besides translational motion, we often come across angular oscillations, e.g. a section of a shaft moving clockwise or anticlockwise in relation to the other cross sections. Such a problem is said to be one of torsional vibration. The whole shaft rotates with uniform speed with superposition of angular accelerations or decelerations of some sections, causing twisting or untwisting of the shaft. This is not perceptible to the naked eye, unlike in the case of longitudinal or lateral oscillations. However this can be discerned as a frequency modulated tone by an experienced ear. Such problems are prevalent in machines with rotating parts such as punch presses, shearing machines, reciprocating compressors, pumps, turbines and reciprocating engines in which the input or output torque is fluctuating. The results described above for translational oscillations are applicable to all SDOF systems with rotation, for which case the mass is to be replaced by a mass moment of inertia and the linear spring by a torsional spring.

Let J be the mass moment of inertia of the rotor about its axis and q the torsional stiffness (torque needed per unit angular displacement) of the shaft in Fig. 2.2. Let θ denote the angular rotation of the rotor with respect to a section of uniform velocity. Applying Newton's laws, the sum of the torques acting on the rotor must be equal to zero. Hence

$$J\ddot{\theta} + q\theta = 0 \tag{2.6}$$

Assuming a harmonic solution as in the case of translation, we get

$$-J\omega_n^2\Theta + q\Theta = 0 \tag{2.7}$$

The natural frequency of torsional oscillation is

$$\omega_n = \sqrt{\frac{q}{J}} \text{ rad/s} \tag{2.8}$$

It is to be noted here that the rotor and the shaft being in series, the component with the smaller torsional stiffness, namely the shaft, is taken as the elastic member and the component with the larger mass moment of inertia, namely the rotor, is assumed to contribute towards the rigidity, though in reality both these members have both inertia and elasticity to varying degrees. Typically the angular
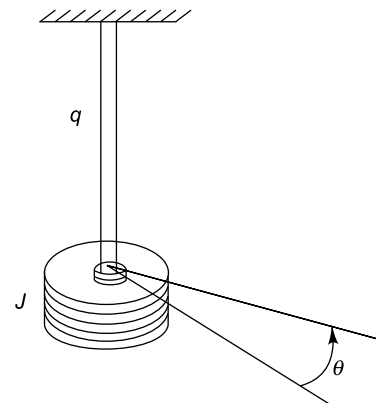
FIGURE 2.2   Single-degree-of-freedom torsional system

displacements θ that we come across in practice are very small, around 2° or 0.05 radians at the most; the shear strain and resulting dynamic shear stress nevertheless can still be large. The lowest torsional natural frequency of practical systems may be of the order 70–2000 rad/s (11–318 Hz approximately). As in the case of lateral vibration, there are cases where the torsional natural frequency of a system is in the vicinity of the forcing frequency, causing resonance. A typical example is a steam turbine for power generation where natural frequencies of some of its components have been observed to be close to the running speed (3000 rpm for 50 Hz).

2.2.3   Damping

In any practical vibratory system, the physical parts of the system contributing to the inertia and elasticity can easily be identified. In addition to mass and stiffness, there is a third very important parameter, which is responsible for the dissipation of energy in a vibrating system. This parameter is called damping and is not so explicitly seen, since most mechanical structures are inherently lightly damped. Damping is obvious only in cases like the shock absorber in a vehicle or the dampers in an automatic door closer or a gun recoil system. However, it is clearly seen that the vibration amplitude keeps falling in the case of free vibration after a system has been set into motion and that amplitude at resonance during a forced vibration does not shoot to infinity (because the energy supplied by the excitation is balanced by loss of energy), indicating that there is some mechanism present in all vibratory systems for the dissipation of energy. Ignoring damping can give rise to grossly wrong results in a forced vibration study, especially at frequencies close to the natural frequencies, though not at frequencies far away from resonance.

Damping is brought about by a large number of phenomena like intermolecular friction or air resistance. It is difficult to get a closed form expression for the combined effect of these phenomena and one has to resort to simplified models or experimentally obtained values instead. Three damping models, which are commonly used are as follows:

 (i) Viscous damping model
 (ii) Structural damping model
 (iii) Coulomb damping model.

2.2.4   Free Damped Vibration

In the model shown in Fig. 2.1, there is no damping. To simulate practical situations though, one needs to include damping as described in the earlier section. The viscous damping model is the most popular one, the reason being that the introduction of this linear damper is very simple and it also adequately explains the response observed experimentally. Besides, it gives rise to a linear differential equation with constant coefficients, the solution to which is well known and simple. Referring to the SDOF system with damping and the free body diagram shown in Fig. 2.3, the equation of motion is a homogeneous differential equation of the form

$$m\ddot{x} + c\dot{x} + kx = 0 \tag{2.9}$$
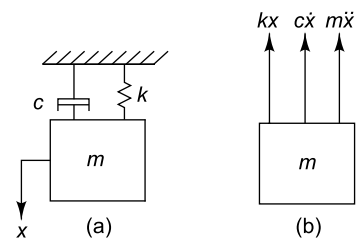
FIGURE 2.3   Single-degree-of-freedom system with damping: (a) schematic drawing, (b) free body diagram

Traditionally a solution of the form shown below is assumed

$$x(t) = X \exp(i\omega t) \tag{2.10}$$

On plugging in this solution into Eq. (2.9), the roots of the quadratic equation obtained are given by

$$\omega_{1,2} = \frac{-\dfrac{c}{m} \pm \sqrt{\left(\dfrac{c}{m}\right)^2 - \dfrac{4k}{m}}}{2} = \sqrt{\frac{k}{m}}\left[\frac{-c}{2\sqrt{km}} \pm \sqrt{\left(\frac{c}{2\sqrt{km}}\right)^2 - 1}\right] \tag{2.11a}$$

Since $\omega_n = \sqrt{k/m}$ the free undamped natural frequency, the damped natural frequency is given by

the equation below for $\left(\dfrac{c}{m}\right)^2 < \dfrac{4k}{m}$.

$$\omega_{1,2} = \omega_n\left[\frac{-c}{2\sqrt{km}} \pm i\sqrt{1 - \left(\frac{c}{2\sqrt{km}}\right)^2}\right] \tag{2.11b}$$

$c/2\sqrt{km}$ is dimensionless and can be expressed as a quantity $\zeta = c/c_c$ known as damping ratio, where c_c is known as critical damping coefficient and is equal to $2\sqrt{km}$. Then

$$\omega_{1,2} = -\zeta\omega_n \pm i\omega_n\sqrt{1 - \zeta^2} \tag{2.12}$$

Hence the solution to Eq. (2.9) is given by

$$x = e^{-\zeta\omega_n t}\left[c_1 \cos(\omega_n\sqrt{1-\zeta^2}\,t) + c_2 \sin(\omega_n\sqrt{1-\zeta^2}\,t)\right] \tag{2.13a}$$

The constants c_1 and c_2 may be obtained from the initial displacement and initial velocity. With initial conditions $x(0)$ and $\dot{x}(0)$, the above equation can be simplified as

$$x = e^{-\zeta\omega_n t}\left[x(0)\cos\left(\omega_n\sqrt{1-\zeta^2}\,t\right) + \frac{\dot{x}(0) + \zeta\omega_n x(0)}{\omega_n\sqrt{1-\zeta^2}}\sin\left(\omega_n\sqrt{1-\zeta^2}\,t\right)\right] \tag{2.13b}$$

For most engineering problems considering materials like mild steel or aluminium, $\zeta << 0.1$ and hence $\sqrt{1-\zeta^2}$ can be taken as 1. Vibrations with such low ζ are said to be underdamped and the corresponding motion is oscillatory. The response can then be simplified to

$$x = e^{-\zeta\omega_n t}(c_1 \cos\omega_n t + c_2 \sin\omega_n t) \tag{2.14}$$

ζ is found to be frequency dependent and of the order of 10^{-1}–10^{-3} for materials like mild steel and aluminium. It may also be noted that when $\zeta = 1$, $c = 2\sqrt{km}$. This damping is known as critical damping. When $\zeta = 1$, the two roots of Eq. (2.11) are equal. The system comes to rest in the minimum possible time with no oscillation as given by the solution

$$x = e^{-\omega_n t}(c_1 + c_2 t) \tag{2.15}$$

Critical damping is incorporated in systems where oscillations are to die down within the shortest period once the system is excited, as in the case of the recoil system in the firing of a bullet or closing of

a door using an automatic door closer. In instruments, eddy current damping is deliberately introduced to have a damping ratio of 0.65–0.7 to increase the useful frequency range of accelerometers and vibrometers. The case corresponding to $\zeta > 1$ is said to be overdamped or non-oscillatory motion; the free vibration record shows an exponentially decreasing function of time and is aperiodic for this case.

2.2.4.1   *Determination of damping ratio ζ by logarithmic decrement method*   One of the simplest ways of determining the damping ratio of a system is by the logarithmic decrement method applied to the decaying time response of the system. This decaying response curve for an MDOF system contains a mixture of modes. The unfiltered signal may be used as such for the determination of damping ratio of the first mode. However, to find the damping associated with the higher modes using this method, the time decay data will have to be band-pass filtered before applying the logarithmic decrement method and is therefore a little cumbersome. It might be more convenient to use the frequency domain based half power method instead.

Equation (2.13) has been plotted in Fig. 2.4. Here the amplitudes of successive waves are separated in time by $T = 2\pi/\omega_d$ and the envelope of the decaying oscillation is an exponential curve, which lends itself to a convenient definition of damping. The ratio of any two successive amplitudes is given by
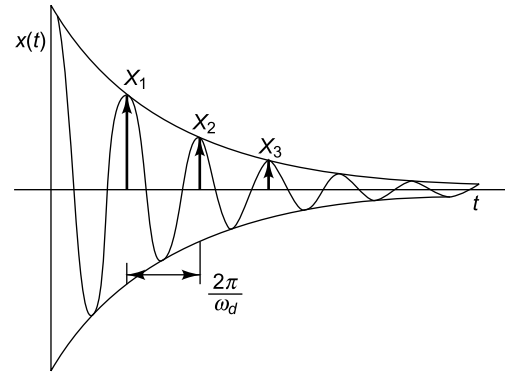
Figure 2.4   Damped system response

$$\frac{X_1}{X_2} = e^{\zeta\omega_n\left(t-t+\frac{2\pi}{\omega_d}\right)} = e^{2\pi\zeta} \tag{2.16}$$

where X_1 and X_2 are the amplitudes of the first and second cycles, respectively. The natural logarithm of the ratio of any two successive displacements known as logarithmic decrement is given by

$$\delta = \ln\left(\frac{X_1}{X_2}\right) = \frac{2\pi\zeta}{\sqrt{1-\zeta^2}} \tag{2.17a}$$

The logarithmic decrement can also be defined as

$$\delta = \frac{1}{n}\ln\left[\frac{X_1}{X_{n+1}}\right] \tag{2.17b}$$

where X_{n+1} is the amplitude of the $(n + 1)^{\text{st}}$ cycle. Equation (2.17b) is especially convenient for computation of δ and hence ζ from measurements, since measuring two successive amplitudes (as required in Eq. 2.17a) without error may be difficult for the reason that the two values are very close. The relationship between the logarithmic decrement and the damping ratio of the system, ζ is

$$\zeta = \frac{\delta}{\sqrt{(2\pi)^2 + \delta^2}} \simeq \frac{\delta}{2\pi} \tag{2.17c}$$

2.2.5 Damping Models

In the mathematical formulation discussed so far, damping has been modelled as arising out of a viscous force, and hence proportional to velocity. However, this does not mean that in reality the physical damping mechanism is viscous in nature; it is simply a convenient mathematical model which may not be a realistic representation. Many practical structures exhibit Coulomb or structural damping resulting from looseness of joints, internal strain and other complex phenomena. However, most of these mechanisms can be modelled by an equivalent viscous damping component, which can be conveniently used to characterize a system when using a linear model.

2.2.5.1 Viscous damping model   The most common damping model used is that of viscous damping described by a force proportional to the velocity of vibration. This model is based on the viscous resistive force of the oil when a piston moves in an oil filled cylinder. The damping force is expressed by the equation

$$F_d = c\,\dot{x} \tag{2.18}$$

where c is a constant of proportionality called damping coefficient and has the dimension of force per unit velocity (Ns/m). This damping model may not always result in the best idealization, since it assumes that the damping is frequency independent.

2.2.5.2 Structural/hysteretic damping   Many real life structures exhibit a frequency dependent damping behaviour. Stiffness of materials like rubber can be expressed in the form $k(1 + ih)$, h being the coefficient of hysteretic damping present in the system. The equation of motion when such materials are used as a spring is given by

$$m\,\ddot{x} + k\,x\,(1 + ih) = 0 \tag{2.19}$$

When this is compared with Eq. (2.9), it is seen that the term $c\omega$ is replaced by kh for harmonic vibration. This damping model poses difficulties for a rigorous free vibration analysis and is generally used for forced vibration analysis. Table 2.3 shows the structural damping for typical materials that we come across in engineering practice. These values are rough indicative values and may change depending on the actual constitution of the material.

TABLE 2.3   Structural damping of common materials

Material	*h*
Aluminium	0.0002–0.001
Steel	0.01–0.06
Cast iron	0.03–0.06
Concrete	0.15
Masonry structures	0.3–0.6
Rubber	0.1–0.5
Cork	0.1–0.2

2.2.5.3 Coulomb damping   In such a damping model, the damping force results from the friction due to the sliding of two dry surfaces and is equal to μN where μ is the coefficient of friction and N is the normal force. Once the motion is initiated, this force is independent of velocity. The sign of the damping

force is opposite to that of velocity. The decay in vibration amplitude per cycle may be determined using work—energy principle. Figure 2.5 shows the half cycle starting at time t, when velocity is zero and amplitude is X_1. X_{-1} is the amplitude at the end of this half-cycle. Equating the work done on mass m to the change in kinetic energy,

$$\frac{1}{2}k(X_1^2 - X_{-1}^2) - F_d(X_1 + X_{-1}) = 0 \tag{2.20a}$$

or $\quad \frac{1}{2}k(X_1 - X_{-1}) = F_d \tag{2.20b}$

This equation shows that the decrease in vibration amplitude per half cycle is $\dfrac{2F_d}{k}$ and therefore for the full cycle is

$$X_1 - X_2 = \frac{4F_d}{k} \tag{2.21}$$

It is clearly seen from Fig. 2.5 that the vibration amplitude decreases linearly with time.
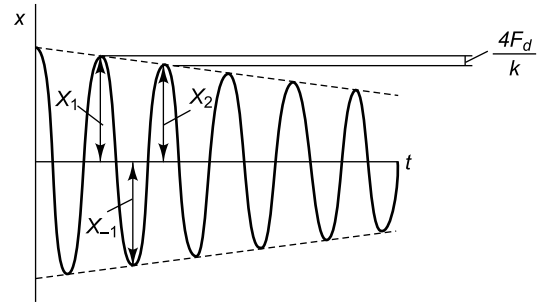
FIGURE 2.5   Vibration decay with Coulomb damping

2.2.5.4   *Energy dissipated by damping and equivalent viscous damping*   The energy dissipated per cycle of oscillation due to a damping force F_d can be calculated as

$$W_d = \oint F_d dx \tag{2.22}$$

The force-displacement curve will depend on the type of damping present, but will always enclose an area called the hysteresis loop, which is proportional to the energy lost per cycle W_d. Such a loop is shown in Fig. 2.6(a) for the case of viscous damping. Figure 2.6(b) shows the loop for a Voigt model where the spring is parallel to the damper (Fig. 2.3a). In the case of harmonic vibration of the form described in Eq. (2.2) with viscous damping

$$W_d = \oint c\dot{x}\,dx = \oint c\dot{x}^2 dt = \pi c\omega X^2 \tag{2.23}$$
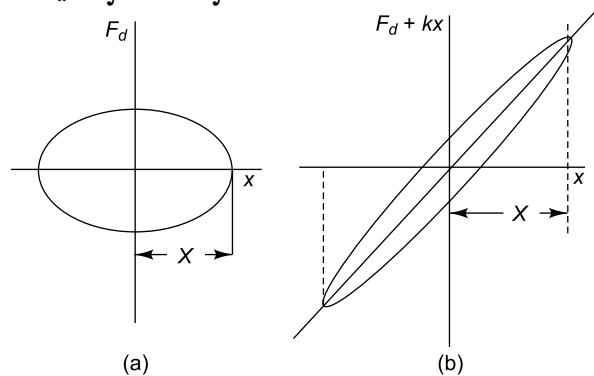
FIGURE 2.6   Hysteresis loops for energy dissipated due to viscous damping: (a) viscous damping model, (b) Voigt model

For any damping model other than viscous, the equivalent damping C_{eq} can be obtained by equating the energy dissipated by the viscous damping to that of the non-viscous damping force. Harmonic motion is implied in this equation.

$$\pi C_{eq} \omega X^2 = W_d \qquad (2.24)$$

Here W_d corresponds to the work done per cycle for the particular type of damping force considered.

2.2.6 Forced Vibration

As in the case of free vibration, here also we shall study the undamped SDOF system first, followed by the damped system. The simplest case of forced vibration is that with harmonic excitation, for which the steady state response is also harmonic at the forcing frequency.

F cos ωt (External force)

FIGURE 2.7   Forced vibration of undamped system

2.2.6.1   *Without damping*   Let us now consider a simple undamped SDOF system subjected to excitation forces which may come from say, a drive motor in a machine. The governing differential equation for such a system (Fig. 2.7) would then be given by

$$m\ddot{x} + kx + F \cos \omega t \qquad (2.25)$$

The solution to this equation has two parts: we have the homogeneous differential equation, which is the complementary function corresponding to the free damped vibration of the system when $F(t)$ is equal to zero, and which would disappear in a matter of seconds (even for very small values of ζ). The second part of the solution is the particular integral, corresponding to $F(t)$ not equal to zero and is the response to the excitation forces and is called the steady-state response. The particular solution is obtained by substituting $x = X \cos \omega t$ in Eq. (2.25) to give

$$-m\omega^2 X + kX = F \qquad (2.26)$$

Hence
$$X = \frac{F}{k - m\omega^2} = \frac{(F/k)}{(1 - \eta^2)} \qquad (2.27)$$

where
$$\eta = \frac{\omega}{\omega_n} = \frac{\omega}{\sqrt{k/m}} \qquad (2.28)$$

η is known as the frequency ratio, the ratio of the external forcing frequency to the natural frequency of the system. Figure 2.8 shows the response as a dynamic magnifier, i.e. ratio of dynamic displacement to static displacement as a function of the frequency ratio. The static deflection $X_{st} = (F/k)$ corresponds to $\eta = 0$. When $\eta = 1$, the forcing frequency is equal to the natural frequency; this corresponds to a condition of resonance and the dynamic magnifier becomes ∞. As η becomes very large, i.e. as the forcing frequency becomes much higher than the natural frequency, the dynamic deflection tends to zero.

2.2.6.2   *With damping*   Referring to Fig. 2.9 (a), the governing differential equation can be written as

$$m\ddot{x} + c\dot{x} + kx = F \cos \omega t \qquad (2.29)$$

Assuming a harmonic solution of the form

$$x = X \cos (\omega t - \phi) \qquad (2.30)$$
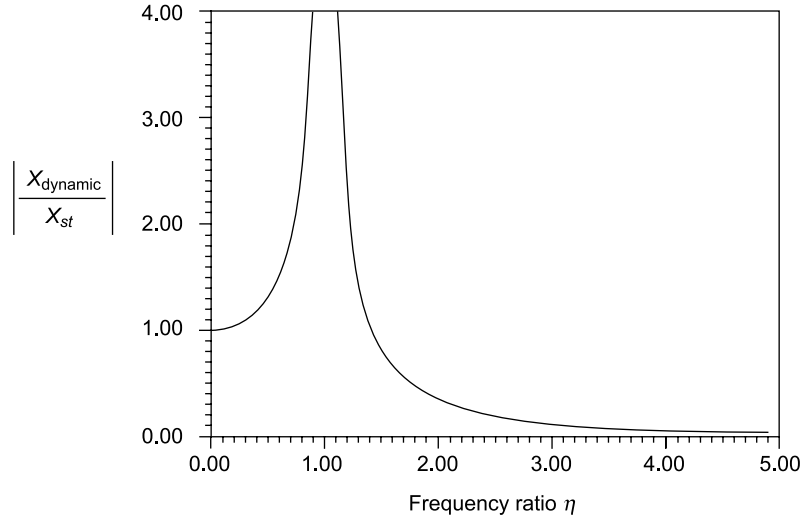
FIGURE 2.8 Dynamic magnifier for various frequency ratios

Eq. (2.30) can be rewritten as

$$(-m\omega^2 + ic\omega + k)X = F \tag{2.31}$$

Figure 2.9(b) shows the vectors representing the inertia, damping and elastic forces, as well as the external force in the above equation.

Hence the particular solution can be expressed as

$$X = \frac{F \cos \omega t}{(k - m\omega^2 + ic\omega)} \tag{2.32}$$

Dividing the numerator and denominator by k and using the non-dimensional values $\zeta = c/c_c$, $\eta = \omega/\omega_n = \omega/\sqrt{k/m}$, we get the expression for the dynamic magnifier as

FIGURE 2.9 Forced vibration of damped SDOF system: (a) schematic of SDOF system, (b) vector diagram showing various forces

$$\frac{X_{\text{dynamic}}}{X_{\text{static}}} = \frac{X}{(F/k)} = \frac{1}{\sqrt{(1-\eta^2)^2 + (2\zeta\eta)^2}} \tag{2.33a}$$

Here X_{static} is the static displacement. The angle of lag ϕ is given by

$$\tan \phi = \frac{2\zeta\eta}{(1-\eta^2)} \tag{2.33b}$$

Figure 2.10 indicates the variation of dynamic magnifier and phase angle ϕ as a function of η. At $\eta = 1$ (resonance), the angle of lag is always 90° for any damping ratio. For values of $\eta << 1$, the phase angle is close to 0°, whereas it is close to 180° for $\eta >> 1$. At resonance the dynamic magnifier is equal to $(1/2\zeta)$.
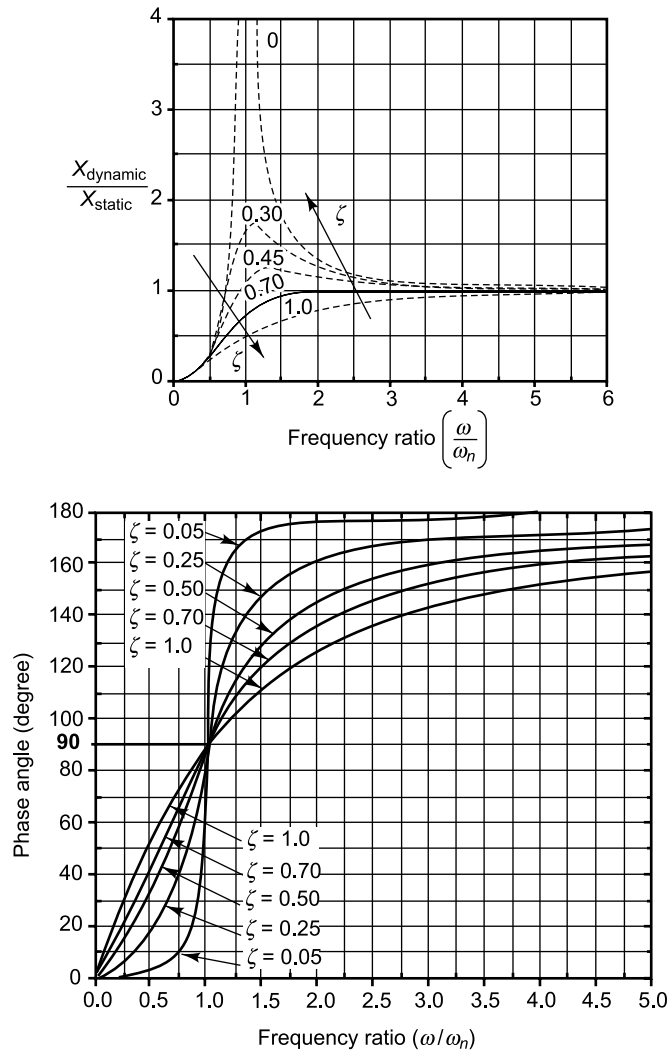
Figure 2.10   Dynamic magnifier and phase angle for forced vibration with damping

For hysteretic damping, the dynamic magnifier $(1/2\zeta)$ at resonance gets replaced by $(\sqrt{km}/2c) = (1/h)$; or in other words, it is nothing but the reciprocal of the hysteretic damping coefficient.

2.2.6.3   *Evaluation of damping by half-power method*   One common way of determining damping ratio is by measuring the frequency bandwidth, between points on the response curve, for

which the response is some fraction of the resonant response of the system. The convention is to consider two frequencies on either side of resonant frequency where the amplitude of the response is $1/\sqrt{2}$ times the maximum response. The bandwidth corresponding to these frequencies is referred to as the 'half-power bandwidth,' a term borrowed from the analysis of electrical systems in which the amplitude squared is a measure of the electric power. This amplitude ratio of $1/\sqrt{2}$ corresponds to a change in amplitude of $20\log_{10}\left(1/\sqrt{2}\right) = -3.01$ dB or in other words a reduction of 3 dB. Thus a measurement associated with an amplitude ratio of $1/\sqrt{2}$ is frequently referred to as '3 dB bandwidth' or half power bandwidth and is given by $\Delta\omega = \omega_2 - \omega_1$ (Fig. 2.11). The damping ratio can be got from the following formula:

$$\zeta = \frac{\Delta\omega}{2\omega_n} \tag{2.34a}$$

and

$$\zeta = 2h \tag{2.34b}$$

More generally, the damping of a system can be determined from the bandwidth between any two points A and B associated with any amplitude ratio $1/n$, where n is greater than 1. Although there are practical limits of measurements, values of n greater than $\sqrt{2}$ or less than $\sqrt{2}$ can occasionally be used.
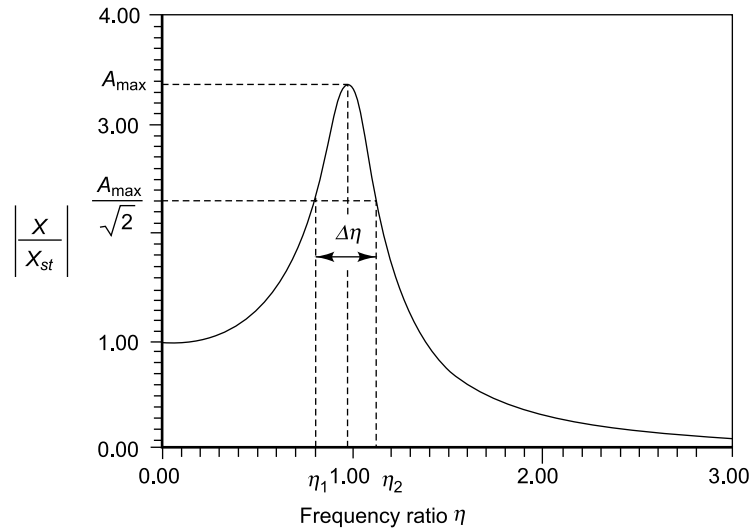
FIGURE 2.11   Determination of damping from half-power method

2.2.7   Response with Unbalanced Forces

One of the most important applications of forced vibration is the case of vibration due to rotating unbalances. All manufactured rotors have some amount of eccentricity. There is no rotor which has zero eccentricity. This may be due to the presence of keys, casting defects and due to non-uniformity in welding. Besides, when three or more parts are assembled to form a combined rotor, the individual eccentricities due to the permitted tolerances in assembly may add up to worsen the quality of manufacture

of the combined rotor. Let us consider the rotor of mass m shown in Fig. 2.12. Let it have an eccentricity of e mm (a quality of balance ωe for a rated speed of ω rad/s).
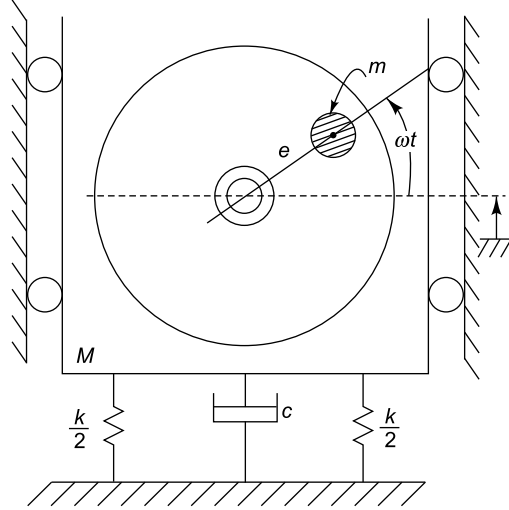
FIGURE 2.12   Unbalanced rotor

Let the displacement of the non-rotating mass $M - m$ be x. Then the displacement of the rotor is obtained as $x + e \sin \omega t$. The equation of motion is given by

$$(M - m)\ddot{x} + m\frac{d^2}{dt^2}(x + e \sin \omega t) = -c\dot{x} - kx \qquad (2.35a)$$

This can be rewritten as

$$M\ddot{x} + c\dot{x} + kx = me\omega^2 \sin \omega t \qquad (2.35b)$$

This is the expression for forced vibration with excitation force amplitude $me\omega^2$. The steady state displacement response can hence be written as

$$X = \frac{me\omega^2}{[(k - M\omega^2)^2 + (\omega c)^2]^{1/2}} \qquad (2.36)$$

This can be reduced to

$$\frac{MX}{me} = \frac{\eta^2}{[(1 - \eta^2)^2 + (2\zeta\eta)^2]^{1/2}} \qquad (2.37)$$

The variation of MX/me with η is shown in Fig. 2.13 for zero damping. When $\eta = 0$ (when the rotor is not rotating), $X = 0$. When $\eta = 1$, the response is infinity.
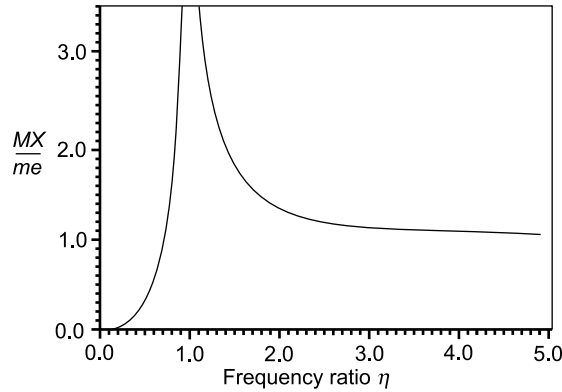
FIGURE 2.13   Response of an unbalanced rotor

2.2.7.1   Critical speed and quality of balance   Manufactured rotors are grouped under different qualities of balance, depending on the order of eccentricity present and specific expectations of these rotors. International standards on quality are shown in Table 2.4. The most superior quality is G 0.4 and the most inferior G 4000. The number 4 or 4000 stands for the peripheral velocity of the centre of gravity of the rotor which is equal to ωe in mm/s, where ω is the rotor speed in rad/s and e the eccentricity in mm. A rotor of quality G 1 running at a speed of 400 rpm (approximately 40 rad/s) must have been manufactured with an eccentricity not exceeding (1/40), i.e., 0.025 mm. Likewise a rotor of quality G 40 running at a speed of 300 rpm (30 rad/s) can have a maximum eccentricity of (40/30), i.e. 1.33 mm. For rotors running at high speeds to meet precision applications, the eccentricity permitted is very low (Fig. 2.14).

TABLE 2.4   Quality of balance for rigid rotors: ISO 1940

Quality grade	ωe (mm/s)	Application
G 4000	4000	Crank shafts of rigidly mounted slow marine diesel engines with uneven number of cylinders
G 1600	1600	Crank shaft drives of rigidly mounted large two cycle engines
G 630	630	Crank shaft drives of rigidly mounted large four cycle and elastically mounted marine diesel engines
G 250	250	Crank shaft drives of rigidly mounted fast four cylinder engines
G 100	100	Crank shaft drives of fast diesel engines with six and more cylinders, engines for cars, trucks and locomotives
G 40	40	Car wheels, wheel rims, crank shafts of elastically mounted fast four cycle engines with six and more cylinders
G 16	16	Drive shafts with special requirements as in cars, trucks, locomotives, gasoline and diesel engines, crushing machines, agricultural machinery
G 6.3	6.3	Parts of process plant machines, turbine gears, centrifugals, impellers and armatures
G 2.5	2.5	Gas and steam turbines, turbo generators, compressors, small electric armatures
G 1	1	Tape recorders, phonographs, grinding machines, small armatures with special requirements
G 0.4	0.4	Spindles, discs and armatures of precision grinders and gyroscopes
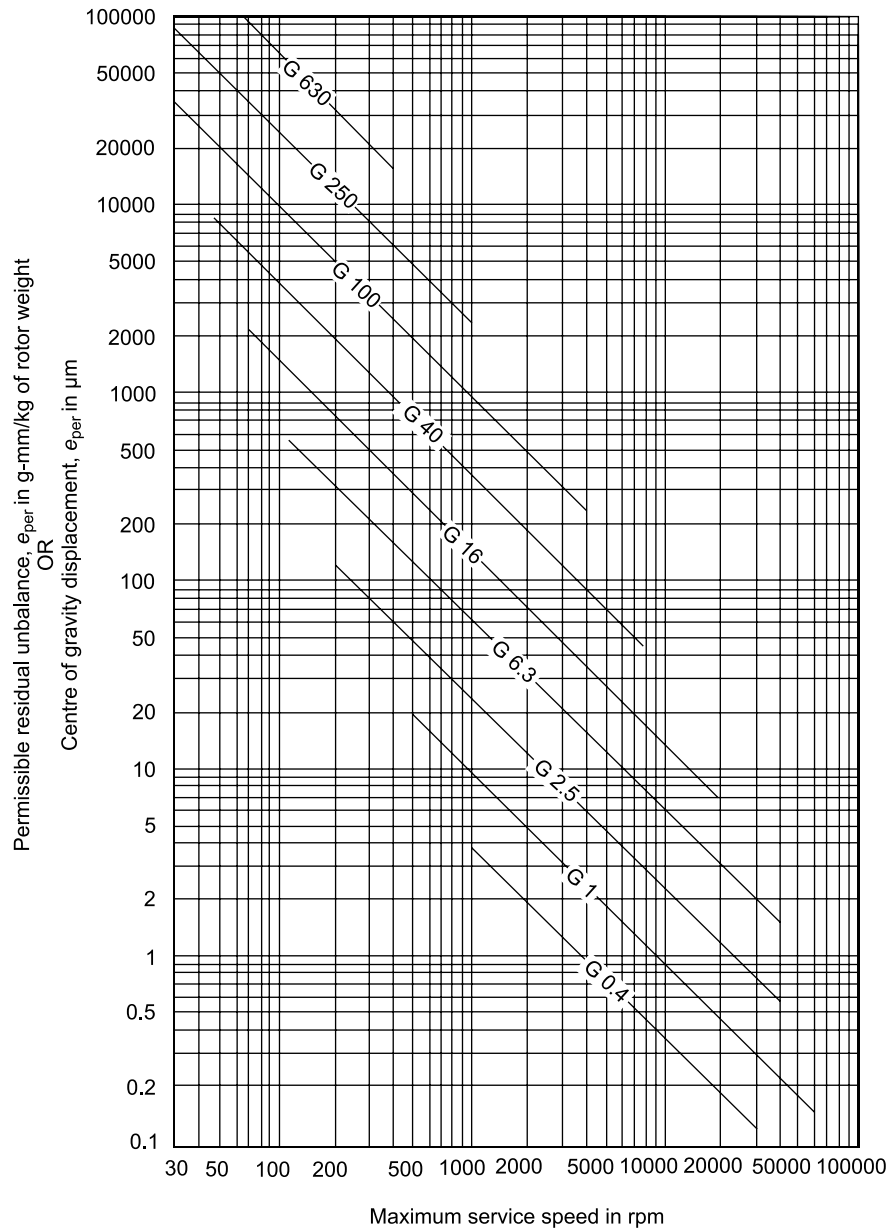
FIGURE 2.14 Maximum permissible residual unbalance corresponding to various balancing quality grades (Courtesy of ISO 1940/1)

2.2.8 Vibration Isolation

Vibration isolation is a very important problem to be dealt with in dynamic systems and is of concern in steady state disturbances. There are two types of isolation problems present:

(i) Force isolation, in which it is to be ensured that a machine with inherent vibratory forces is mounted in such a way that it does not transmit the disturbances to the surroundings and

(ii) Displacement isolation, in which it is to be made sure that a sophisticated instrument which is to be placed in a vibratory environment is not affected by the base excitation. Ensuring good ride comfort to passengers inside a vehicle subjected to undulations from rough roads is an example of this category.

2.2.8.1 *Force transmissibility*   Consider a compressor or an engine mounted on a bed experiencing a vibratory force which in the simplest form may be assumed to be harmonic of the form $F \cos \omega t$ (Fig. 2.15). It is to be ensured that these vibrations are not transmitted through the bed to other nearby machines. Let the mass of the bed be m and let the spring stiffness and damping coefficient of the elastic mount on which it is supported be k and c respectively. Let the displacement response of the bed be $X \cos(\omega t - \phi)$. Then we know that the response X is of the form given in Eq. (2.32).
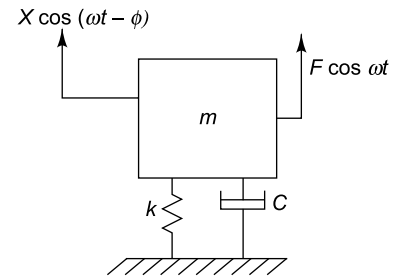
FIGURE 2.15   Force transmissibility

Dividing numerator and denominator of the response by k to non-dimensionalize the equation, we get

$$X = \frac{(F/k)}{\sqrt{(1-\eta^2)^2 + (2\zeta\eta)^2}} \tag{2.38}$$

Force transmitted to the ground by the mount is $kx + c\dot{x}$. The magnitude of this force is given by

$$F_T = \frac{\dfrac{F}{k}\sqrt{k^2 + (c\omega)^2}}{\sqrt{(1-\eta^2)^2 + (2\zeta\eta)^2}} \tag{2.39}$$

On non-dimensionalizing this equation and expressing it as the ratio of force transmitted to the force produced, we get

$$\frac{F_T}{F} = \frac{\sqrt{1 + (2\zeta\eta)^2}}{\sqrt{(1-\eta^2)^2 + (2\zeta\eta)^2}} \tag{2.40}$$

This quantity is known as force transmissibility and its variation with respect to η is shown in Fig. 2.16. Transmissibility is found to be maximum when $\eta = 1$, irrespective of the value of damping ratio, and shoots to infinity when there is no damping. Good vibration isolation, or in other words, reduced transmissibility, i.e. transmissibility less than 0.1, is obtained only for large values of η, typically $\eta > 3$. Considering that one often does not have control over the excitation frequencies, reduction in transmissibility can be achieved by making the natural frequency of the system a fraction of the forcing frequency, at least one-third. Since the forcing frequencies of most machines are known, one can calculate the required low natural frequencies of the combined machine-bed-isolator systems. ω_n can be made small by either increasing the mass of the machine bed or by choosing an isolator of low stiffness. Another very important observation to be made is that in the frequency range of effective vibration isolation, an undamped spring is better than a damped spring as seen from Fig. 2.16, though it is

required to have high damping in the vicinity of resonance for reduced response. This may be achieved through the use of vibration stops.
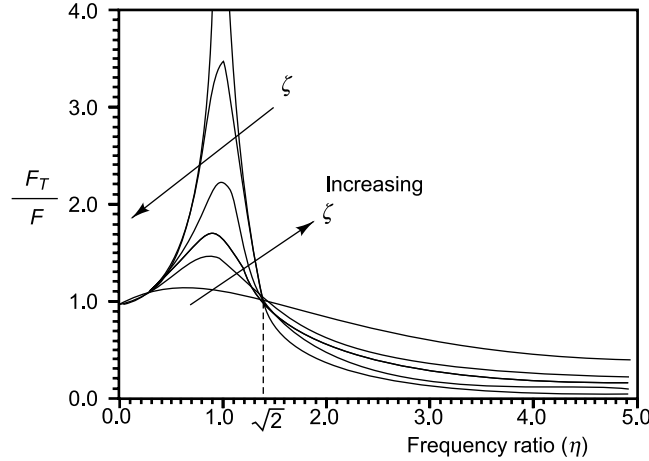
FIGURE 2.16 Transmissibility as a function of η

2.2.8.2 Displacement transmissibility The case of displacement isolation will now be considered. This is shown in Fig. 2.17. Let a system be subjected to a displacement excitation represented by $y = Y \cos \omega t$. Let m be the mass of the system and let the stiffness and damping of the isolators be k and c respectively. Let the response of the system at any instant of time be x. Let the relative displacement be $z = (x - y)$. Referring to Fig. 2.17, the equation of motion of the system is given by

$$m\ddot{x} + c(\dot{x} - \dot{y}) + k(x - y) = 0 \qquad (2.41)$$

Since $\qquad z = x - y, \ \dot{z} = \dot{x} - \dot{y}, \ \ddot{z} = \ddot{x} - \ddot{y} \qquad (2.42)$

Substituting Eq. (2.42) in (2.41) we have

FIGURE 2.17 Displacement transmissibility

$$m\ddot{z} + c\dot{z} + kz = m\omega^2 Y \cos \omega t \qquad (2.43)$$

This is the equation for a forced vibration problem. Hence the solution for z is given by $z = Z \cos (\omega t - \phi)$ and

$$\frac{Z}{\left(m\omega^2 Y / k \right)} = \frac{1}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \qquad (2.44)$$

Rewriting, $\qquad \dfrac{Z}{Y} = \dfrac{\eta^2}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \qquad (2.45)$

The plot for F_T/F in Fig. 2.16 as a function of η, for various values of ζ holds good for Z/Y also. All the salient points put forth for force transmissibility are valid here also. A typical example of displacement isolation is the case of ride comfort in automobiles which have road excitation frequencies anywhere up

to 50 Hz depending on the vehicle speed. Their lowest natural frequency is designed to be around 1 Hz to ensure good ride comfort.

Wherever the exciting force is purely in the vertical direction, one can conveniently use non-metallic isolators like cork in compression, besides helical springs with vertical axis. If the disturbance is emanating from a rotating unbalance with axis horizontal, the exciting force has components of equal intensity in the vertical direction and in a horizontal direction, both at right angles to the axis of rotation. For horizontal exciting forces, rubber in shear in the horizontal plane or inclined rubber isolators can be used. Rubber bushings are well suited for torsional and axial excitation.

2.3   TWO-DEGREE-OF-FREEDOM SYSTEM

A 2-degree-of-freedom system requires 2 independent coordinates to describe its motion and is the simplest of *N*-degree-of-freedom systems. These could be translational or rotational.

2.3.1   Translatory System

Consider a two-degree-of-freedom undamped vibratory system consisting of two springs and two masses as shown in Fig. 2.18. The masses could very well be the sprung and unsprung masses of a vehicle and the stiffness those of the suspensions and tyres. Let x_1 and x_2 denote the vertical displacements measured from the static equilibrium positions of the two masses m_1 and m_2, respectively.

If the system is given some arbitrary initial conditions, say, an initial displacement to mass m_2, the system will perform free vibrations which are periodic in nature. A harmonic analysis of the motion of masses m_1 and m_2 shows that it is made up of vibrations of two frequency components. Using Newton's laws, the differential equations of motion may be written as:

FIGURE 2.18   2-DOF model

$$m_1\ddot{x}_1 + \left(k_1 + k_2\right)x_1 - k_2 x_2 = 0$$

$$m_2\ddot{x}_2 + k_2 x_2 - k_2 x_1 = 0 \tag{2.46}$$

These are coupled equations in two unknowns x_1 and x_2. Equation (2.46) can be put in matrix form as

$$\left[M\right]\{\ddot{x}\} + \left[K\right]\{x\} = \{0\} \tag{2.47}$$

where
$$\left[M\right] = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix} \qquad \left[K\right] = \begin{bmatrix} \left(k_1 + k_2\right) & -k_2 \\ -k_2 & k_2 \end{bmatrix} \tag{2.48}$$

The solution may be assumed to be harmonic of the form

$$\begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix} \exp\left(i\omega t\right) \tag{2.49}$$
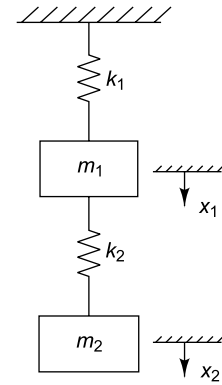
Substituting this solution in Eq. (2.47) we get

$$
\begin{bmatrix} \left(k_1 + k_2 - m_1\omega^2\right) & -k_2 \\ -k_2 & \left(k_2 - m_2\omega^2\right) \end{bmatrix} \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}
\tag{2.50a}
$$

i.e.
$$
\begin{vmatrix} \left(k_1 + k_2 - m_1\omega^2\right) & -k_2 \\ -k_2 & \left(k_2 - m_2\omega^2\right) \end{vmatrix} = 0
\tag{2.50b}
$$

The solution to Eq. (2.50a) assumes harmonic motion for both the masses m_1 and m_2 at frequency ω, but with different amplitudes, X_1 and X_2. For a normal mode of oscillation, each mass undergoes harmonic motion of the same frequency, passing through the equilibrium position simultaneously. For a non-trivial solution for Eq. (2.50a), Eq. (2.50b) has to be satisfied.

i.e.
$$
m_1 m_2 \omega^4 - \omega^2 \left[\left(k_1 + k_2\right) m_2 + k_2 m_1 \right] + k_1 k_2 = 0
\tag{2.51}
$$

This results in a quadratic equation in ω^2, the roots of which are

$$
\omega_{1,2}^2 = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}
\tag{2.52}
$$

where

$$
A = m_1 m_2, \, B = -\left[\left(k_1 + k_2\right) m_2 + k_2 m_1 \right]
$$
$$
C = k_1 k_2
\tag{2.53}
$$

Strictly there are four roots for ω, i.e. $\pm\omega_1$ and $\pm\omega_2$ and the system will perform harmonic motion when the frequency ω in Eq. (2.51) has one of the values given by Eq. (2.52). Hence the complete solution has four vector constants and is as shown in Eq. (2.54).

$$
\begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \left\{\phi_1\right\} e^{+i\omega_1 t} + \left\{\phi_2\right\} e^{-i\omega_1 t} + \left\{\phi_3\right\} e^{+i\omega_2 t} + \left\{\phi_4\right\} e^{-i\omega_2 t}
\tag{2.54}
$$

When $\omega = \omega_1$, the first of Eqs. (2.50a) becomes

$$
(k_1 + k_2 - m_1\omega_1^2)X_1 - k_2 X_2 = 0
\tag{2.55}
$$

Therefore, when $\omega = \omega_1$, X_1 and X_2 have the following ratio

$$
\left.\frac{X_2}{X_1}\right|_{\omega=\omega_1} = \frac{k_1 + k_2 - m_1\omega_1^2}{k_2}
\tag{2.56}
$$

If X_1 is assigned an arbitrary value of unity,

$$
X_2 = \frac{(k_1 + k_2 - m_1\omega_1^2)}{k_2}
\tag{2.57}
$$

The vector $\begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(1)}$ for the first mode can be written as

$$\left\{ \begin{matrix} X_1 \\ X_2 \end{matrix} \right\}^{(1)} = \left\{ \begin{matrix} 1 \\ (k_1 + k_2 - m_1\omega_1^2)/k_2 \end{matrix} \right\} \tag{2.58}$$

Similarly when $\omega = \omega_2$, we can get a similar equation relating X_1 and X_2 from the first of Eqs. (2.50a) as

$$\frac{X_2}{X_1}\bigg|_{\omega=\omega_2} = \left\{ \frac{(k_1 + k_2 - m_1\omega_2^2)}{k_2} \right\} \tag{2.59}$$

Arbitrarily assigning a value of unity to X_1 as before, we get the vector of displacements $\left\{ \begin{matrix} X_1 \\ X_2 \end{matrix} \right\}^{(2)}$ for the second mode as

$$\left\{ \begin{matrix} X_1 \\ X_2 \end{matrix} \right\}^{(2)} = \left\{ \begin{matrix} 1 \\ (k_1 + k_2 - m_1\omega_2^2)/k_2 \end{matrix} \right\} \tag{2.60}$$

The two vectors described in Eqs (2.58) and (2.60) give the displacement configurations of the system when it is performing purely harmonic motion and are called the normal modes of the system. The solution in Eq. (2.54) may now be expressed as

$$\left\{ \begin{matrix} x_1(t) \\ x_2(t) \end{matrix} \right\} = \left\{ \begin{matrix} X_1 \\ X_2 \end{matrix} \right\}^{(1)} (C_1 e^{i\omega_1 t} + C_2 e^{-i\omega_1 t}) + \left\{ \begin{matrix} X_1 \\ X_2 \end{matrix} \right\}^{(2)} (C_3 e^{i\omega_2 t} + C_4 e^{-i\omega_2 t}) \tag{2.61}$$

where the constants C_i, $i = 1, 2, 3, 4$ are obtained by substituting the initial conditions of the motion. Equation (2.61) can be expressed in the alternate form

$$\left\{ \begin{matrix} x_1(t) \\ x_2(t) \end{matrix} \right\} = \left\{ \begin{matrix} X_1 \\ X_2 \end{matrix} \right\}^{(1)} (D_1 \cos\omega_1 t + D_2 \sin\omega_1 t) + \left\{ \begin{matrix} X_1 \\ X_2 \end{matrix} \right\}^{(2)} (D_3 \cos\omega_2 t + D_4 \sin\omega_2 t) \tag{2.62}$$

where

$$D_1 = (C_1 + C_2), \quad D_2 = i(C_1 - C_2)$$
$$D_3 = (C_3 + C_4), \quad D_4 = i(C_3 - C_4) \tag{2.63}$$

Assuming the initial displacements to be $x_1(0)$ and $x_2(0)$ for masses m_1 and m_2 and the initial velocities to be zero, we get

$$\left\{ \begin{matrix} x_1(0) \\ x_2(0) \end{matrix} \right\} = \left\{ \begin{matrix} X_1 \\ X_2 \end{matrix} \right\}^{(1)} D_1 + \left\{ \begin{matrix} X_1 \\ X_2 \end{matrix} \right\}^{(2)} D_3$$

$$\left\{ \begin{matrix} 0 \\ 0 \end{matrix} \right\} = \omega_1 \left\{ \begin{matrix} X_1 \\ X_2 \end{matrix} \right\}^{(1)} D_2 + \omega_2 \left\{ \begin{matrix} X_1 \\ X_2 \end{matrix} \right\}^{(2)} D_4 \tag{2.64}$$

From Eq. (2.64) we get

$$D_1 = \frac{x_1(0)\,X_2^{(2)} - x_2(0)\,X_1^{(2)}}{X_1^{(1)}X_2^{(2)} - X_2^{(1)}X_1^{(2)}}$$

$$D_3 = \frac{X_1^{(1)}x_2(0) - X_2^{(1)}x_1(0)}{X_1^{(1)}X_2^{(2)} - X_2^{(1)}X_1^{(2)}} \tag{2.65}$$

and $D_2 = D_4 = 0$. Hence the solution in Eq. (2.62) can be written as

$$\begin{Bmatrix} x_1(t) \\ x_2(t) \end{Bmatrix} = \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(1)} D_1 \cos\omega_1 t + \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(2)} D_2 \cos\omega_2 t \tag{2.66}$$

It can be seen from Eq. (2.65) that if

$$\begin{Bmatrix} x_1(0) \\ x_2(0) \end{Bmatrix} = \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(1)} \tag{2.67}$$

then $D_1 = 1$, $D_3 = 0$ and the complete motion is in the first mode and is harmonic with frequency ω_1. If on the other hand

$$\begin{Bmatrix} x_1(0) \\ x_2(0) \end{Bmatrix} = \begin{Bmatrix} X_1 \\ X_2 \end{Bmatrix}^{(2)} \tag{2.68}$$

then $D_1 = 0$, $D_3 = 1$, and the motion is in the second mode and is harmonic with frequency ω_2. For arbitrary initial conditions, it is evident from Eq. (2.66) that the motion will be the superposition of two harmonic motions at frequencies ω_1 and ω_2 with both the modes participating.

2.3.2 Coupled Translation and Rotation

In this section, equations corresponding to a system with coupled translation and rotation (as in pitch and bounce modes) are considered. The differential equations of motion for the 2-degree-of-freedom system in general are coupled, in that both coordinates appear in each equation. The equations can be expressed in matrix form and the type of coupling present is evident from them. Mass or dynamical coupling exists if the mass matrix is not diagonal, whereas stiffness or elastic coupling exists if the stiffness matrix is non-diagonal.

Consider the undamped 2-degree-of-freedom system as shown in Fig. 2.19(a) with coordinates x and θ, where x is the linear displacement of the centre of mass G, J is its mass moment of inertia and θ is the rotation of the bar. Let the coordinates be defined through the centre of mass which is not coincident with the geometric centre, i.e. $l_1 \neq l_2$ as shown in Fig. 2.19(b).

With the translational and rotational coordinates described through the centre of gravity (Fig. 2.19(a) and (b)), the equations have coordinate coupling and are of the form

$$\begin{bmatrix} m & 0 \\ 0 & J \end{bmatrix} \begin{Bmatrix} \ddot{x} \\ \ddot{\theta} \end{Bmatrix} + \begin{bmatrix} (k_1 + k_2) & (k_2 l_2 - k_1 l_1) \\ (k_2 l_2 - k_1 l_1) & (k_1 l_1^2 + k_2 l_2^2) \end{bmatrix} \begin{Bmatrix} x \\ \theta \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \tag{2.69}$$
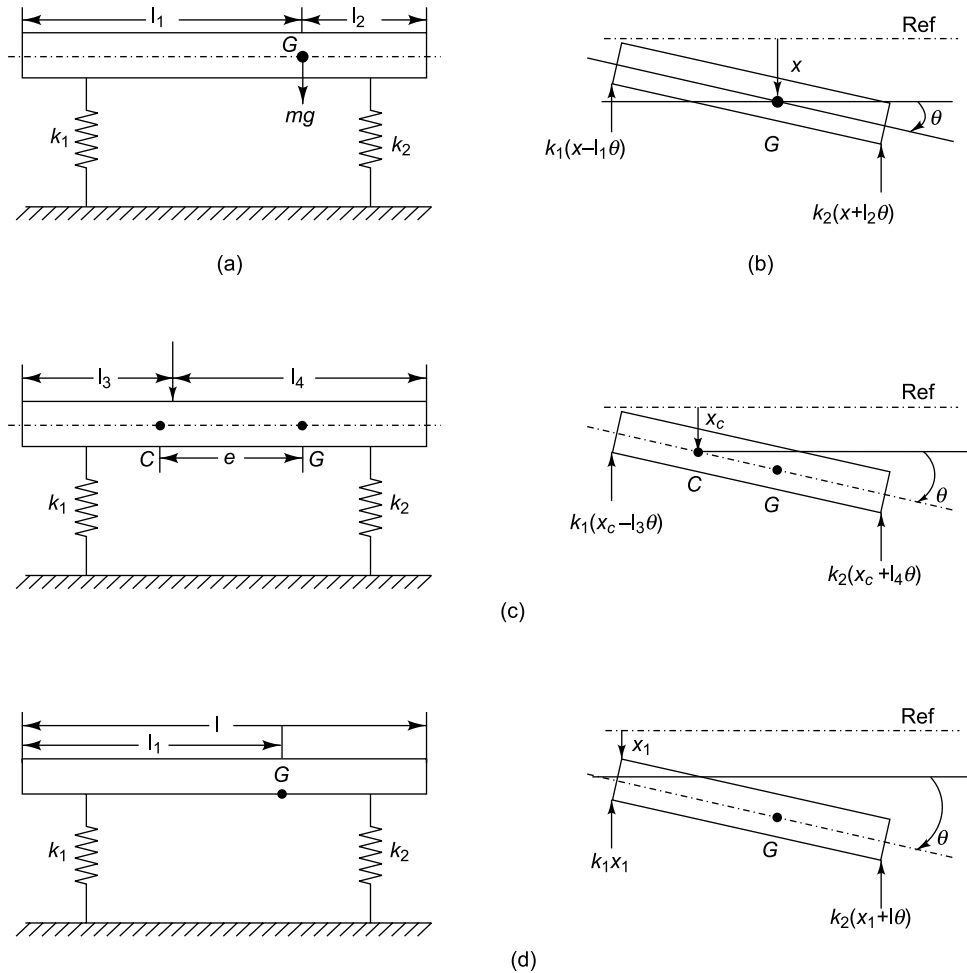
FIGURE 2.19   Two-DOF system with coordinate coupling: (a) two DOF system, (b) static coupling, (c) dynamic coupling, (d) static and dynamic coupling

Stiffness or static coupling exists in this case and the stiffness matrix is non-diagonal. Assuming harmonic motion, the equations of motion of the system can be written as

$$\begin{bmatrix} (k_1 + k_2 - \omega^2 m) & -(k_1 l_1 - k_2 l_2) \\ -(k_1 l_1 - k_2 l_2) & (k_1 l_1^2 + k_2 l_2^2 - \omega^2 J_c) \end{bmatrix} \begin{Bmatrix} x \\ \theta \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \tag{2.70}$$

Mass or dynamic coupling exists if the mass matrix is non-diagonal. Such coupling is obtained by defining the coordinates through a point *C* shown in Fig. 2.19(c), such that a force applied perpendicular to the bar produces pure translation, i.e. $k_2 l_4 = k_1 l_3$. Here static coupling is eliminated and the equations of motion in terms of coordinates x_c and θ can be shown to be

$$\begin{bmatrix} m & me \\ me & J_c \end{bmatrix} \begin{Bmatrix} \ddot{x}_c \\ \ddot{\theta} \end{Bmatrix} + \begin{bmatrix} (k_1 + k_2) & 0 \\ 0 & (k_1 l_3^2 + k_2 l_4^2) \end{bmatrix} \begin{Bmatrix} x_c \\ \theta \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \tag{2.71}$$

It is also possible to have both forms of coupling. If coordinate x is chosen at the end of the bar such that $x = x_1$ and θ is the rotation of the bar as before and as shown in Fig. 2.19(d), the matrix equation is

$$\begin{bmatrix} m & ml_1 \\ ml_1 & J_1 \end{bmatrix} \begin{Bmatrix} \ddot{x}_1 \\ \ddot{\theta} \end{Bmatrix} + \begin{bmatrix} (k_1 + k_2) & k_2 l \\ k_2 l & k_2 l^2 \end{bmatrix} \begin{Bmatrix} x_1 \\ \theta \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix} \tag{2.72}$$

It is possible to find a coordinate system that has neither form of coupling. The two equations are then decoupled and each equation can be solved independent of each other. Coordinates that result in such decoupling are called principal coordinates or normal coordinates. Though it is possible to decouple the equations of motion for an undamped system, it is not always the case for a damped system. The system may have zero dynamic and static coupling, but the coordinates are still coupled through the damping matrix. If the damping is proportional to the stiffness or the mass matrix, the equations become decoupled.

2.4 MULTI-DEGREE-OF-FREEDOM SYSTEM

A system requiring more than one coordinate to describe its motion is called a multi-degree-of-freedom system. Most practical engineering structures are continuous, have infinite degrees of freedom and give rise to inhomogeneous differential equations. However, often we approximate them to systems with only a finite number of degrees of freedom. The size of the model chosen depends on the frequency range of interest in vibration analysis (dictated by the frequency range of excitation) and the vibration modes that are of significance in this frequency range. The effectiveness of the analysis depends on the assumptions made in the model developed and on the number and type of degrees of freedom chosen. An NDOF system has N natural frequencies and N normal modes or in other words N eigenvalues and N eigenvectors which are obtained from the solution of N simultaneous equations of motion of the system. Figure 2.20 represents an NDOF viscously damped system consisting of mass, stiffness and damping elements. Such a model is typically obtained by lumping the properties of a continuous structure or from a finite element (FE) discretization of the structure. All concepts as applicable to SDOF systems are applicable to MDOF systems also. For systems with a large number of degrees of freedom, matrix methods are essential.
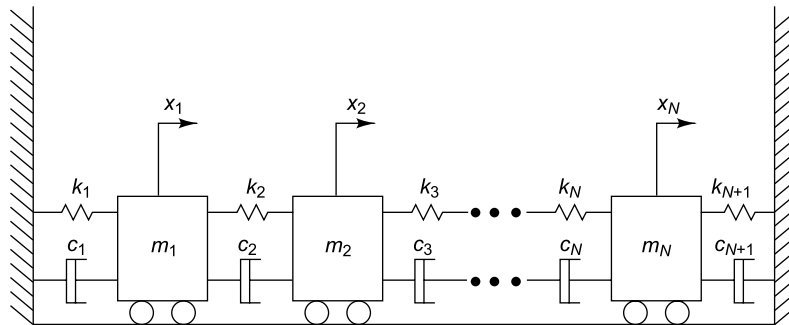
FIGURE 2.20 NDOF vibratory model

2.4.1 Free Vibration of Undamped Multi-Degree-of-Freedom System

Let us first consider the undamped case. Normal mode vibrations depend only on the masses and stiffness of the system and their distribution. When a system vibrates at any of its normal modes, all points in the system undergo simple harmonic motion that passes through their equilibrium positions simultaneously. For this, a system must be given specific initial conditions corresponding to its normal mode to excite it. With general initial conditions such as an impulse, the vibrating system may vibrate at all normal modes simultaneously. The N equations of motion for the undamped system may be represented as

$$m_1 \ddot{x}_1 + \left(k_1 + k_2 \right) x_1 - k_2 x_2 = 0$$
$$m_2 \ddot{x}_2 - k_2 x_1 + \left(k_2 + k_3 \right) x_2 - k_3 x_3 = 0$$

$$\begin{matrix} . & . & . & . \\ . & . & . & . \\ . & . & . & . \end{matrix}$$ (2.73)

$$m_N \ddot{x}_N - k_N x_{N-1} + \left(k_N + k_{N+1} \right) x_N = 0$$

They can be rewritten in matrix form as

$$\left[M \right]\{\ddot{x}\} + \left[K \right]\{x\} = 0$$ (2.74)

Here $[M]$ is the $N \times N$ mass matrix and $[K]$ is the $N \times N$ stiffness matrix. $\{x\}$ is the $N \times 1$ coordinate displacement vector.

The solution is assumed to be harmonic as

$$\{x\} = \{X\} \exp(i\omega t)$$ (2.75)

The solution satisfies the differential equation if

$$[K - \omega^2 M] \{x\} = \{0\}$$ (2.76)

Equation (2.76) is a system of N homogeneous simultaneous equations and for the existence of a non-trivial solution, we must have

$$|K - \omega^2 M| = 0$$ (2.77)

This is a polynomial in ω^2 of order N and there will be exactly N roots of ω^2, for which Eq. (2.77) is satisfied. If the matrices $[K]$ and $[M]$ are positive definite (or semi-definite) which is satisfied if the matrices are symmetric, then real solutions exist for ω^2 given by $\omega_1^2, \omega_2^2, \ldots\ldots\ldots\ldots\omega_N^2$, which are the squares of the natural frequencies of the system. These solutions are called the eigenvalues. Substituting any one of these natural frequencies into Eq. (2.76), we can obtain the system configuration at that natural frequency which is a vector solution given by $\{\phi_k, k = 1, 2, \ldots\ldots N\}$ and which represents the mode shapes or eigenvectors of the system. There are N such natural mode shapes for the structure corresponding to the N natural frequencies. It is clear from the type of solution assumed in Eq. (2.75) that the structure can vibrate harmonically at each of these N natural frequencies, with the structural configuration at those frequencies defined by the corresponding mode shapes.

2.4.2   Free Vibration of Multi-Degree-of-Freedom System with Damping

Equation (2.78) represents a set of N coupled second-order differential equations for an MDOF system with viscous damping.

$$m_1 \ddot{x}_1 + (c_1 + c_2)\dot{x}_1 - c_2 \dot{x}_2 + (k_1 + k_2) x_1 - k_2 x_2 = 0$$

$$m_2 \ddot{x}_2 - c_2 \dot{x}_1 + (c_2 + c_3)\dot{x}_2 - c_3 \dot{x}_3 - k_2 x_1 + (k_2 + k_3) x_2 - k_3 x_3 = 0$$

$$\begin{matrix} \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot & \cdot & & \cdot \end{matrix}$$

$$m_N \ddot{x}_N - c_N \dot{x}_{N-1} + (c_N + c_{N+1})\dot{x}_N - k_N x_{N-1} + (k_N + k_{N+1}) x_N = 0$$

(2.78)

It can be rewritten in matrix form as follows:

$$[M]\{\ddot{x}\} + [C]\{\dot{x}\} + [K]\{x\} = \{0\} \tag{2.79}$$

Here the matrices $[M]$, $[K]$ and $[C]$ are the mass, stiffness and damping matrices, respectively, describing the spatial properties of the discretized structure. Hence this model is called the spatial model. $\{\ddot{x}\}$, $\{\dot{x}\}$ and $\{x\}$ correspond to the acceleration, velocity and displacement vectors. Equation (2.79) can be solved if two initial conditions of motion are specified for each of the N coordinates.

2.4.3   Orthogonality of Natural Modes

The mode shape vectors satisfying the eigenvalue problem possess 'orthogonality' properties which are due to the symmetry of the $[K]$ and $[M]$ matrices. The ith natural mode corresponding to the ith natural frequency is a column vector with N elements, each element representing the displacement of the corresponding DOF in that mode. Each of the mode shapes satisfies the equations of motion individually. Hence for the ith and jth modes, we have

$$-\omega_i^2 [M]\{\phi_i\} + [K]\{\phi_i\} = \{0\}$$

$$-\omega_j^2 [M]\{\phi_j\} + [K]\{\phi_j\} = \{0\}$$

(2.80)

Pre-multiplying the first equation by ϕ_j^T and the second by ϕ_i^T and writing the resulting expression without matrix brackets, we have

$$\omega_i^2 \phi_j^T M \phi_i = \phi_j^T K \phi_i$$

$$\omega_j^2 \phi_i^T M \phi_j = \phi_i^T K \phi_j$$

(2.81)

Taking transpose of both sides of the second of Eqs (2.81) above, we obtain

$$\omega_j^2 \phi_j^T M^T \phi_i = \phi_j^T K^T \phi_i \tag{2.82}$$

For most practical structures, the mass and stiffness matrices are symmetric and therefore $M^T = M$, $K^T = K$. Using this in Eq. (2.82) we obtain

$$\omega_j^2 \phi_j^T M \phi_i = \phi_j^T K \phi_i \qquad (2.83)$$

Subtracting Eq. (2.83) from the first equation of (2.81), we get

$$\left(\omega_i^2 - \omega_j^2\right) \phi_j^T M \phi_i = 0 \qquad (2.84)$$

However, $\omega_i \neq \omega_j$; therefore

$$\phi_j^T M \phi_i = 0 \quad \text{for} \quad i \neq j \qquad (2.85a)$$

Substituting Eq. (2.85a) in the first of Eqs (2.81), it follows that

$$\phi_j^T K \phi_i = 0 \quad \text{for} \quad i \neq j \qquad (2.85b)$$

However, if $i = j$, then

$$\phi_i^T M \phi_i = \mu_i$$
$$\phi_i^T K \phi_i = \kappa_i \qquad (2.86a)$$

$$\frac{\phi_i^T K \phi_i}{\phi_i^T M \phi_i} = \frac{\mu_i}{\kappa_i} = \omega_i^2 \qquad (2.86b)$$

Equations (2.85) and (2.86) define the orthogonality conditions for normal modes. μ_i and κ_i are called the generalized or modal mass and stiffness corresponding to the ith mode.

2.4.4   Free Vibration of Multi-Degree-of-Freedom System in Terms of Modal Coordinates

The free vibration response of an MDOF system can be expressed in terms of the normal modes as mentioned earlier. It is to be noted that the absolute magnitudes of the eigenvectors of the ith mode ϕ_i, $i = 1, 2,..., N$ are not known, but their directions are. Since the absolute magnitudes are not known and depend on the initial conditions, one can normalize the mode shape vectors with respect to the modal masses and these 'normal mode shapes' are then called the normalized or weighted eigenvectors, ψ_i, $i = 1, 2,..., N$.

$$\psi_i = \frac{\phi_i}{\sqrt{\mu_i}}, \, i = 1, 2, ..., N \qquad (2.87)$$

A modal coordinate is associated with every mode. The physical coordinate $\{x\}$ is related to the modal coordinate $\{q\}$ by the equation

$$\{x(t)\} = [\phi]\{q(t)\} \qquad (2.88)$$

Here $[\phi]$ is the $N \times N$ matrix called the modal matrix and is obtained by arranging all the eigenvectors side by side. The vector corresponding to the ith column of this matrix is the modal vector corresponding to the ith mode. Substituting for $\{x\}$ from Eq. (2.88) into Eq. (2.74) and omitting the brackets, we get

$$M\phi\ddot{q} + K\phi q = 0 \tag{2.89}$$

Pre-multiplying Eq. (2.89) by ϕ^T we get

$$\phi^T M\phi\ddot{q} + \phi^T K\phi q = 0 \tag{2.90}$$

Using the orthogonality conditions for normal modes, Eq. (2.90) becomes

$$\mu\ddot{q} + \kappa q = 0 \tag{2.91}$$

Here μ and κ are diagonal matrices containing generalized masses and stiffness. Equation (2.91) constitutes a set of N uncoupled equations of the form

$$\mu_i\ddot{q}_i + \kappa_i q_i = 0, \qquad i = 1, 2, ..., N \tag{2.92}$$

Equation (2.91) simply corresponds to N SDOF systems as μ and κ are diagonal matrices. Therefore, one can solve for $\{q(t)\}$ using SDOF theory and then use the transformation relation in Eq. (2.88) to calculate the physical motion vector $\{x(t)\}$. The response coordinates $\{q(t)\}$ are called the principal coordinates or normal coordinates.

Instead of the transformation used in Eq. (2.88), if we use one involving the normalized eigenvectors of the form

$$\{x(t)\} = [\psi]\{q(t)\} \tag{2.93}$$

then Eq. (2.92) takes on the form

$$\ddot{q}_i + \omega_i^2 q_i = 0, \qquad i = 1, 2, ..., N \tag{2.94}$$

Here again the equation corresponds to N SDOF systems. If we include the effect of the damping matrix however, as in Eq. (2.79), we will find that Eqs (2.89) and (2.92) are not diagonal. However, a form of viscous damping, known as proportional or Rayleigh damping results in diagonalization of the damping matrix also. Rayleigh damping assumes that the damping matrix, $[C]$ is given by $\alpha[M] + \beta[K]$, where α and β are proportionality constants. Then the normal mode matrix $[\psi]$, which diagonalized the equations of motions along with the earlier transformation represented by Eq. (2.93), can now diagonalize the damping matrix $[C]$ also, to yield N uncoupled SDOF damped equations as shown below

$$\ddot{q}_i + 2\zeta_i\omega_i\dot{q}_i + \omega_i^2 q_i = 0, \qquad i = 1, 2, ..., N \tag{2.95}$$

Here $2\zeta_i\omega_i$ is called the modal damping and is defined by the equation

$$2\zeta_i\omega_i = \alpha + \beta\omega_i^2 \tag{2.96}$$

It is worthwhile knowing that the values of α and β may be obtained, if experimental values of ζ_i and ω_i are got for any two modes of the system through experiments.

The solution corresponding to each of Eqs (2.91) can be obtained by plugging in the initial conditions $\{x(0)\}$ and $\{\dot{x}(0)\}$ in the coordinates $\{q\}$, which are obtained from the transformation Eq. (2.88).

$$\{q(0)\} = [\phi]^{-1}\{x(0)\}$$

$$\{\dot{q}(0)\} = [\phi]^{-1}\{\dot{x}(0)\} \tag{2.97}$$

The solution to Eq. (2.91) is then

$$q_i(t) = q_i(0)\cos\omega_i t + \frac{\dot{q}_i(0)}{\omega_i}\sin\omega_i t \qquad (2.98)$$

Here ω_i is the ith natural frequency which is related to the modal mass and stiffness matrices as described in Eq. (2.86). After solving for q_i, $i = 1, 2, 3, \ldots, N$, the response in coordinates $\{x(t)\}$ is obtained using Eq. (2.88).

2.4.5   Forced Response to Harmonic Excitation: Frequency Response Function

Once the free vibration response is known, the next step is determination of the steady state response of the structure to sinusoidal excitation at different frequencies. This is called its frequency response function (FRF). Knowledge of the mass, stiffness and damping properties of the structure can be used to obtain the FRF or transfer function. The response of the MDOF system is derived for the proportional viscous damping case alone. The governing equations are obtained based on the assumption that each mass is acted upon by a force $f_i(t)$, $i = 1, 2,\ldots, N$. The equations of motion with forced excitation can be written as

$$[M]\{\ddot{x}\} + [C]\{\dot{x}\} + [K]\{x\} = \{f(t)\} \qquad (2.99)$$

The eigenvalues and normalized eigenvectors obtained by solving the homogeneous eigenvalue Eq. (2.74), ignoring the damping terms are ω_i and ψ_i, $i = 1, 2, 3,\ldots, N$. Using the transformation in Eq. (2.93) and substituting in Eq. (2.99) we get

$$[M][\psi]\{\ddot{q}\} + [C][\psi]\{\dot{q}\} + [K][\psi]\{q\} = \{f(t)\} \qquad (2.100)$$

Pre-multiplying both sides of Eq. (2.100) by ψ^T and using the principle of orthogonality of the eigenvectors, we get the equation corresponding to the jth coordinate, q_j as

$$\{\ddot{q}_j\} + [2\zeta_j\omega_j]\{\dot{q}_j\} + [\omega_j^2]\{q_j\} = \{\sigma_j(t)\} \qquad (2.101)$$

where $\{\sigma\} = [\psi]^T\{f(t)\}$ is the generalized force vector.

Equation (2.101) constitutes N number of uncoupled equations of motion corresponding to N SDOF systems. The complete solution in the jth coordinate, q_j is

$$q_j = e^{-\zeta_j\omega_j t}\left(A_j\cos\omega_{d_j}t + B_j\sin\omega_{d_j}t\right) + q_{js}(t) \qquad (2.102)$$

where

$$\omega_{d_j} = \omega_j\left(1 - \zeta_j^2\right)^{1/2} \qquad (2.103)$$

and $q_{js}(t)$ is the steady state solution in the jth coordinate and can be readily obtained when the form of forcing function $\sigma_j(t)$ is known. If $\sigma_j(t)$ is harmonic and is expressible as

$$\sigma_j(t) = \sigma_{jo}\cos\omega t \qquad (2.104)$$

then it can easily be shown that $q_{js}(t)$ can be expressed as

$$q_{js}(t) = \frac{\sigma_{jo}\cos\omega t}{\omega_j^2 - \omega^2 + i2\zeta_j\omega_j\omega}, \quad j = 1, 2,\ldots, N \qquad (2.105)$$

The amplitude of response in the *j*th coordinate to harmonic excitation of unit amplitude in the *j*th coordinate or in other words the term $H_{jj}(\omega)$ in the frequency response function (FRF) is given by

$$H_{jj}(\omega) = \left[\omega_j^2 - \omega^2 + i2\zeta_j\omega_j\omega \right]^{-1} \tag{2.106a}$$

The total steady state response amplitude in the *j*th coordinate due to the contribution from all the *N* modes is given by

$$Q_{js} = \frac{\sum_{r=1}^{N}(_r\psi_j)\sigma_{ro}}{\omega_j^2 - \omega^2 + i2\zeta_j\omega_j\omega}, \; j = 1, 2, ..., N \tag{2.106b}$$

Here $(_r\psi_j)$ is the response of the *j*th coordinate in the *r*th mode. The total steady state response in the *j*th coordinate to a unit harmonic force in the *k*th coordinate is given by

$$H_{jk}(\omega) = \sum_{r=1}^{N} \frac{(_r\psi_j)(_r\psi_k)}{\left[\omega_j^2 - \omega^2 + i2\zeta_j\omega_j\omega \right]} \tag{2.107}$$

$H(\omega)$ is the called the receptance matrix or frequency response function matrix and has all the dynamic characteristics of the system. Each element in the receptance matrix $H_{jk}(\omega)$ describes the relation between the response at location *j* due to an excitation at location *k*. The form given in Eq. (2.107) is an efficient way to obtain the response. More details regarding obtaining and interpreting the FRF matrix is discussed in Chapter 10.

FURTHER READINGS

1. Anderson, R.A., *Fundamentals of Vibrations*, Macmillan, New York, 1967.
2. Bishop, R.E.D., Gladwell, G.M.L. and Michaelson, S., *The Matrix Analysis of Vibration*, Cambridge University Press, London, 1965.
3. Bottega, W.J., *Engineering Vibrations*, CRC Press Online, 2006.
4. Clough, R.W. and Penzien, J., *Dynamics of Structures*, McGraw-Hill Inc. New York, 1993.
5. Den Hartog. J.P., *Mechanical Vibrations*, Dover Publications Inc., New York, 1985.
6. Dimarogonas, A.D. and Haddad, S., *Vibration for Engineers*, Prentice-Hall, New Jersey, 1992.
7. Dukkipati, R.V., *Vibration Analysis*, Alpha Science International, Ltd., Harrow, United Kingdom, 2004.
8. Ginsberg, J.H., *Mechanical and Structural Vibrations: Theory and Applications*, John Wiley & Sons, Inc., New York. 2001.
9. Graham Kelly, S., *Schaum's Outline of Theory and Problems of Mechanical Vibrations*, McGraw-Hill, New York, 1996.
10. Harker, R.J., *Generalised Methods of Vibration Analysis*, Wiley, New York, 1983.
11. Huston, D.V., *Applied Mechanical Vibrations*, McGraw-Hill, New York, 1981.
12. Inman, D.J., *Engineering Vibration*, Prentice Hall, Upper Saddle River, NJ, 2001.
13. Jacobsen, L.S. and Ayre, R.S., *Engineering Vibrations*, McGraw-Hill, New York, 1958.
14. Lalanne, C., *Mechanical Vibration & Shock*, Taylor & Francis, North America, 2002.

15. Lalanne, M., Berthier, P. and Der Hagopian, J., *Mechanical Vibration for Engineers*, Wiley, New York, 1983.
16. Meirovitch, L., *Analytical Methods in Vibrations*, Macmillan Co., New York, 1967.
17. Meirovitch, L., *Elements of Vibration Analysis*, McGraw-Hill, New York, 1986.
18. Meirovitch, L., *Methods of Analytical Dynamics*, McGraw-Hill, Inc., New York, 1970.
19. Myklestad, N.O., *Fundamentals of Vibration Analysis*, McGraw-Hill, New York, 1956.
20. Ramamurti, V., *Mechanical Vibration Practice with Basic Theory*, Narosa Publishing House, New Delhi, 2000.
21. Rao, S.S., *Mechanical Vibrations*, Addison-Wesley, Reading, MA, 1986.
22. Shabana, A.A., *Theory of Vibration: An Introduction*, Springer, New York, 1996.
23. Shabana, A.A., *Vibration of Discrete and Continuous Systems*, Springer, 1997.
24. Srinivasan, P., *Mechanical Vibration Analysis*, Tata McGraw-Hill Publishing Company Limited, New Delhi, 1990.
25. Svetlitsky. V.A., Chechin, V.A. and Merzon, G.I., *Engineering Vibration Analysis: Worked Problems*, Vol. 1, Springer Verlag, New York, 2004.
26. Svetlitsky, V.A, Lidvansky, A.S. and Mukhamedshin, R.A., *Engineering Vibration Analysis: Worked Problems*, Vol. 2, Springer, New York, 2004.
27. Thomson, W.T. and Dahleh, M.D., *Theory of Vibration with Applications*, Pearson Education India, India, 2003.
28. Thomson, W.T., *Theory of Vibration with Applications*, Taylor & Francis Inc., Bristol, USA, 2004.
29. Tse, F.S., Morse, I.E., and Hinkle, R.T., *Mechanical Vibrations—Theory and Applications*, Allyn and Bacon, Boston, 1978.
30. Weaver, W., Timoshenko S. and Young D.H., *Vibration Problems in Engineering*, Wiley-IEEE, New York, 1990.

Chapter

3

Vibration Transducers

3.1 INTRODUCTION

We are often faced with the task of measuring the vibration level in a machine to decide if it is within allowable limits. For this we need to use a vibration transducer. Considering that the market is flooded with a plethora of transducers, choosing the appropriate transducer for the application at hand becomes difficult, and requires answers to the following questions:

(i) Do I need to measure displacement, velocity or acceleration?
(ii) Can I use contact type of transducers without introducing mass loading?
(iii) Do I need any additional signal conditioning devices to go with the transducers?
(iv) What are the frequency range, dynamic range and sensitivity required?
(v) Is the choice of the transducer dictated by any environmental constraints like dust, electromagnetic fields, humidity, shock levels, etc.?
(vi) What is my budget?

One should arrive at the appropriate transducer only after one has the answers to all these questions. Otherwise the very purpose of measurement may be defeated. If one were to make vibration measurements on a ball mill rotating at 100 revolutions per minute (rpm) using a piezoelectric accelerometer, one would find that all the low frequency vibrations have not been picked up. If one were to use a heavy accelerometer to measure the vibrations of a light panel, one would find that the mass loading effects have altered the system dynamics completely. The transducer being the front end in any measurement system, care has to be taken in selecting an appropriate one. All further signal conditioning and analysis will be a waste if the signals have not been correctly captured in the first place. This necessitates that the user have a thorough understanding of the various types of transducers available in the market, their principles of operation, their advantages and limitations, typical specifications, etc. This chapter aims at providing all this information.

3.2 CLASSIFICATION OF TRANSDUCERS

Vibration transducers can be classified into the following categories based on the following considerations:

(i) Parameter being measured
  • Displacement
  • Velocity
  • Acceleration

(ii) Electrical output
- Passive
- Active

(iii) Type of output
- Relative
- Absolute

(iv) Proximity of the transducer to the vibrating object
- Contact
- Non-contact.

3.2.1 Choice of Vibration Parameter—Displacement, Velocity and Acceleration

In its simplest form, a vibration displacement signal may be represented by the harmonic function

$$x(t) = A \sin \omega t \tag{3.1}$$

By differentiation, we have velocity and acceleration given by

$$\dot{x}(t) = A \cos \omega t \tag{3.2}$$

$$\ddot{x}(t) = -A\omega^2 \sin \omega t \tag{3.3}$$

Equations (3.1)–(3.3) show how displacement, velocity and acceleration components of a complex signal are dependent on the angular frequency ω. Thus the amplitude of velocity is proportional to frequency and that of acceleration to square of frequency. Any vibratory system may be represented by an equation of the form given below, as discussed in Chapter 2:

$$m\ddot{x} + c\dot{x} + kx = F(t) \tag{3.4}$$

Here the terms on the left-hand side are those of inertia, damping and elastic forces in that order. By substituting Eqs (3.1), (3.2) and (3.3) into (3.4), we find that the relative magnitudes of the various forces in Eq. (3.4) are frequency dependent. An important observation to be made is that at high frequencies, the force component due to acceleration may dominate, resulting in large forces, even with very low displacement amplitudes, on account of the ω^2 effect. It is indeed these high frequency components, which remain undetected by displacement criteria and cause sudden catastrophic fatigue failures. One can therefore conclude that displacement alone, as often applied, can be a poor measure of vibration severity, since this should always be considered in conjunction with frequency. To overcome this difficulty, maximum velocity criteria are used so as to include the frequency parameter and the total dynamic force concept in the measurement of machinery vibration.

The frequencies of interest in machinery vibration analysis form a wide continuous spectrum from the infrasound range of a few Hz through the ultrasonic range approaching 100 kHz. Within this range are present subharmonic 'whirl' frequencies of the order of 10s of Hz, once per revolution frequencies from say 20 to 700 Hz, vane or blade or gear tooth passing frequencies from 2 to 30 kHz and finally antifriction bearing frequencies from 20 to 100 kHz. With such a vast range of frequencies, it is imperative that the vibration engineer acquires and presents data in an effective and meaningful manner. Figure 3.1 gives vibration displacement and acceleration as a function of frequency for a constant velocity level of

7.5 mm/s. This velocity corresponds to a vibration displacement of 1.194 mm (1194 μm) at 1 Hz which amounts to a very small acceleration of 47.1 mm/s^2 (0.0048 g). On the other hand, at 1000 Hz, the same vibration velocity of 7.5 mm/s corresponds to a high acceleration of 47,100 mm/s^2 (4.8 g) with a displacement of only 1.194×10^{-3} mm (1.194 μm).
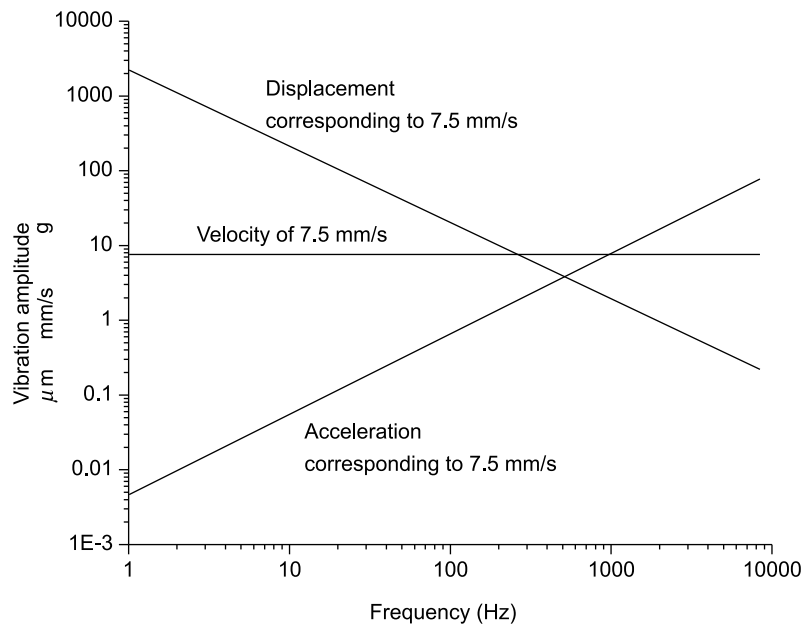
FIGURE 3.1 Displacement, velocity and acceleration nomograph

Thus, it is clear from the figure that transducers are like windows through which portions of the frequency spectrum may be observed. In general, displacements are large at low frequencies and displacement sensors, or vibrometers as they are called, are preferred for these frequencies. Acceleration values are large at high frequencies and for such applications accelerometers are preferred. Velocity measurements are useful at intermediate frequencies where the displacements are likely to be small, or where the frequency range of measurements is not known beforehand. The displacement pickup has a maximum frequency range from 0 to about 1000 Hz, the velocity pickup from 10 to about 2500 Hz and the accelerometer from 20 Hz to well above 50 kHz.

3.2.2 Active and Passive Transducers

Two different principles are involved in the process of converting non-electrical variables into electrical signals. One of these is energy conversion; transducers based on this principle are called active transducers. In theory, active transducers can utilize any known physical principle for converting non-electrical energy, typically mechanical energy, into electrical energy. It is a characteristic of active transducers that, frequently, but not always, the same transduction principle used to convert from a non-electrical form of energy, can also be used in the reverse direction to convert electrical energy into non-electrical forms. Examples are the electrodynamic transducer and microphone, which give a voltage output proportional to vibration or pressure fluctuations. Table 3.1 shows some methods of energy conversion used in active transducers.

TABLE **3.1**   Energy conversion used in active vibration transducers

Original form	Transduced form	Device or effect	Reversible
Mechanical	Electrical	Magnetic induction	Yes
		Electric induction	Yes
Pressure	Electrical	Piezoelectric	Yes
Light radiation	Electrical	Photoelectric	No
Acoustic	Electrical	Microphone	Yes
Electrical	Acoutic	Loudspeaker	Yes

Passive transducers use a different principle to generate electrical signals other than voltage or current. The principle involved here is control of a DC excitation voltage or modulation of an AC carrier signal. The mechanical signal is converted into a parameter such as a change in a circuit resistance or inductance or capacitance. Examples of passive transducers are strain gauges, linear variable differential transformers (LVDT) and capacitance pickups, which give output resistance, inductance and capacitance, respectively, proportional to the input signal.

3.2.3   Absolute and Relative Measuring Instruments

Fixed reference instruments or absolute measuring devices have one terminal of the instrument attached to a point which is fixed in space, whereas the other terminal is attached to the vibrating object. The measurements made using such transducers are independent of the user. Many of them incorporate a so-called seismic mass, which offers a fixed frame of reference with respect to which measurements are made. Relative measuring instruments on the other hand, display different values with different users since measurements are not made with respect to a fixed frame. Quite often the readings depend on the spring force with which the user keeps the transducer in contact with the vibrating object.

3.2.4   Contact and Non-contact Transducers

Contact type of transducers may be used for conducting vibration tests on heavy specimens such as machine foundations, large machines, road vehicles, ships, bridges and buildings, but care must be exercised in using these on lightweight structures since they are likely to cause mass loading effects. Hence non-contact transducers have to be used for the same. Besides, contact pickups, heavy or light, cannot be used on rotating machines and non-intrusive measurement techniques have necessarily to be used.

3.3   EARLIEST VIBRATION TRANSDUCER: THE HAND VIBROGRAPH

A hand vibrograph (Fig. 3.2) is an easy to handle robust device used for the measurement of vibrations. This instrument is one of the earliest vibration transducers and is essentially of historic importance since it is seldom used. It is a relative measuring instrument, which measures displacement and has no fixed reference point (like a seismic mass) as in the case of absolute measuring instruments (e.g. LVDT). Owing to its compactness and small weight, it is versatile in its application and can be used for measurements without any special preparations.

It consists of a housing having two chambers. In one chamber, a rotating recording drum is present. Paper is wound around this rotating drum, so that the pen or stylus will scribe the vibration records on

it. The pen is coupled to the follower pin or feeler through a system of levers. The pin is attached to a spring present in another chamber. It protrudes out of the housing and during operation is pressed against the vibrating object in the direction of motion so that it can freely move up and down between its stops. The vibrations are then magnified and recorded on the rotating drum. When the vibrating object moves with a particular acceleration in the downward direction, the spring force has to overcome the inertia force of the follower pin and system of levers, to keep the pen pressed against the object, so that it will not chatter. A greater spring force admits measurement of higher accelerations.

The vibrograph can be held to the vibrating object in two ways and the frequency response of the recording depends on this, since it is a function of the relative movement between the pin and the housing. When holding the vibrograph with both hands, it acts as a vibrometer. This differs from a standard vibrometer as it forms an inert mass itself and the holding mass together with the spring causes the flexibility effect. It is recommended to use the instrument only above the natural frequency, i.e. typically above 3–5 Hz in which case the recorded amplitude will always be the product of vibration displacement and amplification. If clamped in a vice, the instrument becomes a simple displacement instrument; this requires a little more preparatory work and in this mode it can also be used for low frequencies down to 0 Hz. The limit of acceleration, as in the case of vibrograph application, depends upon the force of tension or rigidity of the clamping. The upper frequency limit is dependent upon the rigidity of the clamping. In the case of ideal rigid clamping, the record will be proportional to the vibration amplitude throughout the usable frequency range.
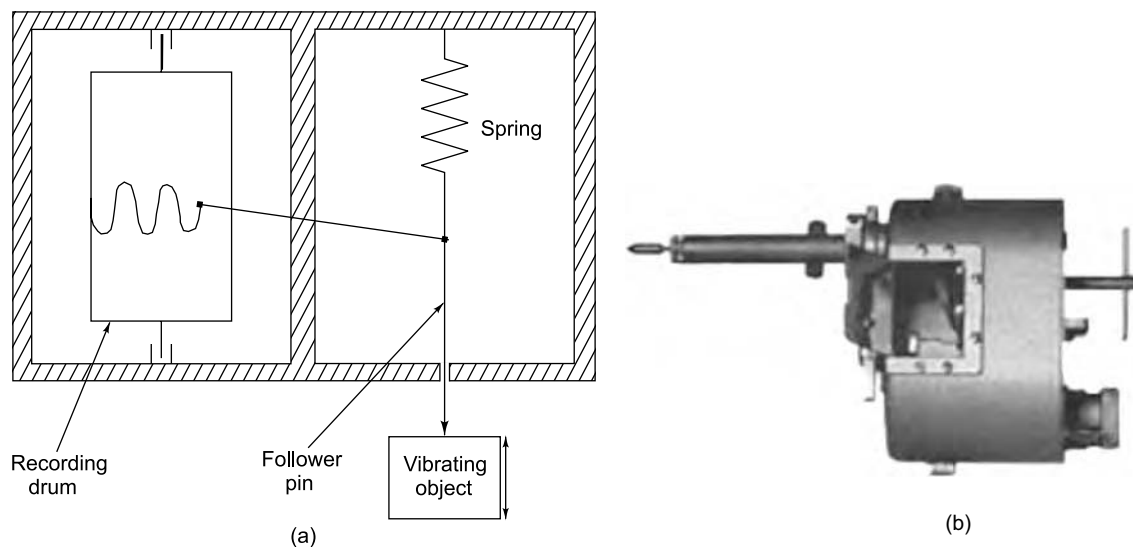
FIGURE 3.2 Hand vibrograph: (a) schematic drawing, (b) photograph
(Courtesy of http://www.star-instruments.co.uk)

Typical specifications

Frequency range:	5–100 Hz
Amplitude range:	25 μm–2 mm
Magnification:	1, 5, 20, 50

Maximum acceleration:	20 g
Weight:	15 N

Advantages

  (i) It is a compact device.

 (ii) It can be made to work in any desired direction and does not require any preparatory measures for attaching it to the vibrating object.

(iii) It is simple to use with built-in battery.

Disadvantage

  Readings taken by different operators can be different due to differing spring forces.

Applications

  (i) It can be used for measurement of critical speeds, unbalances, etc. on steam turbines.

 (ii) Rough preliminary displacement measurements can be made on machines, machine tools like lathe, grinding machine, power plants, automobiles and buildings.

(iii) It provides early warning of breakage or excessive wear.

3.4 ABSOLUTE MEASURING TRANSDUCERS OR SEISMIC TRANSDUCERS

The absolute measuring or seismic transducer can be modelled as shown in Fig. 3.3. It has a seismic mass, i.e. one that remains stationary in space over the usable frequency range of the transducer. Depending on the frequency range utilized, displacement, velocity or acceleration is indicated by the relative motion of the suspended mass with respect to the casing.
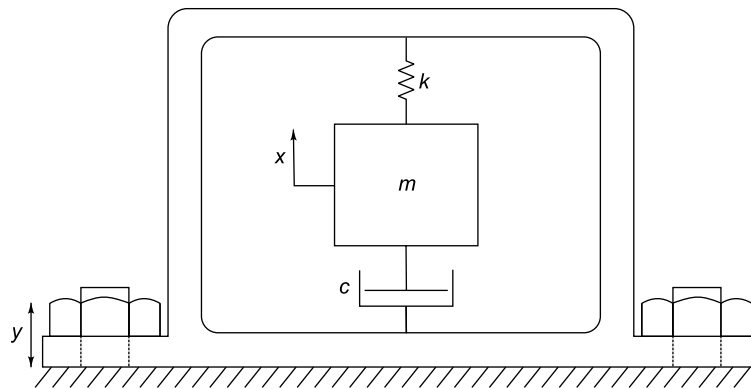
FIGURE 3.3　Vibratory model of seismic transducer

  The equation of motion of the system shown in Fig. 3.3 may be written as

$$m\ddot{x} = -c(\dot{x} - \dot{y}) - k(x - y) \tag{3.5}$$

where *x* and *y* are the displacements of the seismic mass and vibrating body, respectively, both measured with respect to an inertial reference. Letting the relative displacement of the mass *m* with respect to the case attached to the vibrating body be

$$z = (x - y) \tag{3.6}$$

and assuming sinusoidal motion $y = Y \sin \omega t$ of the vibrating body, we obtain the equation

$$m\ddot{z} + c\dot{z} + kz = m\omega^2 Y \sin \omega t \tag{3.7}$$

The steady-state solution is

$$z = Z \sin (\omega t - \phi) \tag{3.8}$$

The frequency response function Z / Y is a complex quantity with magnitude and phase as shown below:

$$Z = \frac{m\omega^2 Y}{\sqrt{(k - m\omega^2)^2 + (c\omega)^2}} \tag{3.9a}$$

$$= \frac{Y(\omega/\omega_n)^2}{\sqrt{\{[1 - (\omega/\omega_n)^2]\}^2 + [2\zeta(\omega/\omega_n)]^2}} \tag{3.9b}$$

or

$$\frac{Z}{Y} = \frac{\eta^2}{\sqrt{(1 - \eta^2)^2 + (2\zeta\eta)^2}} \tag{3.9c}$$

where $\eta = \omega/\omega_n$ = the frequency ratio and ζ is the damping ratio. The phase lag is given by

$$\tan \phi = \frac{\omega c}{k - m\omega^2} = \frac{2\zeta(\omega/\omega_n)}{1 - (\omega/\omega_n)^2} \tag{3.10}$$

Figure 3.4 shows a plot of Eqs (3.9c) and (3.10) as a function of η. It is evident from the figure that when $\eta = \omega/\omega_n \gg 1$, i.e. when the natural frequency ω_n of the instrument is low in comparison to the vibration frequency ω to be measured, the ratio Z/Y approaches unity and the relative displacement Z approaches Y irrespective of the value of damping ratio ζ. This means that the mass remains undisturbed in space. Hence the relative displacement between the casing and the mass is the true displacement of the casing. Likewise, the relative velocity between the casing and the mass is the true velocity of the casing. Since these transducers work satisfactorily for frequency ratios greater than 3, they must necessarily have a low natural frequency, requiring incorporation of a large mass and soft spring in the transducer.

The choice of damping in such instruments is very critical as seen from the plot of Z/Y in Fig. 3.4. When ζ is of the order of 0.6–0.7, the percentage error in Z as compared to Y is very small. Introduction of artificial damping improves the usable lower frequency of the vibrometer, making it usable right from η equal to around 2. Besides, this value of ζ also offers a reasonably small settling time and low phase distortion.
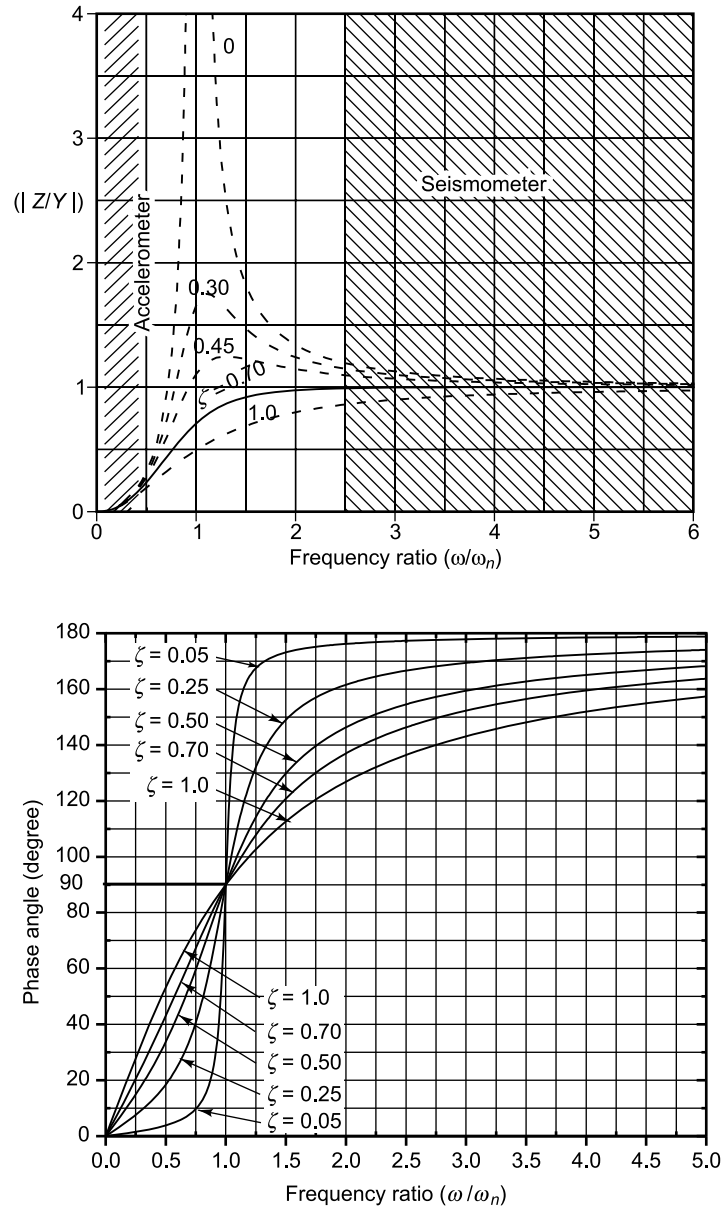
FIGURE 3.4 Response of seismic transducer

On the contrary, for values of η far less than 1, $Z/Y = \eta^2$. The acceleration of the casing will then be proportional to the relative displacement of the seismic mass with respect to the casing. These instruments are known as accelerometers. They have very high natural frequencies which are obtained using small masses and very stiff springs and are used up to around 0.6 times their natural frequencies.

For such transducers the frequency response function is

$$\frac{Z}{Y} = \frac{\omega^2}{\omega_n^2 \sqrt{(1-\eta^2)^2 + (2\zeta\eta)^2}}$$

(3.11a)

It is clear from Eq. (3.11b) that the denominator approaches unity for $\omega/\omega_n \to 0$.

$$\frac{1}{\sqrt{\{[1-(\omega/\omega_n)^2]\}^2 + [2\zeta(\omega/\omega_n)]^2}}$$

(3.11b)

Hence

$$Z = \frac{\omega^2 Y}{\omega_n^2} = \frac{\text{Acceleration}}{\omega_n^2}$$

(3.12)

Thus, Z becomes proportional to the acceleration of the motion to be measured with a factor of $1/\omega_n^2$. The useful range of the accelerometer can be seen from Fig. 3.5, which is a magnified plot of the term in Eq. (3.11b) for various values of damping ratio ζ as a function of η. The corresponding phase plot is the same as shown in Fig. 3.4.

Figure 3.5 shows that the useful frequency range of the undamped accelerometer is rather limited. When the damping ratio is between 0.6 and 0.7, the correction factor is within 4% of unity for η up to 0.5. As in the case of the displacement transducer, the usable frequency range of the accelerometer is extended by choosing a damping ratio between 0.6 and 0.7; in this case it is extended up to 0.5 times its natural frequency. Besides, $\zeta = 0.6 – 0.7$ ensures a good transient response. Another consideration is that this value of ζ also minimizes phase distortion for complex waves by ensuring linear phase, i.e. $\phi = \pi\omega/2\omega_n$ is satisfied with good accuracy.
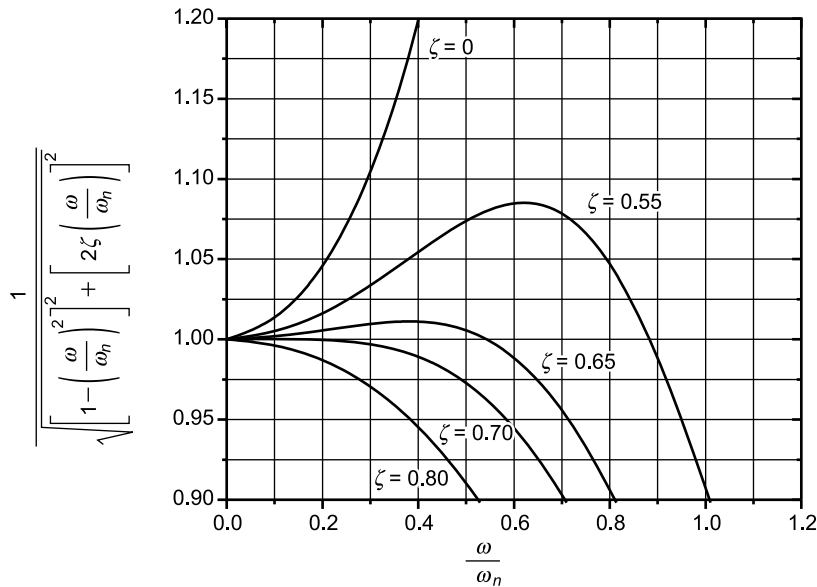
FIGURE 3.5   Acceleration error versus frequency

Let the vibration sensed by an accelerometer be

$$y(t) = Y_1 \sin \omega_1 t + Y_2 \sin \omega_2 t + Y_3 \sin \omega_3 t + \ldots \tag{3.13a}$$

Then the output of a seismic accelerometer with $\zeta = 0.7$ would be

$$z(t) = \left(\frac{\omega_1^2}{\omega_n^2} \right) Y_1 \sin \omega_1 \left(t - \frac{\pi}{2\omega_n} \right) + \left(\frac{\omega_2^2}{\omega_n^2} \right) Y_2 \sin \omega_2 \left(t - \frac{\pi}{2\omega_n} \right)$$

$$+ \left(\frac{\omega_3^2}{\omega_n^2} \right) Y_3 \sin \omega_3 \left(t - \frac{\pi}{2\omega_n} \right) + \ldots \tag{3.13b}$$

Such linear phase thus leads to the same time lag $\pi / 2\omega_n$ for all frequency components, ensuring that there is no distortion.

3.4.1 Seismic Displacement Transducer of the Inductive Type

The inductive pickup or LVDT is a passive transducer consisting essentially of a spring-mass damper system as shown in Fig. 3.6. This transducer can be used for the measurement of medium to large displacements at low frequencies.
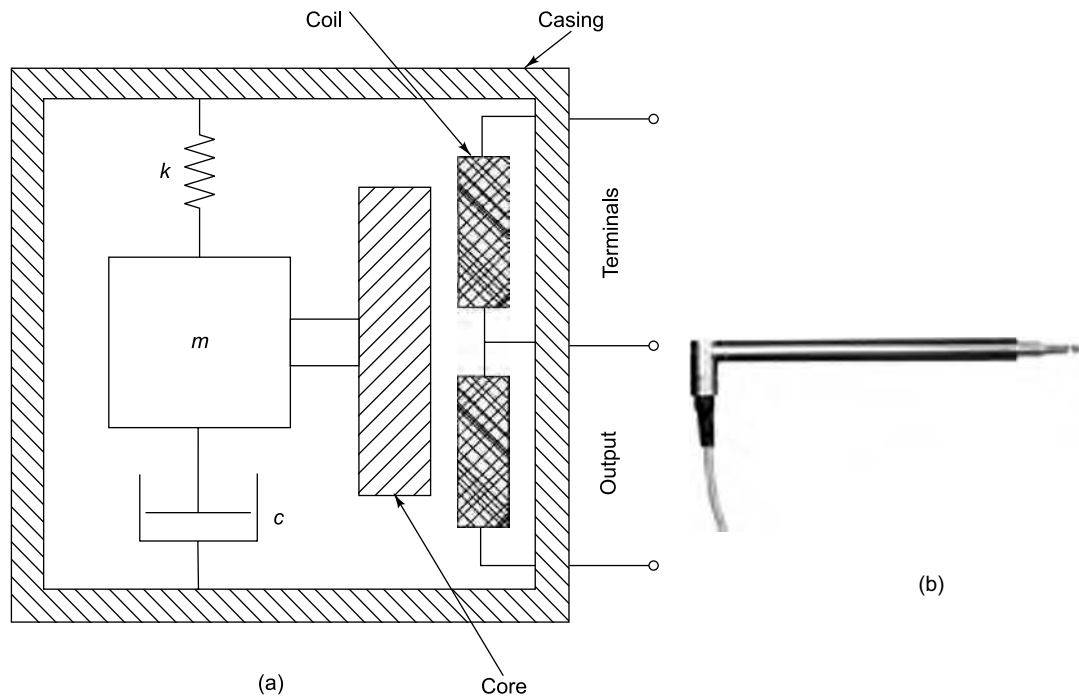
FIGURE 3.6 Inductive transducer: (a) schematic, (b) photograph (Courtesy of HBM, Germany)

The housing of the pickup is fixed to the object whose vibrations are to be sensed. The larger LVDTs (mass > 1 kg) are simply made to rest on the vibrating object after appropriate levelling, while the plungers of the smaller ones may be kept pressing against the vibrating object after the transducer is fixed to a stationary stand. The pickup is so designed that the mass remains more or less fixed in space for the specified frequency range of operation and hence the core connected to the mass serves as a satisfactory stationary reference. The core also provides the magnetic flux path linking the two inductive coils which are connected to the housing of the pickup and follow its motion. The two coils are connected in phase opposition to the centre, or null position, so that the output is the difference of these voltages. Hence the vibration to be measured is sensed as a relative motion between the core and the coils and is proportional to the change in inductance due to the relative motion. The null position of an LVDT is extremely stable and repeatable.

By designing the transducer to have a large mass and a soft spring, the change in inductance may be made proportional to the displacement of the vibrating object. The two inductive coils constitute two arms of an AC Wheatstone's bridge, the remaining two arms and the required AC excitation, conditioning amplifier and phase sensitive detector being provided by a carrier frequency amplifier. Initially the bridge is balanced to give zero output. Any vibration will now cause a change in inductance and hence bridge balance gets affected. The voltage corresponding to the amount of bridge unbalance can be measured on an oscilloscope or recorder. The inductive LVDT has a low natural frequency ω_n and can be used in the frequency range ω such that $1.414 \leq \omega / \omega_n \leq \infty$. The transducer is provided with a damping factor between 0.6 and 0.7 so that the measured relative motion is equal to the displacement of the vibrating object in the frequency range mentioned.

Specifications of typical LVDTs

		Large	Small
Natural Frequency	:	1 Hz	3 Hz
Frequency range	:	1–200 Hz	5–65 Hz
Vibration amplitude	:	±25 mm	±5 mm
Nominal sensitivity	:	80 m V/V	80 m V/V

Advantages
   (i) Relatively low cost.
  (ii) Solid and robust, capable of working in a wide variety of environments.
 (iii) No frictional resistance, since the iron core does not contact the transformer coils, resulting in an infinite (very long) service life.
  (iv) High signal to noise ratio (SNR) and low output impedance.
   (v) Negligible hysteresis.
  (vi) Excellent resolution limited only by the resolution of the amplifiers.
 (vii) Short response time, limited by the inertia of the iron core and the rise time of the amplifiers.

Disadvantages
   Readings affected by extraneous magnetic fields.

Applications
   (i) Vibration monitoring of huge structures, buildings, railway carriages, bridges, machine foundations, ship hull.
  (ii) Low frequency measurements.

3.4.2 Seismic Velocity Transducer: Electrodynamic Pickup

We have seen in Section 3.4 that the relative displacement sensed by a seismic transducer is related to the actual input displacement by Eq. (3.9). Since input velocity $\dot{Y} = \omega Y$ and the relative velocity sensed by the transducer is $\dot{Z} = \omega Y$, it follows that

$$\frac{\dot{Z}}{\dot{Y}} = \frac{\omega Z}{\omega Y} = \frac{Z}{Y} = \frac{\eta^2}{\sqrt{(1-\eta^2)^2 + (2\zeta\eta)^2}} \qquad (3.14)$$

The electrodynamic velocity transducer is a moving coil or moving magnet pickup of the seismic type. Also called a velometer, this transducer works on the principle of magnetic induction. If an electrical conductor is moved in a magnetic field in such a way that the magnetic flux through the conductor is changed, a voltage is induced in the conductor which is proportional to the rate of change of the magnetic flux. Conversely, if a current is sent through the same conductor, a mechanical force is exerted upon it, proportional to the current and the magnetic field. The resultant conversion from electrical to mechanical energy, or vice-versa depends on the polarities of voltage and current on the electrical side, or the directions of force and motion on the mechanical side, respectively. All electrical motors and generators and a host of other devices, such as solenoids and loudspeakers, utilize this principle.

In the electrodynamic vibration transducer, the relative motion Z in Eqs (3.9) and (3.14) is usually converted to an electric voltage by making the seismic mass a magnet moving relative to coils fixed in the case as shown in Fig. 3.7. When the pin fixed to the coil moves up and down within the magnetic field, a voltage is generated across the coil in proportion to the rate of cutting of the magnetic field or the velocity experienced by the vibrating body. The output velocity can be differentiated or integrated to get acceleration or displacement, respectively. The output of the instrument, i.e. the induced voltage e is as shown in Eq. (3.15a):

$$e = Blv \qquad (3.15a)$$

where B is the effective flux density in the winding in Wb/m^2 or Tesla, v is the relative velocity between the coil and magnet in m/s, e is the induced open circuit electromotive force (emf) in V and l is the length of conductor cutting the flux in m.
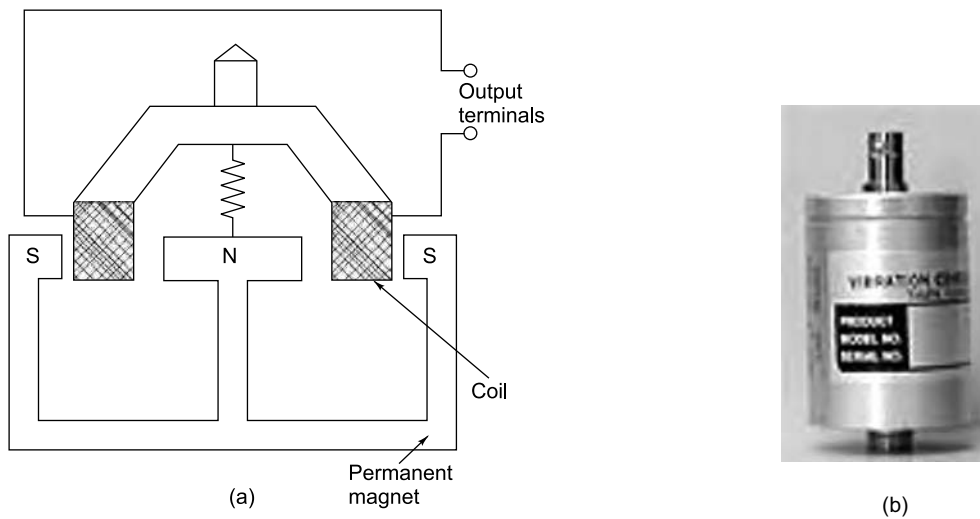
FIGURE 3.7   Electrodynamic transducer: (a) schematic, (b) photograph (Courtesy of http://www.vibcons.com)

The springs are initially compressed so that the vibrating object has to exert additional force to lower the pin into free motion in the magnetic path. The sensitivity (*S*) is given by

$$S = \frac{de}{dv} = Bl \tag{3.15b}$$

To increase the sensitivity, *B* and *l* or *l* should be increased. The value of *B* depends on the material of the permanent magnet. Further increase in sensitivity can be obtained only by increasing the length of the coil. But increase in length results in increase in weight of the coil, which may give rise to loading effect. Greater length with reduced loading effect can be obtained for a given mass of the coil by reducing its cross-section. This in turn results in a higher output resistance of the coil, necessitating the use of a voltmeter with a much higher resistance for reading the output voltage of the pickup. Thus, increasing the sensitivity is possible to a limited extent only.

Advantages
This is one of the most versatile transducers for machinery condition monitoring, for the simple reason that it costs less than other vibration transducers and can be easily installed on machines. It does not need any additional signal conditioning amplifiers and its output may be directly fed to an oscilloscope/ recorder/signal analyser. Besides, it gives vibration levels in terms of velocity, which is the parameter in general used to specify vibration severity criteria. Most velocity pickups have a frequency response in the range of 10–5000 Hz.

Disadvantage
This transducer is affected by extraneous magnetic fields.

Typical specifications
Sensitivity	= 100 mV/(cm/s)
Compression force	= 8 N
Maximum acceleration	= 20 g
Displacement	= 2 mm peak to peak
Frequency range	= 10–1000 Hz

Applications
  (i) Machinery vibration diagnostics
 (ii) Modal analysis

3.4.3   Seismic Inductive Accelerometer

The inductive accelerometer (Fig. 3.8) is similar in construction to the LVDT as shown in Fig. 3.6(a), the major difference being that the transducer is designed to have a small mass and a stiff spring such that the change in inductance is proportional to the acceleration of the vibrating object. It is a passive transducer consisting essentially of a spring mass damper system. This pickup is of the contact type and may be stud-mounted or fixed to the vibrating object using a magnetic base. The inductive accelerometer has a high natural frequency ω_n and can be used in the frequency range ω such that $0 \leq \omega/\omega_n \leq 0.7$. The measured relative motion for a seismic accelerometer is given by Eqs (3.9)–(3.12). The transducer is provided with a damping ratio between 0.6 and 0.7 so that the measured relative motion is proportional to the acceleration of the vibrating object in the desired frequency range. In contrast to the piezoelectric accelerometer which is essentially meant for high frequency applications, this accelerometer is ideally

suited for acceleration measurements at low frequencies. It may be recalled from Section 3.2.1 that generally displacement transducers are suggested for low frequency measurements.

<div align="center">**FIGURE 3.8**   Photograph of inductive accelerometer (Courtesy of HBM, Germany)</div>

Advantages and disadvantages

This transducer has to be used in conjunction with a carrier frequency amplifier, as in the case of the LVDT. It responds to static and dynamic measurements and has continuous resolution and a fairly high output. An accelerometer has the advantage that its output can be integrated once to get velocity and twice to get displacement. Its disadvantage is that external magnetic fields may cause erratic performance.

Specifications of inductive accelerometer

Natural Frequency   : 500 Hz
Nominal acceleration   : ± 1000 m/s^2
Nominal output signal : ± 80 m V/V

Applications

 (i) Measurement of ground motion
(ii) Measurement of low frequency accelerations (<1000 Hz) for modal analysis or machinery diagnostics

3.5   DISPLACEMENT TRANSDUCERS

Displacement measurements go hand in hand with pressure and force measurements. Displacement sensors are those that sense the variation of the position of a body. For vibration measurements, the vibration magnitude is measured in terms of the displacement of a machine or structure and is displayed in terms of millimetres (mm) or µm of displacement. There are a large number of displacement measuring devices (vibrometers) working on different principles. Among the electrical devices, all the three passive elements, namely, resistance, inductance and capacitance, are used for transducing displacement signals. The passive type of displacement transducer which is sensitive to changes in inductance and is called the LVDT has already been discussed in Section 3.4.1.

3.5.1   Eddy Current Transducer

This is a passive, non-contact type of transducer and uses the effect of eddy (circular) currents to sense the proximity of non-magnetic but conductive materials. An eddy current is a local electric current induced in a conductive material by the magnetic field produced by an active coil. This local electric current in turn induces a magnetic field opposite in sense to the one from the active coil and reduces the inductance of the coil. When the distance between the target and the transducer changes, the impedance of the coil changes correspondingly and can be detected by a bridge circuit. The eddy currents are confined to shallow depths near the conductive target surface. Their effective depth is given by

$$\delta = \frac{1}{\sqrt{\pi f \mu \sigma}} \tag{3.16}$$

where f is the excitation frequency of the circuit, μ is the magnetic permeability of the target material and σ is its conductivity. The target material must be at least three times thicker than the effective depth of the eddy currents.

The pickup consists of a ferromagnetic core on which two coils are wound: one (active) main coil which senses the motion of a conducting target and a second (balance) coil which serves to complete a bridge circuit and provide temperature compensation. Bridge excitation is a high-frequency (about 1 MHz) alternating current. Magnetic flux lines from the active coil pass into the conductive target surface, producing eddy currents in the target; the density of the currents is greatest at the surface and becomes negligibly small about three 'skin depths' below the surface. As the target comes closer to the probe, the eddy currents become stronger, changing the impedance of the active coil and causing a bridge unbalance which is a function of target position. This unbalance voltage is demodulated, low-pass filtered (and sometimes linearized) to produce a DC output proportional to target displacement. The high excitation frequency not only allows the use of thin targets, but also provides good system frequency response (up to 100 kHz). Figure 3.9 shows a schematic of the transducer with the bridge circuit arrangement and a photograph.

Probes are commercially available with full-scale ranges from about 0.25 to 30 mm (probe diameter 2–76 mm), non-linearity of 0.5% and a maximum resolution of 0.0001 mm. Targets are not supplied with the probes since the majority of applications involve non-contact measurement of existing machine parts, the part itself serving as target. Since target material, shape, etc. influence output, it is necessary to statically calibrate the system with the specific target to be used. For non-conductive targets, one must fasten a piece of conductive material of sufficient thickness to the surface. Commercially available adhesive-backed aluminium foil tape is convenient for this purpose. The recommended measuring range of a given probe begins at a 'standoff' distance equal to about 20% of the probe's stated range, i.e. a probe rated at 0–1 mm range should be used at target to probe distances of 0.2–1.2 mm.

Advantages

  (i) High sensitivity
 (ii) Flexibility to be used for static and dynamic measurement
(iii) Non-contacting measurement
 (iv) High resolution
  (v) High frequency response

Disadvantages

  (i) Effective distance is limited to small range.
 (ii) The relationship between the distance and the impedance of the coil is non-linear and temperature dependent. Fortunately, a balance coil can compensate for the temperature effect.
(iii) This pickup works only on conductive materials with sufficient thickness. It cannot be used for detecting the displacement of non-conductive materials or thin metalized films. However, a piece of conductive material with sufficient thickness can be mounted on non-conductive targets to overcome this drawback.
 (iv) Calibration is generally required, since the shape and conductivity of the target material can affect the sensitivity.

(a)

(b)

(c)
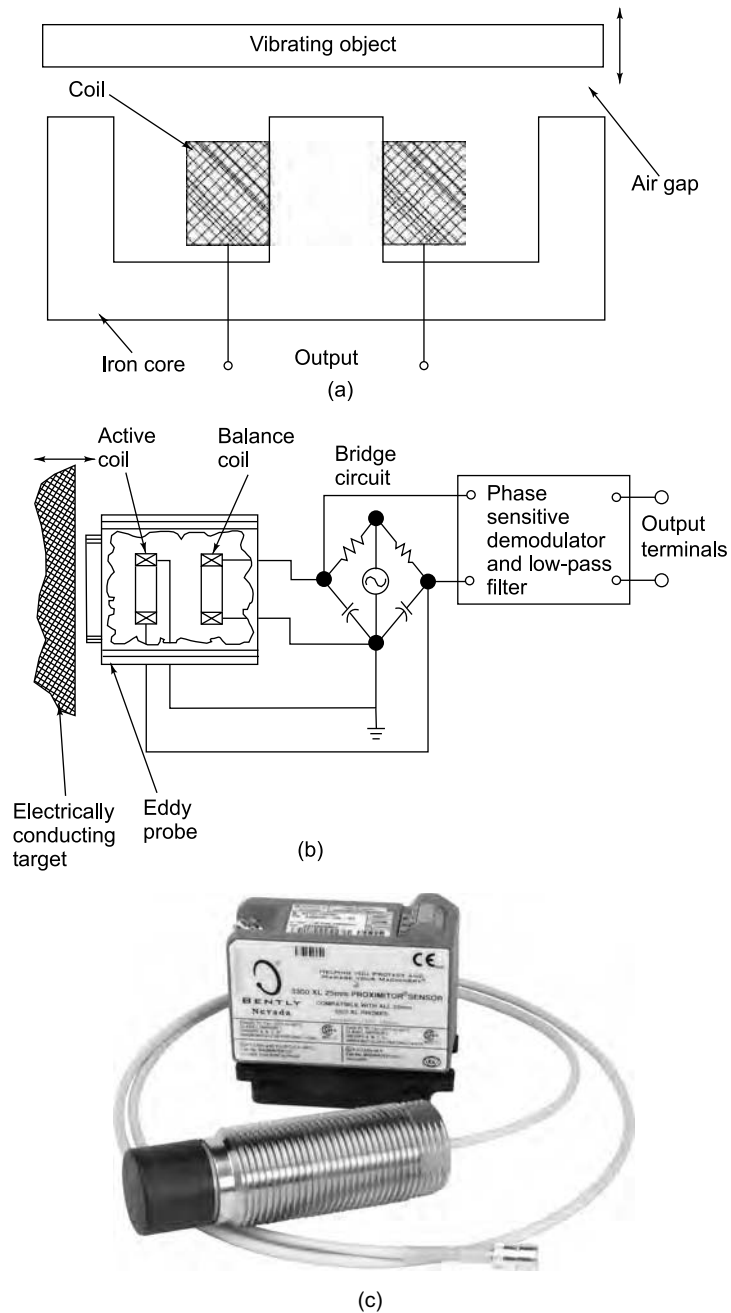
FIGURE 3.9   Eddy current transducer (proximity probe): (a) schematic of transducer, (b) transducer with bridge circuit, (c) photograph (Courtesy of http://www.ge-energy.com)

Typical specifications

Size	: About 2–75 mm in diameter, 20–40 mm long
Range	: 0.25–30 mm
Resolution	: Up to 0.1 μm
Non-linearity	: 0.5%
Bridge circuit frequency	: 50 kHz to 10 MHz

Applications

For measuring orbital motions of rotating shafts and various centring and alignment operations, special four-probe systems are available.

3.5.2 Capacitance Pickup/Condenser Vibrometer

The capacitance transducer is a passive, non-contact type of vibrometer. This parallel plate transducer (Fig. 3.10) works on the principle that the reactance of an ideal plate capacitor is inversely proportional to the distance between the plates. Its capacitance, C in pF is given as

$$C = 0.00886 \ \varepsilon A/d \tag{3.17}$$

where d is the distance between plates (mm), A is the overlapping area (mm^2) and ε is the dielectric constant (1.0006 for air).

From this equation, it is seen that the capacitance is dependent on the three quantities mentioned above. By varying any of these quantities and measuring the capacitance, we can find out the effect of change of that quantity. Variation in spacing of the parallel plates is often used for motion detection if the change in spacing is less than the electrode size and gives a conveniently large value of capacitance at small spacing. Signal conditioning is required to compensate for the parabolic capacitance-displacement relationship and this is easily done by measuring impedance rather than capacitance. The capacitance transducer is generally used for measurements in air and special care should be taken for measurements in liquids.
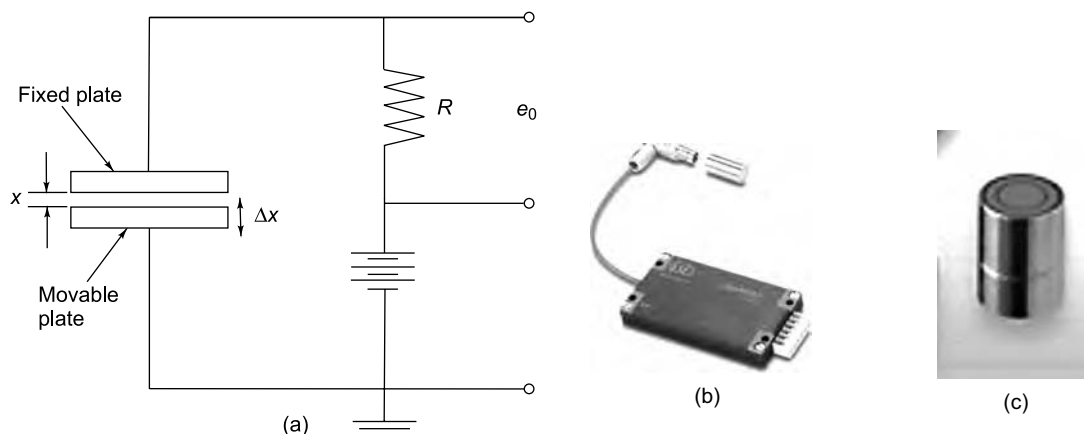
FIGURE 3.10 Capacitance transducer: (a) schematic, (b) photograph of transducer with measuring unit, (c) close up of transducer (Courtesy of http://www.micro-epsilon.co.uk)

Advantages
   (i) It has excellent linearity over the entire dynamic range when area is changed.
  (ii) The technology is low cost, is stable and uses simple conditioning circuits; often the offset and gain adjustments needed for most sensor types are not required.
 (iii) It offers freedom of electrode materials and geometry for demanding environments and applications.
 (iv) Fractional change in capacitance can be made large.
  (v) Capacitive sensors can be made to respond to displacements in one direction only.
 (vi) Capacitors are noiseless and hence excellent S/N ratio can be obtained.

Disadvantage
Special care and calibration is required for dielectric media other than what is given in the manual.

Typical specifications
   Measuring range   : 0–1 mm
   Frequency response : 0–1 kHz
   Sensitivity   : 10 V/mm

Typical applications
For non-contact measurements in lightweight and rotating machinery.

3.5.3 Fibre Optic Probe

The intensity modulated fibre optic displacement transducer is extensively being used for displacement measurements in vibration monitoring, what with its attractive features like simplicity, good performance, versatility and low cost. This sensor is based on the principle that, all other parameters being held constant, the light reflected back from a target surface varies with the distance that the transmitted light travels between the light source and the detector. Figure 3.11 shows a schematic diagram of this probe.
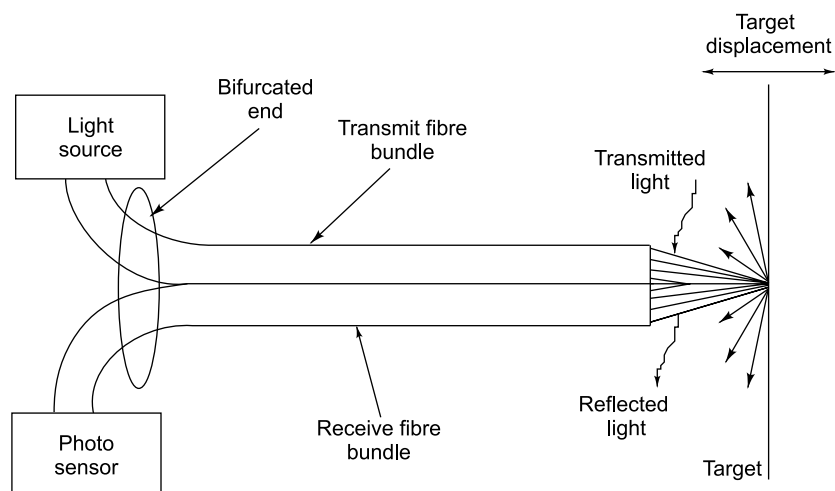
FIGURE 3.11   Schematic of an external intensity modulated fibre optic displacement sensor

The basic principle employed here is the use of a pair of fibre optic elements, one for carrying light from a source to an object/target whose displacement is to be measured and the other for receiving the light reflected from the target and carrying it back to a photo sensitive detector. By comparing the intensity of the transmitted light with that of the reflected light, the distance between the sensor and the target surface is determined. Figure 3.12 shows the relationship between the sensor output and distance from the target surface. This curve exhibits the typical 'front slope' and 'back slope' behaviour of such photo-detectors. For small sensor to target distances, more of the reflected light is captured by the receiving fibres producing the highly sensitive 'front slope'. As the target displacement increases beyond the 'optical peak', the intensity of the recaptured light decreases and produces a relatively wider 'back slope' in the output curve. Both regions are used in the calibration of the meter, the former for small gaps and the latter for large gaps. The gap at which the maximum, or zero slope occurs, provides a convenient and readily usable calibration reference position at which the output signal can be normalized to obtain a consistent sensitivity factor relatively independent of the colour or finish of the surface of the target or object under measurement.
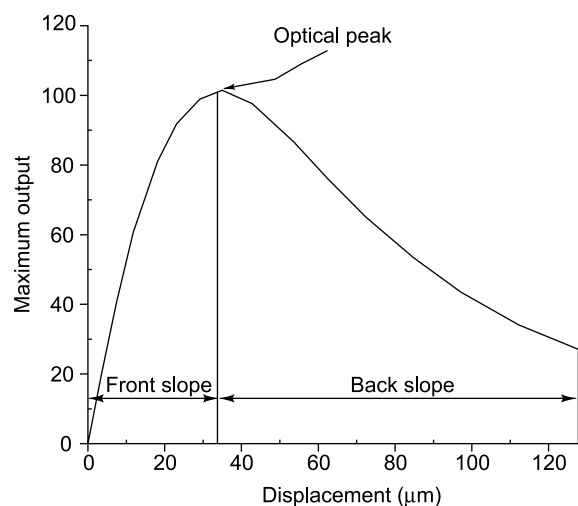
FIGURE 3.12   Fibre optic probe characteristics

In these sensors, light propagates by total internal reflection within the glass fibre, which is a flexible strand of glass or plastic, capable of transmitting light along its length by maintaining near total internal reflection of the light impinging at its input end. The most commonly used fibres are of the "step index" type and consist of an inner core of high refractive index (N_1) to carry the light flux and a surrounding concentric cladding of lower refractive index (N_2) such that total internal reflection occurs. Individual fibres usually fall in the range of about 25 μm diameter to 250 μm diameter, although recent advances in fibre optic manufacturing technology have extended the size up to about 1.5 mm. Transmission efficiency is dependent upon the composition and purity of the glass used in the core and cladding and on the quality of the optical finish on the end surfaces of the fibres. Most commercial sensors use multiple transmit and receive fibres to obtain the high levels of intensity at the photo-detectors needed to ensure acceptable levels of performance. Fibre optic transducers are available in a great variety of sizes,

configurations and fibre distribution patterns giving the user a broad choice of sensing range, resolution, frequency response and physical shape. Figure 3.13(a) shows the photograph of a typical fibre optic meter and Fig. 3.13(b) the probe tip.

(a)   (b)
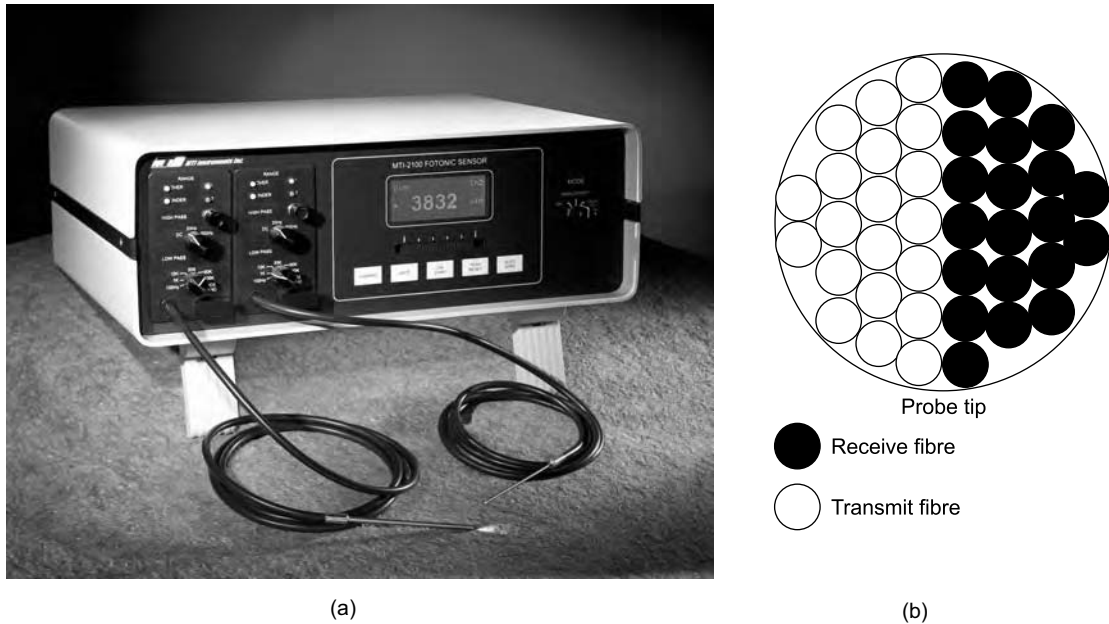
FIGURE 3.13   Fibre optic probe: (a) photograph (Courtesy of http://mtiinstruments.com), (b) probe tip

Advantages

  (i) A major advantage of the fibre optic transducer is its ability to operate directly with a large variety of surfaces and materials from conductors to insulators.

 (ii) It has inherent simplicity, versatility, and ease of use, small physical size, fast response, absence of mass loading due to its non-contacting nature and immunity to electromagnetic interference.

(iii) It possesses many outstanding characteristics, such as high frequency response, high resolution, high sensitivity, small measurement footprint and reasonable cost.

Disadvantages

  (i) The operating principles of these sensors make the experimental data susceptible to various measurement errors due to improper orientation of the probe, poor finish of the surface of the target, vibration of the probe clamp, etc.; particular attention has to be paid to these aspects.

 (ii) These measurement errors result in varying sensitivity and affect the linearity between measurement points, which translates into inconsistencies in the measured dynamic data.

(iii) The use of these sensors in the laboratory is somewhat more complicated than that of other transducer systems and extra care must be exercised in the experimental setup and data interpretation.

(iv) The fibre optic transducer is to be calibrated for use in different media such as air, water, and motor oil.

Applications
   (i) Modal analysis of small, lightweight parts or mechanisms such as read/write head of hard disk drive.
  (ii) Performance investigation of rolling element bearings.
 (iii) Measurement of very high frequency, small amplitude vibrations such as ultrasonically driven medical devices or welding equipment.
 (iv) Repeatability, hysteresis and response time measurement of precision mechanisms, or piezoelectric micropositioners.
  (v) Non-intrusive and remote vibration testing, especially in the presence of electric or magnetic fields.

Typical specifications

Resolution	:	0.3–12 μm
Sensitivity	:	0.012–0.75 μm/mV
Standoff	:	0.05–5 mm
Target footprint	:	0.17–2.2 mm
Linearity	:	0.003–0.015
Frequency response	:	0–70 kHz

3.5.4   Magnetostrictive Displacement Sensors

Ferromagnetic materials, such as iron, nickel, and cobalt, change size and/or shape when placed in a magnetic field. This property is called magnetostriction and is due to the presence of magnetic moments in the material. The material may be visualized as a collection of tiny permanent magnets, or domains, each domain consisting of many atoms. When the material is magnetized, the domains which were originally randomly arranged, are re-oriented with their axes approximately parallel to one another. Alloy selection, thermal annealing, cold working and magnetic field strength control the ordering of the domains.

In these metals, the $3d$ electron shell is not completely filled, allowing for the formation of a magnetic moment. As electron spins are rotated by a magnetic field, coupling between the electron spin and electron orbit causes electron energies to change. The crystal then strains so that electrons at the surface can relax to states of lower energy. When a material has positive magnetostriction, it enlarges when placed in a magnetic field; with negative magnetostriction, the material shrinks. The amount of magnetostriction in base elements and simple alloys is small, of the order of μm/m. The reverse effect is also true: applying stress to a magnetostrictive material changes its magnetic properties (e.g. magnetic permeability). This is called the Villari effect.

Any wire made of a magnetostrictive material displays what is called the Wiedemann effect. When an axial magnetic field is applied (usually from a permanent magnet) to a magnetostrictive wire, and a current is passed through the wire, a twisting occurs at the location of the axial magnetic field. This twisting is due to interaction of the applied axial magnetic field, with the magnetic field along the magnetostrictive wire generated by the current in the wire. The current is generally applied as a short duration pulse, typically 1 or 2 μs. The maximum current density and hence the maximum magnetic field intensity is produced at the wire surface due to skin effect, developing the waveguide twist. Due to the current pulse, the mechanical twisting travels as an ultrasonic wave in the wire, making the wire

behave as a 'waveguide'. The wave travels at the speed of sound in the waveguide material, approximately 3000 m/s.

The operation of a magnetostrictive position sensor is shown in Fig. 3.14. This is a transducer that detects the position of a magnet which moves along the length of the sensing element and is attached to the object, the position of which is to be determined. It uses the Villari effect. A small piece of magnetostrictive material, called the tape, is welded to the waveguide near one end. This tape passes through a coil and is magnetized by a small permanent magnet called the bias magnet. When a current or interrogation pulse passes through the waveguide, a second magnetic field is created radially around the waveguide. The interaction between the magnetic field in the waveguide and that produced by the position magnet generates a strain pulse. This travels at a constant ultrasonic speed from its point of generation to the end of the waveguide where it is transformed into an electric voltage output pulse in the sensor element. The exact position of the magnet is found from the time interval between the initiation of the current pulse and the arrival of the torsional strain pulse. The product of this time interval and the travel speed gives the position. This is detected and conditioned into the desired output by electronic circuitry. A damping element is provided at the end of the waveguide to absorb the strain pulse in the waveguide.
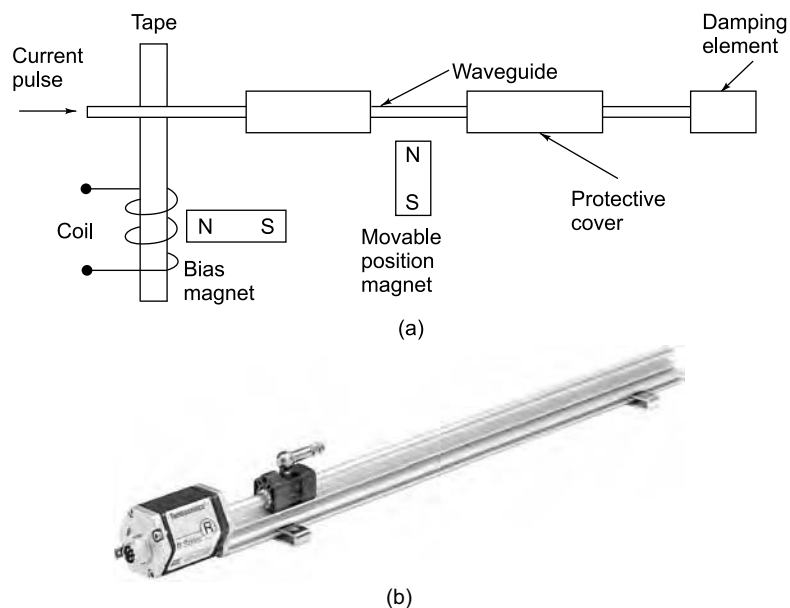
FIGURE **3.14**   Magnetostrictive transducer: (a) principle of operation, (b) photograph
(Courtesy of http://www.mtssensor.de)

Advantages
High resolution, of the order of 10s of μm.

Disadvantages
Not as easy to use as other displacement transducers.

Typical applications
As a linear position sensor for fluid power and process control applications.

Also in automotive, biomedical and military applications.

Typical specifications

    Time response : 0.2–5 ms
    Stroke length   : 50–7000 mm
    Resolution   : 10 μm
    Non-linearity   : < ±0.05%

3.5.5 Holography

Holography is a laser based interferometric technique for non-contact displacement measurements. It is a relatively new measurement procedure, exploiting the characteristics of the coherent, monochromatic light beam created by a laser, for extremely accurate and effective non-intrusive measurement and for response measurement in experimental modal analysis. Holography is essentially a process which enables the recording and storage of a light wave and later the reconstruction of that light wave in complete detail. In this method a deformed surface is illuminated with a coherent light source and the light scattered by the surface is recorded; using this, a complete three-dimensional image of the deformed object can be reconstructed later.

Figure 3.15(a) shows the basic holographic setup for recording. The coherent laser beam is split by a beam splitter into two beams: the reference beam and the illumination beam for the object. The reference beam is reflected onto a holographic plate (recording medium), while the illumination beam is scattered by the object surface to produce the object beam, this beam also being recorded on the holographic plate. Thus at any instant of time, the combination of differing path lengths and incident angles between the reference beam and the object beam will cause constructive and destructive optical interference fringes. This relative interference between the two beams is recorded to produce a photographic image called the hologram. By varying the configuration of the setup so as to vary the angle between the two beams, the distance between the fringes and thus the displacement resolution can be changed. A time varying fringe pattern is produced when one of the beams is reflected from a vibrating target. The fringe pattern is sampled at different frequencies to obtain the dynamic target surface displacement. On re-illumination of the processed holographic plate with a reconstruction beam (reference beam similar to the original beam), diffraction from the fringe pattern on the film reconstructs the original object beam in both intensity and phase as depicted in Fig. 3.15(b). Figure 3.15(c) shows a photograph of the holography equipment.

Of late, electronic-speckle pattern interferometry (ESPI), which is a method for increasing the speed of the basic holographic technique, is gaining attention. This is a state of the art technology which is well suited for non-intrusive measurement of vibration modes in experimental modal analysis. It uses real-time video recording and display over and above the basic interference techniques of holography. The systems presently available are able to produce resonant mode shapes only, the displacement information at a specific point on a target not being easily obtainable.

Typical specifications

    Laser   : 100 mW
    Frequency range   : 20 Hz–50 kHz
    Object size   : 1 cm^2–4 m^2
    Amplitude resolution : <1 nm
    Maximum amplitude   : 10 μm
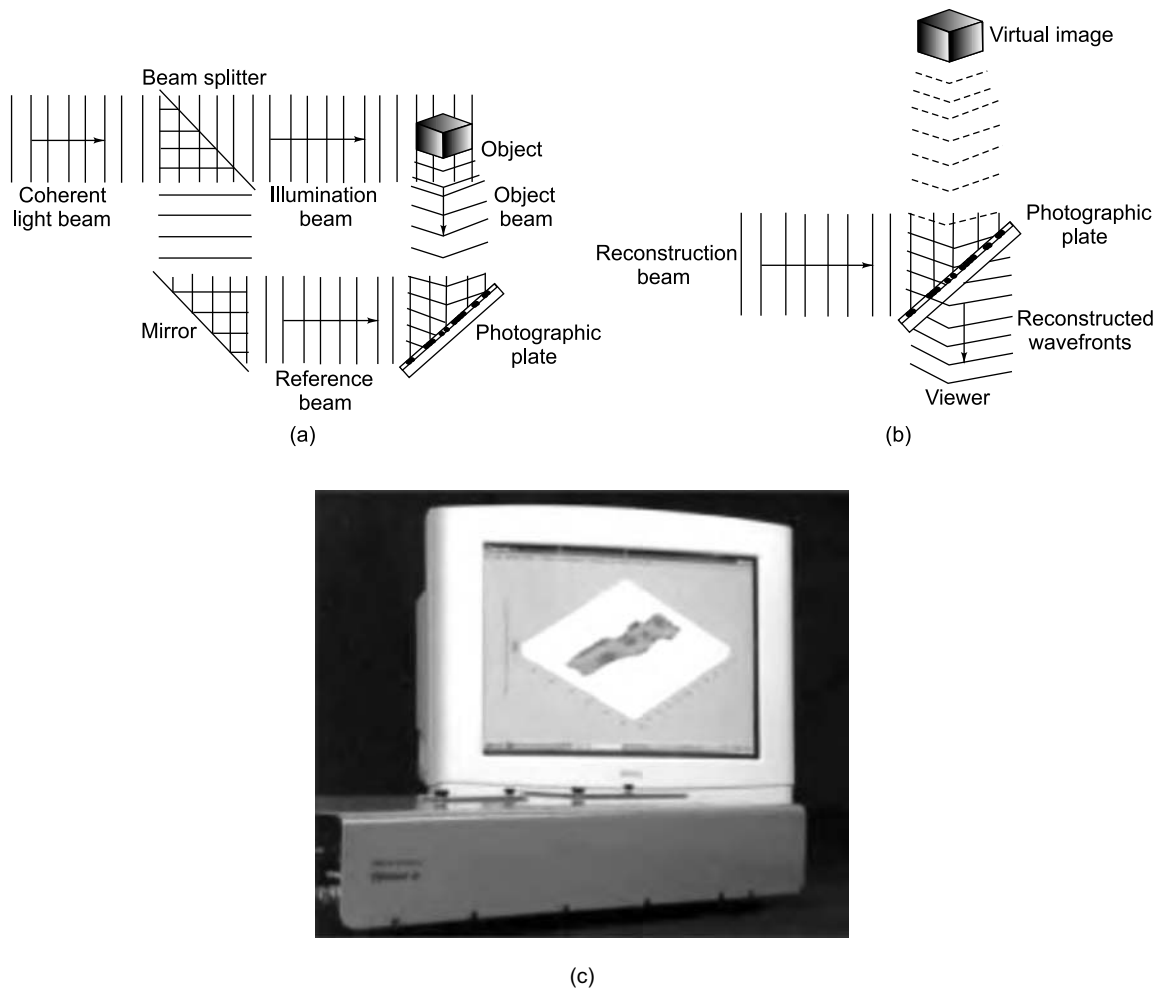
FIGURE 3.15   Basic holographic process: (a) recording process, (b) reconstruction process, (c) holography equipment (Courtesy of http://www.optonor.no)

Advantages

Holography offers the distinct advantage of being capable of measuring the entire three dimensional vibrating surface at one time.

It offers extremely high resolution, up to 2×10^{-6} mm.

Disadvantages

   (i)   The system is very expensive.
  (ii)   Vibration free table is required for mounting the equipment.
 (iii)   Standard holography techniques suffer from slow, complex post-experimental development of the actual three dimensional displacement diagrams.

(iv) Numerical displacement values which are required for modal parameter extraction are difficult to obtain from these holograms.

Applications

For non-destructive testing in the aerospace, automotive and other defence related industries.

3.6 VELOCITY TRANSDUCERS

Velocity transducers are very popular and come with most commercial vibration meters for the reasons that velocity is used for specifying vibration severity criteria and that these transducers cover a fairly wide range of frequencies. Besides, they are also suitable when the measurement frequency range is not known apriori.

3.6.1 Electromagnetic Transducer

This is a very inexpensive and useful vibration pickup of the non-contact type. It is an active transducer. In the electromagnetic transducer, a coil is wound directly on the core of a permanent magnet as shown in Fig. 3.16(a). Figure 3.16(b) shows a photograph of the same. When a ferromagnetic body placed opposite to one of the poles of the magnet is moved with respect to the magnet, a change in the flux ϕ linking the coil is obtained. This causes a voltage proportional to $d\phi/dt$ to be generated in the coil. The voltage may be expressed as

$$e_0 = k\frac{d\phi}{dt} \tag{3.18a}$$

$$= k\frac{d\phi}{dy}\frac{dy}{dt} \tag{3.18b}$$

$$= k\frac{d\phi}{dy}\frac{d(y_0 + x)}{dt} \tag{3.18c}$$

$$= k\frac{d\phi}{dy}\frac{dx}{dt} \tag{3.18d}$$

where y is the instantaneous gap, y_0 the average gap between the face of the magnet and the ferromagnetic body and k is a proportionality constant. As can be seen from Eqs (3.18a) to (3.18d), the induced voltage is directly proportional not only to the velocity of vibration, but is also dependent on the gap y. When the pickup is located near a vibrating ferromagnetic body, a voltage proportional to the velocity of vibration will be induced and as the average distance increases, sensitivity reduces. Thus, $d\phi/dy$ forms the sensitivity factor indicating that the sensitivity is inversely proportional to the gap. In principle, this method can be employed to measure absolute or relative velocities. The air gap should be at least five times the vibration amplitude for distortion to be less than 2%.
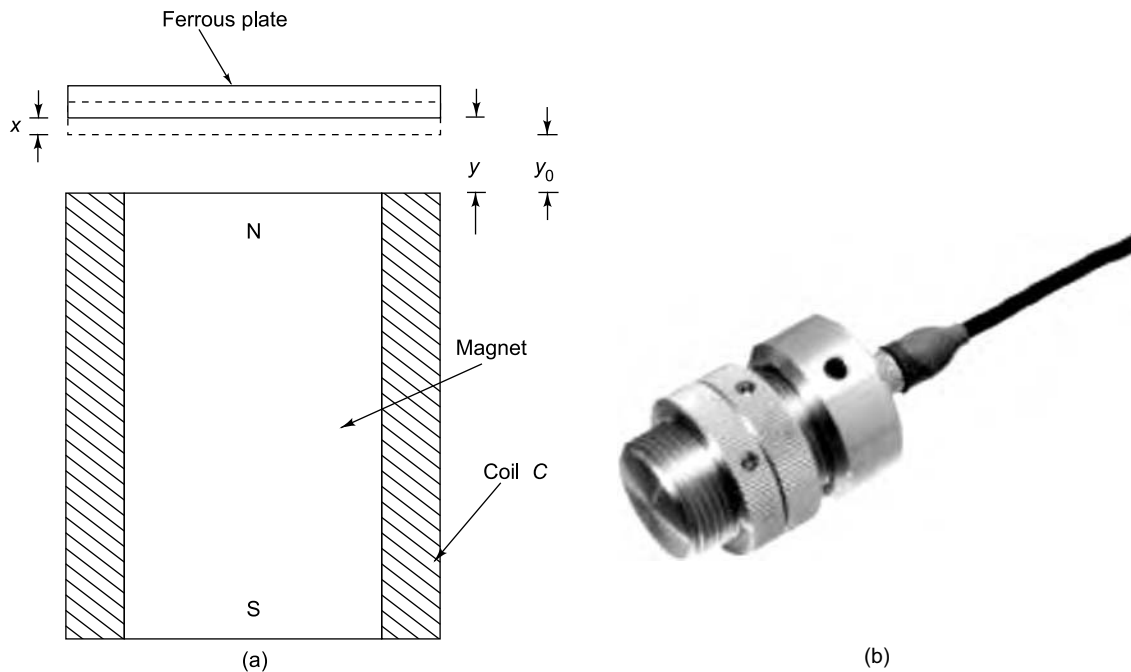
FIGURE 3.16   Electromagnetic pickup: (a) schematic, (b) photograph (Courtesy of Brüel & Kjær, Denmark)

Advantages
 (i) They require no external supplies since they are of the self-generating (active) type.
 (ii) High sensitivity of the order of 200 mV/(cm/s) is obtainable, without amplification.

Disadvantages
   Strong external magnetic fields can adversely affect the performance of the pickup.

Typical specifications
Sensitivity   :   20 mV/(cm/s) at 1 mm air gap
Frequency range :   DC to 1000 Hz
Weight   :   0.5 N

3.6.2   Laser Doppler Vibrometer

A Laser Doppler Vibrometer (LDV) is a non-contact velocity transducer based on the principle of the detection of the Doppler shift of a coherent laser light that is scattered from a moving object. Light from a laser beam is made to fall on a vibrating object, the surface vibrations of which result in scatter or reflection of the light, inducing a Doppler frequency shift on the laser beam. This shift is linearly related to the component of velocity which lies along the axis of the laser beam and is measured using an interferometer. The electronic circuitry converts the Doppler signal to an analogue voltage proportional to the instantaneous velocity of the target. Two moving mirrors driven by galvanometric actuators are used to direct the laser beam to the desired measurement points. Such an instrument is called a Scanning

Laser Doppler Vibrometer (SLDV) and can quickly perform velocity measurements on a grid of points over the structure under test.

The main problem in this method is that the Doppler shifts obtained are usually very small when compared to the very high laser fundamental frequency (approximately 4.74×10^{14} Hz), typically 1 part out of $10^8 - 10^{14}$. The only way to handle such small quantities is to use interferometry. An optical interferometer is therefore used to mix the scattered light coherently with a reference beam; the high frequency oscillations are thus combined and reduced to much lower values that can be dealt with by standard electronics. Such an arrangement can be a Michelson interferometer as shown in Fig. 3.17.
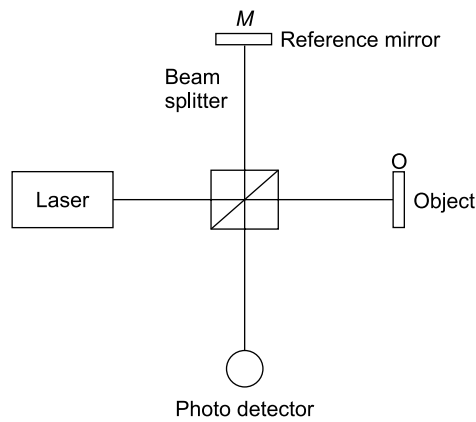
FIGURE 3.17   Michelson interferometer

In this arrangement a laser beam is divided by a beam splitter into a measurement beam and a reference beam. Let the distances travelled by the light from the beam splitter to the reflectors be x_R and x_M for the reference mirror M and object O, respectively. Let the optical phase of the reference beam in the interferometer be $\phi_R = 2kx_R$ and that of the measurement beam be $\phi_M = 2kx_M$ where $k = 2\pi/\lambda$. The time varying relative phase is then $\phi(t) = \phi_R - \phi_M = 2\pi \Delta l / \lambda$ where Δl is the vibrational displacement of the object and λ the wavelength of the laser light. The rate of change of phase is proportional to the rate of change of position, which is the vibrational velocity v of the surface. Let $I(t)$ denote the time dependent intensity at the point where the measurement and reference beams interfere. If Δl changes periodically, the light intensity $I(t)$ also varies in a periodic manner. A phase change ϕ of 2π corresponds to a displacement Δl of $\lambda/2$ giving rise to a Doppler frequency $f_D = 2v/\lambda$. Vibrometers usually employ He–Ne lasers, giving rise to a Doppler frequency shift f_D of about 3.16 kHz for each mm/s.

The optical signal is converted to an electrical signal by a photo detector which measures the intensity of the mixed light and the beat frequency which is equal to the difference in frequencies between the reference and measurement beams. The intensity $I(t)$ is

$$I(t) = I_R I_M + 2K \sqrt{I_R I_M R} \cos (2\pi f_D t + \phi) \qquad (3.19)$$

where I_R and I_M are the intensities of the reference and measurement beams, K is a mixing efficiency coefficient and R is the effective reflectivity of the surface.

It is seen that $I(t)$ is a frequency modulated (FM) signal, and, by demodulating this FM signal, it is possible to obtain the amplitude of v; however information regarding the direction of the surface velocity

is missing because of the cosine function. The most common solution to obtain directional sensitivity is by introduction of an optical frequency shift into one arm of the interferometer to obtain a virtual velocity offset. Figure 3.18 shows the schematic of an LDV setup where an acousto-optic modulator (Bragg cell) is incorporated into one arm of the interferometer.
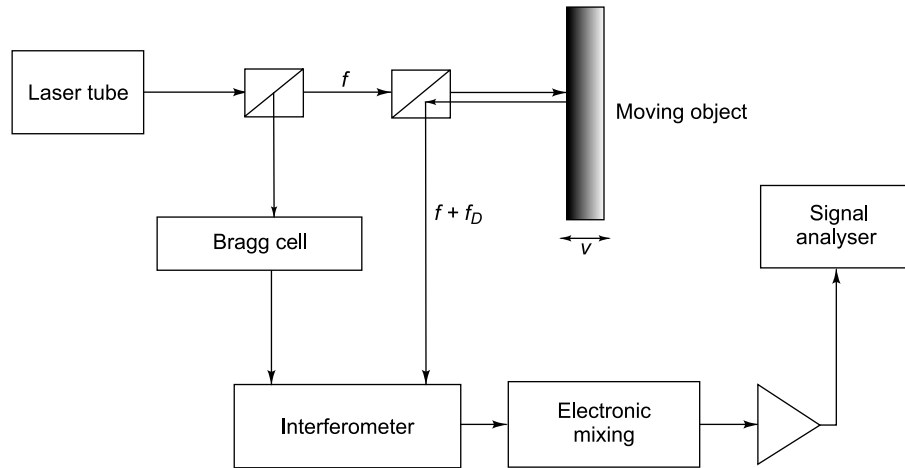
FIGURE 3.18 Schematic of an LDV

The Bragg cell is driven at frequencies of 40 MHz or higher and generates a carrier signal at the drive frequency. The frequency of movement of the object modulates the carrier signal. The object velocity (vector) determines the direction and amount of frequency deviation with respect to the centre frequency f_B.

With the introduction of a shift frequency f_B the intensity at the detector changes to:

$$I(t) = I_R I_M + 2K \sqrt{I_R I_M R} \cos \left(2\pi (f_B - f_D)t + \phi\right) \qquad (3.20)$$

For signal decoding, one can process the phase to produce a displacement output or carry out an FM demodulation to provide the vibrational velocity. Figure 3.19 shows a photograph of a commercially available LDV.

Typical specifications

Laser power	:	< 1 mW
Maximum velocity	:	1 to 10 m/s
Upper frequency limit	:	200 kHz
Resolution	:	1 μm/s
Maximum stand off distance	:	10s of m

Applications

When mounted on to a tripod, it can be easily pointed at a vibrating object. This technique is effectively used in structural dynamic testing, on-line monitoring of industrial plants, acoustics, fault detection, etc. Besides, the coupling of laser vibrometers and scanning systems seems to open up new possibilities, e.g. in the field of measurements in tracking mode on moving objects.

FIGURE 3.19 Laser Doppler vibrometer (Courtesy of Brüel & Kjær, Denmark)

Advantages
  (i) The single-point LDV can measure velocity and displacement of vibrating structures completely without contact.
 (ii) Most LDVs have low power levels so that no special safety measures are required.
(iii) Results are storable in digital formats like BMP and JPG images or AVI movies or TXT text data files.

3.7   ACCELERATION TRANSDUCERS

Accelerometers find widespread applications in the following areas:
- Experimental modal analysis
- High precision internal navigation and guidance systems
- Vibration control
- Automotive industry: release of airbag is the most well known application
- Also for suspension control, brake control, fuel cut off and engine knock off
- Measurement of motion of patients with diseases such as Parkinson's

3.7.1   Piezoelectric Accelerometer

Piezoelectric transducers are the most widely used in the world of shock and vibration, the reason being their very wide frequency response to amplitude and phase. These pickups involve a class of materials which, when mechanically deformed, produce an electric charge. They are very small in size (extending down to a fraction of a gram) and are of the active type, producing charge variations as output. Unlike other sensors, piezoelectric transducers have a reversible effect, deflecting mechanically when subjected to an applied voltage.

The piezoelectric transducer uses Barium Titanate or Lead Zirconate Titanate or Lead Niobate or Lithium Niobate or other suitable piezoelectric crystals as the basic element. This crystal is placed inside a casing with a hard spring pressed against it via a block of mass (Fig. 3.20). This stiff spring-light mass combination has a high natural frequency, around 20,000 Hz and is used as an accelerometer. When the whole assembly is subjected to vibration, the mass exerts a variable force on the piezoelectric crystal and the latter in turn develops a voltage change across its faces. For a wide range of frequencies, the spring mass system develops a voltage or charge proportional to the acceleration to which the transducer is subjected. As the transducer has high impedance, a charge amplifier would be essential to use it to its full capability. The piezo pickup is a contact type of absolute measuring device. It is generally fixed by stud mounting or with a magnetic base. The smaller devices are fixed using bee's wax.

The sensitivity of the crystal accelerometer is given either in terms of charge (pC) per g or in terms of voltage per g (mV/g). Since the voltage is related to the charge by the equation $E = (Q/C)$, the capacitance of the crystal, including the shunt capacitance of the connecting cable, must be specified. Typical sensitivity of a crystal accelerometer is 20 pC/g, with a crystal capacitance of 500 pF. The equation $E = (Q/C)$ then gives 20/500 = 0.04 V/g or 40 mV/g for sensitivity in terms of voltage. This severe loss of signal due to cable shunt capacitance can be avoided by using a charge amplifier.
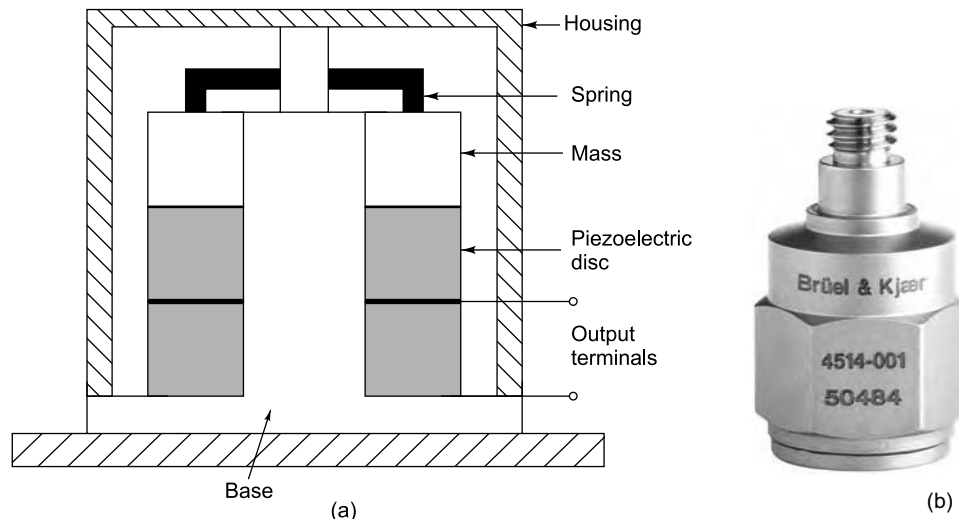
FIGURE 3.20   Piezoelectric transducer: (a) schematic, (b) photograph (Courtesy of Brüel & Kjær, Denmark).

Today we have commercially available 'ICP' accelerometers where ICP, is PCB's registered trademark standing for 'Integrated Circuit-Piezoelectric'. This class of transducers has built-in, signal conditioning electronics which converts the high-impedance charge signal produced by the piezoelectric crystal into a usable low-impedance voltage signal. Such a signal can be easily transmitted over ordinary two-wire or coaxial cables over long distances, to any voltage measuring or recording device. The special electronics typically also has incorporated in it other signal conditioning features, such as gain, filtering, and self-test features. Since these ICP sensors are simple to use, have high accuracy, wide frequency range, and low cost, they are very popular in many shock and vibration applications.

Advantages
  (i)   One of the most popular transducers since it can ideally be used with portable instruments.
 (ii)   Wide frequency range.
(iii)   Very low mass loading due to small weight.
 (iv)   Readings not affected by extraneous magnetic fields.
  (v)   Very useful for experimental modal analysis since this throws open the roving pickup option.

Disavdantages
  (i)   Not suitable for low frequencies.
 (ii)   Not meant for high temperature applications (for this special probes are required).
(iii)   Requires charge amplifier.

Typical specifications
Voltage sensitivity   :   50 mV/g
Charge sensitivity   :   55 pC/g

Resonance frequency	:	30 kHz
Transverse sensitivity	:	<3%
Dynamic range	:	±3000 g
Mass	:	20 g

Typical applications

For general-purpose shock and vibration measurements, modal analysis, machinery diagnostics.

3.7.2 Microelectro-Mechanical Systems (MEMS) Based Accelerometers

Modern machinery diagnostic systems employ a new family of sensors called micromachined accelerometers. Micromachining refers to the technique of manufacturing tiny moving mechanical structures on a single wafer of, say, silicon. Techniques developed from the manufacture of large scale integrated (LSI) semi-conductor electronic components such as deposition, photolithography and etching are employed for making micromachined devices. The new accelerometers include both the signal conditioning circuitry and the sensor, fabricated together on a single monolithic chip at a very low cost with high reliability and make them the perfect candidates for modal testing applications. Previously, modal testing was limited by the cost of the accelerometers and of the associated signal conditioning circuitry, such as charge amplifiers and the high cost of the setup. Apart from the sensors, the cost of calibration was also responsible for the increased cost of testing. The mass of the standard accelerometers, i.e. 100 to 200 gm often produced mass loading effects. Reduced cost of modal testing with improved accuracy is now possible due to MEMS accelerometers, since a larger number of sensors with a lower cost can be used. The sensitivity of the micromachined accelerometers stays within 1% of the full-scale and its shock survival is approximately of the order of 2000 g.

A piezoelectric MEMS sensor (Fig. 3.21a) is very similar in construction to a conventional piezoelectric accelerometer, except that the fabrication is by an etching process. In one design of a MEMS single axis accelerometer, a proof mass is supported on a frame, by means of one or more compliant beams. When acceleration is applied to the frame, each beam is deformed by the force required to accelerate the proof mass, causing it to move relative to the frame. The motion of the proof mass is controlled by the resilient nature of the beams, which apply a restoring force on the mass. Acceleration can be measured by sensing the strain in each beam that supports the proof mass, typically using either piezoelectric or piezoresistive sensors present on each beam. An alternative design of MEMS accelerometer uses a capacitive measurement technique. Here a proof mass is supported by one or more resiliently-deformable beams, the mass carrying one plate of a capacitor and the frame carrying the other plate. Acceleration is sensed by measuring the change in capacitance caused by the deflection of the inertial mass which changes the capacitance between the two plates. A third form of MEMS accelerometer uses a torsion member to constrain a proof mass and a capacitive or servo-capacitive arrangement is used to measure the displacement of the mass when the accelerometer is subjected to acceleration. The torsional stiffness of the support member controls the displacement of the proof mass, or electrostatic forces generated by the servo-capacitors control that displacement. Figure 3.21(b) shows the photograph of a MEMS accelerometer.

Conventional tri-axial accelerometers, usually consist of an assembly of three single axis accelerometers, arranged with their sensing axes orthogonal to each other, making the resultant three-axis accelerometer significantly larger and heavier than intended. MEMS technologies have led to the production of devices which are extremely small and yet are capable of giving very accurate and reliable indications of acceleration

in three dimensions. The required alignment accuracy can be achieved using lithographic etching processes, and no subsequent assembly processes are required to complete the basic structure of the three-axis accelerometer.
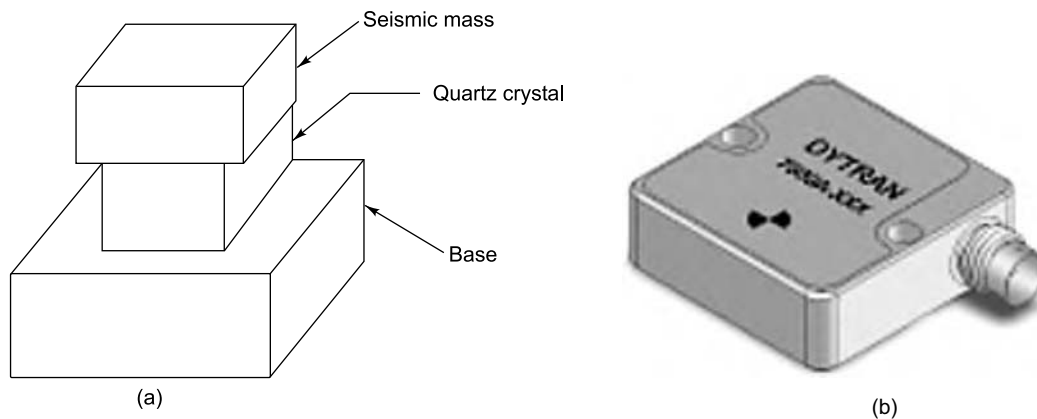
FIGURE 3.21 MEMS accelerometer: (a) schematic, (b) photograph (Courtesy of http://www.dytran.com)

Broadly, there are three types of triaxial MEMS accelerometers. In the first arrangement three separate single-axis MEMS accelerometers are mounted onto three faces of a cube, to measure accelerations in three directions. The whole assembly of the individual wafers and the mounting cube significantly increases the weight of the complete 3-axis accelerometer. Further, difficulties in aligning the three accelerometers with great accuracy results in significant manufacturing difficulties and so there is a cost penalty. In the second arrangement, a MEMS accelerometer with a single mass is used to sense acceleration in three orthogonal directions. Ideally the sensitivities in all directions should be equal, but in practice the out-of-plane response (with respect to the wafer) is usually several times larger than the in-plane response. Isolation of the individual signals for each direction is limited by the accuracy of manufacture of the device and the requirement for equal signals from each axis, leading to cross-axis signals. The performance of such a device is consequently compromised. In the third case, three identical single axis MEMS devices are produced in a single wafer, to sense acceleration in three directions, but a typical three-axis accelerometer manufactured thus cannot produce exactly the same strain distribution in the support beams for the proof mass in response to in-plane and out-of-plane accelerations.

Advantages
  (i) These are high performance, high accuracy, high reliability and low cost accelerometers with low power consumption.
 (ii) They have very low mass, giving rise to very low mass loading.
(iii) They are suitable for measurements in harsh vibration environments, having high shock withstand capability and high temperature operation.
(iv) These include both the signal conditioning circuitry with low noise electronics and the sensor, fabricated together on a single monolithic chip at a very low cost.
 (v) They are suitable for measurements at low frequencies down to DC.
(vi) Besides, they are immune to electromagnetic interference.

Typical applications
These accelerometers are the perfect candidates for modal testing involving roving transducer applications. Other applications include crash testing, air bag testing, robotics, seismic monitoring and tilt measurement.

Specifications

Acceleration range	:	±50 g
Output voltage	:	±2.0 V DC @ ±50 g
Sensitivity	:	40 mV/g, ±5%
Bandwidth	:	DC to 1000 Hz
Cross axis sensitivity	:	< 2%
Shock	:	>1000 g

3.7.3   Servo Accelerometer

These devices were originally designed for use in inertial guidance systems in aircraft and missiles, but have subsequently found application in the measurement of vibrations of large structures which poses special problems. Such structures have low natural frequencies and the associated accelerations are very small. This demands that the accelerometer be very sensitive and suitable for use right from near zero frequency. But unlike many other applications, the motion of large structures is relatively immune to the loading of the accelerometer mass, allowing the use of accelerometers which are heavier than piezoelectric devices. The servo accelerometer or force-balance accelerometer is ideally suited for such applications.

A servo accelerometer measures the acceleration of the structure on which it is mounted by measuring the force required to prevent the proof mass from moving relative to the instrument frame under acceleration. A schematic drawing of the servo accelerometer is shown in Fig. 3.22(a). Figure 3.22(b) shows the photograph of a commercially available servo accelerometer. A displacement transducer is used to sense the relative movement of the mass with respect to the instrument frame. This signal is amplified and fed back as direct current to the force coil suspended in a magnetic field, generating the required restoring force for
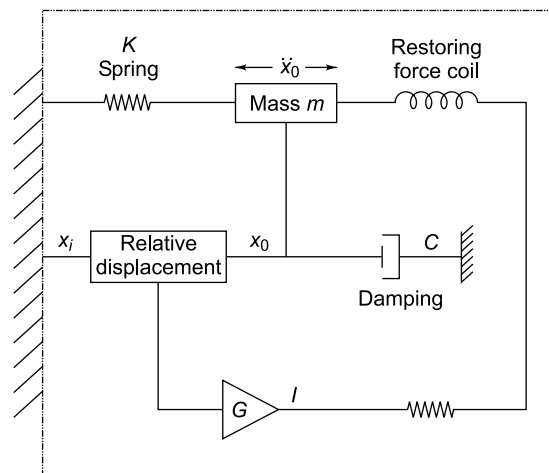
(a)

(b)

FIGURE 3.22   Servo accelerometer: (a) schematic arrangement, (b) photograph (Courtesy of http://www.sherbornesensors.com).

equilibrium. Hence the current required to prevent the mass from moving is a measure of the input acceleration (on the frame) along the direction in which the mass is free to move.

With respect to Fig. 3.22(a) we have,

$$m\ddot{x}_0 = Kd + FI + C\dot{d} \tag{3.21}$$

Here K is the linear spring rate, F is the force produced on the mass per unit current, C is the viscous damping coefficient and I the current. x_i is the displacement of the casing, x_0 is that of the mass and d is the relative displacement given by

$$d = x_i - x_0 \tag{3.22}$$

Therefore,

$$m\ddot{x}_i = Kd + FI + C\dot{d} + m\ddot{d} \tag{3.23}$$

For the condition that $I = Gd$, where G is the gain of the amplifier,

$$m\ddot{x}_i = \frac{m}{G}\ddot{I} + \frac{C}{G}\dot{I} + \left(\frac{K}{G} + F\right)I \tag{3.24}$$

For $K << FG$, i.e. for large values of G, it can be proved that the sensitivity is given by

$$\frac{I}{\ddot{x}_1} = \frac{m}{F} \tag{3.25}$$

Advantages
 (i) It has high accuracy and reasonable frequency response.
 (ii) A flat frequency response up to about 2000 Hz can be obtained in these devices with high sensitivities.
(iii) It can be used for low frequency measurements.

Disadvantages
 (i) Its cost is extremely high as compared to conventional open loop accelerometers.
 (ii) It has limited dynamic range.

Typical specifications
Frequency range	:	0–1000 Hz
Acceleration range	:	±500 g
Sensitivity	:	0.1–10 mV/g
Linearity	:	0.02%

Applications
For vibration measurements of huge structures typically in guidance and navigation systems.

3.8   OTHER VIBRATION TRANSDUCERS

3.8.1   Rotary Variable Differential Transformer (RVDT)/Torsional Vibration Pickup

This is a seismic type of transducer. The principle of operation is similar to that of the LVDT or inductive accelerometer, except that it is used to measure angular displacements. Whereas the LVDT uses a cylindrical iron core, the torsional vibration pickup uses a rotary ferromagnetic core and instead of a linear spring it

uses a torsional spring. The shaft, the torsional vibrations of which are to be measured, is attached to the core and the magnet (which is similar in function to the seismic mass) is free to oscillate. The relative angular motion of the core with respect to the magnet gives an output signal proportional to torsional vibration velocity. Figure 3.23 shows the principle of operation and photograph of this transducer.
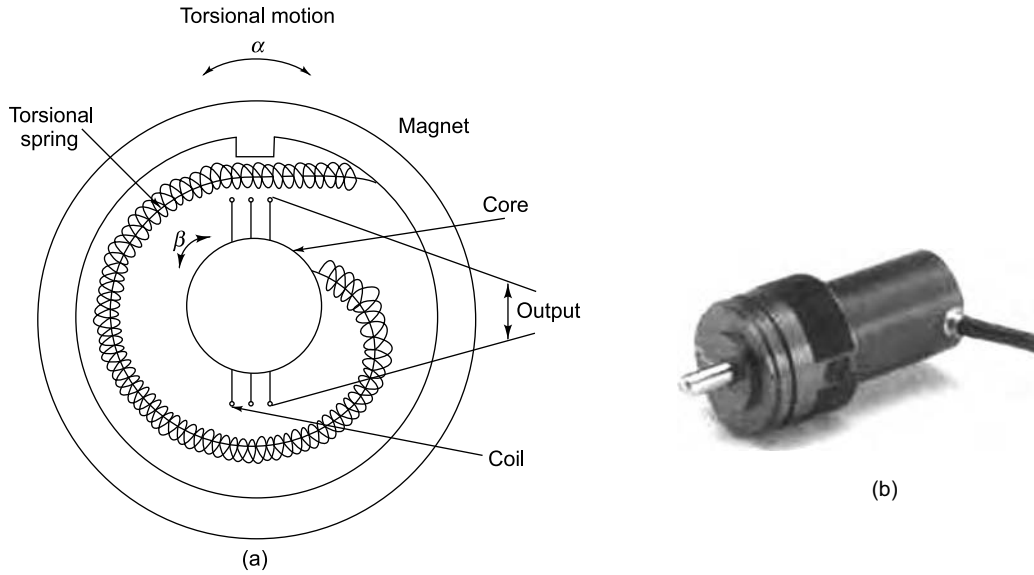
FIGURE 3.23 RVDT: (a) schematic, (b) photograph (Courtesy of http://www.paralab.pt)

Let the angular motions of the magnet and the core be α and β, respectively. Let $\gamma = \alpha - \beta$ be the relative motion between the core and the magnet. The equation of motion of the system shown in Fig. 3.23 may be written as

$$J\ddot{\alpha} = -s(\dot{\alpha} - \dot{\beta}) - q(\alpha - \beta) \tag{3.26}$$

where J, s and q are the mass moment of inertia of the shaft, the torsional damping and the torsional stiffness, respectively. Assuming sinusoidal motion $\beta = B \sin \omega t$ for the torsional motion, we obtain the equation

$$J\ddot{\gamma} + s\dot{\gamma} + q\gamma = J\omega^2 B \sin \omega t \tag{3.27}$$

The steady-state solution is

$$\gamma = \Gamma \sin (\omega t - \phi) \tag{3.28}$$

The frequency response function Γ/B is a complex quantity with magnitude and phase lag ϕ as shown below:

$$\frac{\Gamma}{B} = \frac{\eta^2}{\sqrt{(1-\eta^2)^2 + (2\zeta\eta)^2}} \tag{3.29a}$$

$$\tan\phi = \frac{2\zeta(\omega/\omega_n)}{1-(\omega/\omega_n)^2} \tag{3.29b}$$

where $\eta = \omega/\omega_n$ is the frequency ratio and ζ is the damping ratio.

Equations (3.29a) and (3.29b) are depicted by Fig. 3.4 as a function of η. ζ of the order of 0.65–0.7 is chosen as before.

Advantages

(i) It is relatively low cost.

(ii) It is robust and is capable of working in a wide variety of environments.

(iii) It has a high signal to noise ratio.

(iv) It has negligible hysteresis.

(v) Its theoretical resolution is infinitesimal, being limited only by the resolution of the amplifiers and voltmeters used to process the output signal.

Disadvantages

(i) The fixing and calibration of this torsional transducer requires substantial machinery downtime.

(ii) It is affected by extraneous magnetic fields.

Applications

This transducer is ideal for the following:

(i) Industries such as the automotive, aerospace, petrochemical, marine and power industries, where even small speed variations can lead to poor machine performance, premature failure and human discomfort.

(ii) General torsional vibration measurements and analysis on rotating components in engines, motors, prime movers, pumps, compressors, couplings, shafts and dampers, marine propulsion systems and structural analysis on rotating components.

(iii) Relative torsional vibration measurements during run-up and coast-down, order analysis, dynamic torque-loading influence measurements, shaft speed measurements, etc.

3.8.2 Laser-Based Torsional Vibration Transducer

A laser-based torsional vibration transducer is an easy to use, highly accurate, reliable vibration measuring system. A non-contact torsional vibration transducer using the state-of the-art laser technology may be used for making torsional vibration measurements where it is not feasible to mount a transducer onto a rotating object. Using a dual beam laser transducer, the instantaneous changes in angular velocity, and, by integration, in angular displacement, of a rotating component may be found from the frequency difference of the retro-reflected, Doppler-shifted beams. A schematic of such an instrument is shown in Fig. 3.24(a) and a photograph in Fig. 3.24(b). At the heart of the system is typically a low power (<1.5 mW), Ga–Al–As laser. The laser beam is split into two equal-intensity parallel beams separated by a distance, d given by

$$d = R_A \cos\alpha_A + R_B \cos\alpha_B \tag{3.30}$$

The beams strike the surface of the rotating shaft at points A and B, which have velocities V_A and V_B, respectively. Each beam sees only the velocity in the x-direction.

$$v_A = -V_A \cos\alpha_A - V_x = -\omega R_A \cos\alpha_A - V_x \tag{3.31}$$

$$v_B = V_B \cos \alpha_B - V_x = \omega R_B \cos \alpha_B - V_x$$

and is frequency shifted as shown below.

$$f_A = 2\frac{v_A}{\lambda} \text{ and } f_B = 2\frac{v_B}{\lambda} \tag{3.32}$$
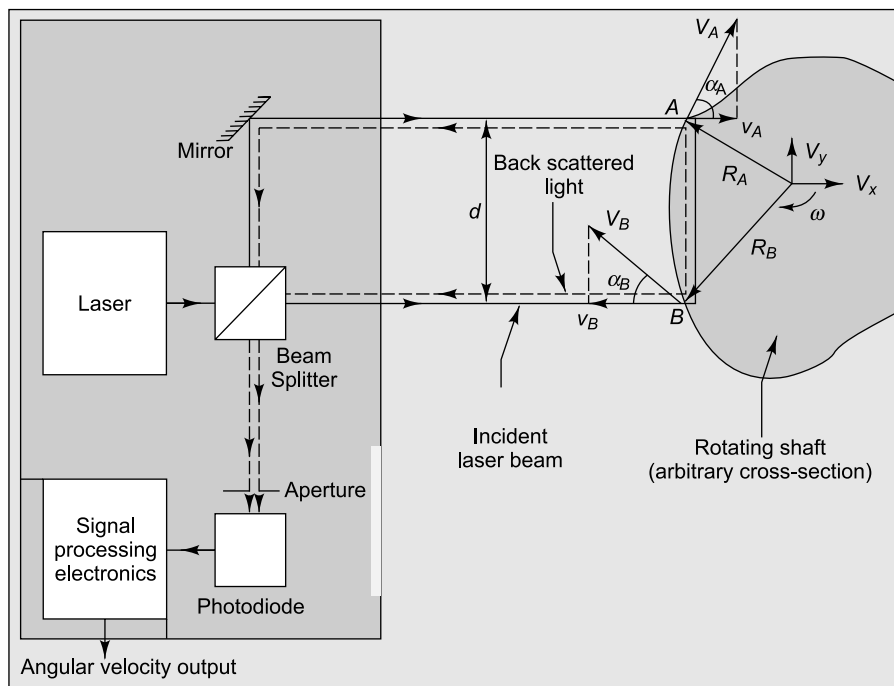
(a)

(b)

FIGURE 3.24 (a) Schematic of laser torsional vibration transducer, (b) photograph of laser torsional vibration transducer (Courtesy of Brüel & Kjær, Denmark)

The reflected beams heterodyne, giving an output current modulated at the beat frequency which is the difference between the frequencies of the Doppler-shifted beams as shown below:

$$f_D = f_B - f_A = \frac{2}{\lambda}(v_B - v_A) = \frac{2\omega d}{\lambda} \qquad (3.33)$$

The equation indicates that the beat frequency is directly proportional to the shaft speed (ω) and is independent of any rigid body motion of the shaft. These equations hold good as long as the plane of the laser beams is perpendicular to the axis of the shaft.

In practice the laser transducer is mounted on its tripod, or held in the hand. A strip of reflective tape is wound around the target. The laser is made to point at the tape and readings are taken or the output is sent to a frequency analyser, together with a tacho signal, to perform order analysis. The optimum measuring distance from the laser transducer to the target is 5–50 cm.

Advantages

 (i) Measurements are independent of target cross-section.
(ii) Remote measurements up to 0.5 m are possible.
(iii) It is simple and quick to use.
(iv) It is compact, lightweight, and portable.
(v) Battery powered, hand-held operation is possible.

Specifications

Angular vibration velocity	:	0.3–10,000 °/s (peak)
Angular vibration displacement	:	0.01–17'
Rotational speed	:	30–7200 rpm
Frequency range	:	0.3–1000 Hz
Laser	:	Ga-Al-As diode producing 780 nm light (Class 3B Laser)
Output power	:	<1.5 mW
Operating distance	:	5–50 cm
Laser spot diameter	:	Less than 1 mm

Applications

The laser-based torsional vibration transducer is useful for all the cases mentioned in Section 3.8.1. Besides it is ideal for non-contact angular vibration velocity and angular vibration displacement measurements.

3.9 SMART SENSORS (PLUG AND PLAY DEVICES) AND TRANSDUCER ELECTRONIC DATA SHEETS (TEDs)

Certain types of transducers are classified as smart sensors. They contain a smart transducer interface module (STIM), which in turn, contains an onboard Electrically Erasable Programmable Read-Only Memory (EEPROM) IC, called Transducer Electronic Data Sheet (TEDS), using which the sensor can identify and describe itself to the network and/or to a smart signal conditioning device, thereby facilitating automatic system configuration. TEDS, or Transducer Electronic Data Sheet, is a set of electronic data in a standardized format defined within the IEEE 1451 standard giving these sensors the name 'Plug and Play' sensors. The TEDS electronically stores information regarding the transducer's characteristics and parameters such as type of device, manufacturer, model number, serial number, calibration date, sensitivity,

reference frequency calibration coefficients and frequency response for a transducer, in addition to interface details such as bridge type, excitation, etc. in terms of a table or an algorithm. It also provides read and write functions for accessing the TEDS and transducer data. The specification also defines the data set, that is, the number of samples acquired for one command, which varies from 0 to 65,535 samples per set. The standard does not specify requirements for signal conditioning, signal conversion, or how applications can use the TEDS data. However, the signal conditioners and other interface hardware used with these smart sensors must provide for an option or include circuits for TEDS communications under the direction of a software module specifically intended for this purpose. Not all sensor manufacturers provide this feature, but those that do are encouraged to follow IEEE 1451.

TEDS capability was originally intended for piezoelectric sensors such as accelerometers and pressure sensors, but it now includes all common analogue sensors and actuators, such as MEMS accelerometers, pressure transducers and temperature sensors with two-wire (analogue) and mixed-mode (analogue and digital) input/output. For two-wire analogue sensors and actuators (called Class 1), the output signal is generally coupled to the signal conditioner or driver, while the TEDS data are enabled and read out with a DC bias voltage applied to the same two wires. The transducers with mixed-mode capability (called Class 2) also communicate digitally with the TEDS memory. The TEDS file may be contained onboard the sensor in the EEPROM, or off board in a reserved file in the data acquisition system. There are services offering conversion of non-TEDS sensors to TEDS "Plug and Play" sensors using a connector, recalibration and a programmed EEPROM. Figure 3.25(a) shows a smart TEDS sensor configuration and Fig. 3.25(b) a photograph of a plug and play transducer.

(a)

(b)
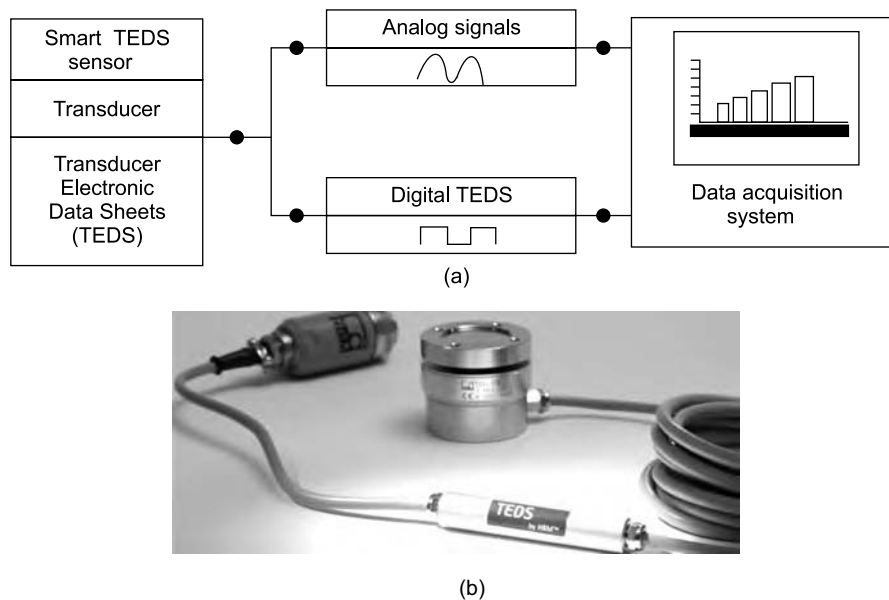
Figure 3.25 TEDS transducer: (a) configuration of a smart TEDS sensor, (b) photograph (Courtesy of HBM, Germany)

Parameters typically included on a TEDS chip

Manufacturer ID	Model number	Version
Serial number	Transducer	Electrical signal type

Minimum physical value   Maximum physical value
Minimum electrical value   Maximum electrical value
Mapping method   Bridge type
Impedance of each bridge element   Response time
Excitation level (nominal)   Excitation level (minimum)   Excitation level (maximum)
Calibration date   Calibration initials   Calibration period (days)
Measurement location

Advantages of TEDS
  (i) Simplified troubleshooting.
 (ii) Reduced safety risks: It is no longer necessary to 'climb on machines' for connection verification.
(iii) Reduced costs for set-up and tear-down.
(iv) No recalibration is needed when replacing sensors.
 (v) Capability of system of recalibrating itself.
(vi) Integration with wireless transceivers.
(vii) Feasibility of adding intelligence to sensors.

3.10   Comparison of Transducers

We have seen a wide range of transducers working on various principles. Non-contact transducers are preferred for many of the vibration measurement and modal analysis applications since they cause minimal mass loading of the structure. There exists a wide range of transducers well-suited for non-contact vibration response measurement as has been discussed in the earlier sections. Table 3.2 gives an assessment of transducers for non-intrusive measurement.

There are also many transducers for contact measurement which can be used for general machinery diagnostics and non-critical applications. Table 3.3 gives a comparison of the performance characteristics of contact type of pickups.

TABLE 3.2   Summary of non-intrusive vibration transducers

Characteristics	Intensity modulated fibre optics	Proximity probes			Laser techniques	
		Electromagnetic	Eddy current	Capacitive	Holography	LDV
Frequency range	DC – 70 kHz	DC – 2 kHz	DC –100s of kHz	DC – 50 kHz	DC – 10 MHz	<1 Hz – 1 MHz
Working or stand-off range	0.05–1.3 mm	0.05–2 mm	0.5–60 mm	$10^{-7} - 1000$ mm	–	10 mm – 18 m
Resolution (at optimum stand-off)	2.5×10^{-4} mm	0.1 mm/s	10^{-5} mm	10^{-4} mm	1 mm	10^{-4} mm
Non-linearity	±1%	±5%	±0.05%	±1%	–	±1%
Minimum target footprint	0.15 mm	2 mm	2 mm	1 mm	N/A	As small as 0.05 mm
Affected by environment	Dust, shock and vibration	Stray magnetic fields	Stray magnetic fields	Grease or dirt, humidity	Dust, shock and vibration	Dust, shock and vibration

TABLE **3.3**   Summary of contact transducers

	LVDT	Electrodynamic velocity pickup	Inductive accelerometer	Piezoelectric accelerometer	MEMS accelerometer	Servo-accelerometer
Frequency range	DC – 200 Hz	5 Hz – 10 kHz	DC – 5 kHz	5 Hz – 50 kHz	DC – 1000 Hz	DC – 2 kHz
Working range	± 50 mm	± 50 mm/s	± 1000 m/s^2	± 5000 g	± 100 g	± 500 g
Resolution	10^{-3} mm	10^{-2} mm/s	10^{-3} g	10^{-3} g	10^{-2} g	10^{-3} g
Non-linearity	± 0.05%	± 1%	± 0.05%	± 1%	± 0.05%	± 0.02%
Minimum target footprint	0.2 mm	2 mm	5 mm	2 mm	2 mm	1 cm
Affected by environment	Stray magnetic fields	Stray magnetic fields	Stray magnetic fields	Temperature	–	Dust

FURTHER READINGS

1. Buzdugan, Gh., Mihãilescu, E. and Rades M., *Vibration Measurement* (*Mechanics: Dynamical Systems*), Springer, Netherland, 1986 .
2. Christian, L., *Mechanical Vibration & Shock*, Taylor & Francis, North America, 2002.
3. Collacott, R.A., *Vibration Monitoring and Diagnosis*, George Godwin Ltd., London, 1979.
4. Cook, R.O. and Hamm, C.W., *Fibre optic lever displacement transducer, Applied Optics*, Vol. **18** (19)–3230–3241, Oct., 1979.
5. Doebelin, E.O., *Measurement Systems: Application and Design*, McGraw Hill Professional, New York, 2004.
6. Ewins. D.J. and Inman, D.J., *Structural Dynamics @ 2000: Current Status and Future Directions*, Research Studies Press Ltd., England, 2001.
7. Gatti, P. and Ferrari, V., *Applied Structural and Mechanical Vibrations—Theory, Methods and Measuring Instrumentation*, (eBook), Taylor & Francis, 1999.
8. Hoogenboom, L., Allen, G.H. and Wang, S., *Theoretical and experimental analysis of a fiber optic proximity probe*, in *Proceedings of the S.P.I.E. Technical Symposium*, East 84, Arlington, Virginia, April 29, Paper #478-25, 1984.
9. McConnell, K.G., *Vibration Testing, Theory and Practice*, John Wiley & Sons, Inc., New York,1995.
10. Murty, D.V.S., *Transducers and Instrumentation*, Prentice-Hall of India Pvt. Ltd., New Delhi, 2004.
11. Nakra, B.C. and Chaudhary, K.K., *Instrumentation Measurement and Analysis*, Tata McGraw Hill, New Delhi, 1985, 2004.
12. Patton, M.E. and Trethewey, M.W., *A technique for non-intrusive modal analysis of very lightweight structures*, in *Proceedings of the 5th International Modal Analysis Conference*, April 6–9, Imperial College of Science, London, England, 1987.
13. Patton, M.E. and Trethewey, M.W., A survey and assessment of nonintrusive-modal-testing techniques for ultralightweight structures, *International Journal of Analytical and Experimental Modal Analysis*, 2(4):163–173; Oct., 1987.

14. Piersol A.G., *Harris' Shock and Vibration Handbook*, McGraw-Hill Professional, (eBook), 2002.
15. Rangan, C.S., Sarma, G.R. and Mani, V.S.V., *Instrumentation Devices and Systems*, Tata McGraw Hill Publishing Company Ltd., N. Delhi, 1983.
16. Rao, B.K.N., *The Handbook of Condition Monitoring*, Elsevier, Oxford, 1996.
17. Redl, W.A., *Noise and Vibration Measurement: Prediction and Mitigation*, American Society of Civil Engineers, New York, 1985.
18. Reeves, C.W., *The Vibration Monitoring Handbook*, Coxmoor, Oxford, 1999.
19. Reza Moheimani, S.O. and Fleming, A.J., *Piezoelectric Transducers for Vibration Control and Damping* (*Advances in Industrial Control*), Springer, USA, Germany, 2006.
20. Smith, J.D., *Vibration Measurement and Analysis*, Butterworth, London, 1989.
21. Thomson, W.T. and Dahleh, M.D., *Theory of Vibration with Applications*, Pearson Education India, India, 2003.
22. Wilson, J.S., *Sensor Technology Handbook*, Elsevier, Newnes, Oxford, 2005.
23. Wowk, V.P.E. *Machinery Vibration: Measurements and Analysis*, McGraw-Hill, New York, 1991.
24. http://www.bksv.com.
25. http://www.dytran.com.
26. http://www.ge-energy.com.
27. http://www.geocities.com/styrene007/sensors/SEMINAR.html.
28. http://www.hbm.com.
29. http://www.hydraulicspneumatics.com.
30. http://www.micro-epsilon.co.uk.
31. http://www.mtiinstruments.com.
32. http://www.mtssensor.de.
33. http://www.optonor.no.
34. http://www.paralab.pt.
35. http://www.pcb.com.
36. http://www.sensorland.com.
37. http://www.sensorsportal.com.
38. http://sensors-transducers.globalspec.com.
39. http://www.sherbornesensors.com.
40. http://www.star-instruments.com.
41. http://news.thomasnet.com.
42. http://www.vibcons.com.

Vibration Excitation Techniques

Chapter

4

4.1 INTRODUCTION TO VIBRATION EXCITATION

It is often required to study the dynamic behaviour and resonant characteristics of structures and machines. Whether we are concerned with printed circuit boards or suspension bridges, high-speed printer mechanisms or satellite launchers, good dynamic response is fundamental to sustained and satisfactory operation. Vibration exciters are used for studying the effects of vibration on a structure or for evaluating the physical properties of materials or structures. They may be used in any of the following phases: product development, simulation, production, dynamic testing or life evaluation.

There is a large variety of excitation techniques which can be used for obtaining structural response in vibration/modal testing. The choice of a particular excitation technique depends on the size and fixity condition of the structure, the excitation signal to be imparted, the required frequency range, the sensing mechanism and the data analysis procedure. Some procedures require specific excitation characteristics and the induced vibration level may dictate the use of specific transducers. Basically excitation techniques can be divided into two categories, contacting (intrusive) and non-contacting (non-intrusive). The former type involves connection of the exciter to the test structure throughout the test. The range of exciters available includes shakers of hydraulic and electrodynamic type and unbalance mass exciters. There are also mechanical shakers of the displacement type, as well as inertial exciters and electromagnetic shakers, especially for rotating structures. While the electrodynamic or electrohydraulic shakers are usually used to directly transmit force to the structures through pushrods or stingers, bulk-fixing of items to these shakers is also possible. In this case, the test item totally sits on the shaker table and this type of testing is more applicable to fatigue life and resonance behaviour studies of the system and/subsystems fixed to the base of the exciter. Fully active actuators, such as electromagnetic shakers, piezoceramics and films, magnetostrictive and electrohydraulic devices, can also be used to excite a structure. The measurement of forces transmitted to the structures by these shakers also requires careful consideration.

In the non-contacting type of excitation technique, the exciter is not in contact with the test structure during the test or is in contact for a very short duration only, as in the case of an impact hammer. Most of the non-contact structural excitation methods currently used are typically extensions of traditional modal excitation procedures such as operational, step relaxation, hammer impact and acoustic methods; they do not possess the ability to apply a force input with a controlled time history to a point on the structure—this is a drawback.

The choice of excitation can make the difference between a good measurement and a bad one. Selection of excitation should be approached from both the type of excitation techniques available and best suited for the

application (for example, shaker excitation) and the type of function desired (for example, sinusoidal or random noise) because they are interrelated. The excitation technique is the physical mechanism used to provide the force input to the structure and the excitation function is the mathematical signal used for the input.

4.2 VIBRATION EXCITERS OF THE CONTACT TYPE

4.2.1 Electrodynamic Vibration Shaker

This is the most common type of exciter and is extensively used for the study of structural dynamics, which is essential for understanding and evaluating the performance of any engineering product. The electrodynamic exciter is essentially the electrodynamic transducer used in the reverse direction. The popularity of these shakers in vibration testing is due to the fact that they offer a wide range of forces— from less than 5 to 20,000 N and a broad range of frequencies—from near 0 to over 10 kHz. These exciters may be given sinusoidal, random or transient signals for excitation and are invariably driven by an audio frequency power amplifier.

In principle the electrodynamic exciter is similar to a common loudspeaker, where the motion is produced by a current passing through a coil in a magnetic field. But in general, shaker coils use heavier conductors than speakers and are more robust to allow flow of heavier currents. Hence shaker coil resistance is generally lower than that of common speakers. Just as in the case of an electrodynamic pickup, the springs are initially compressed, forcing the vibrating object to exert additional force to lower the pin into free motion in the magnetic path. Figure 4.1(a) shows the principle of operation of the shaker and Fig. 4.1(b) the photograph of a typical exciter. At the heart of the shaker is the coil of wire, suspended in a radial magnetic field. When a current is passed through this coil, a longitudinal force which resists the motion of the conductor is produced. This is proportional to the magnetic flux passing through the coil, to the current flowing through it and to the number of coil turns within the flux. The force provided by the shaker is given by the equation

$$F = Bil \tag{4.1}$$

where F is the force produced in N, B is the magnetic flux density in Wb/m^2, i is the current through the coil in A and l is the length of conductor cutting the flux in m.
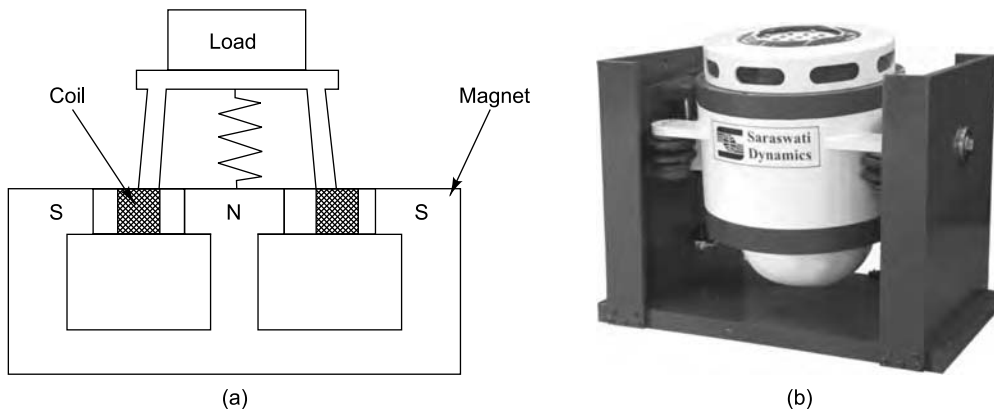
FIGURE 4.1 Electrodynamic shaker: (a) schematic drawing, (b) photograph (Courtesy of Saraswati Dynamics Private Ltd., India).

This force is transmitted through the spring to a table structure to which the test specimen may be fixed. The radial magnetic field is accomplished by building a magnetic circuit of permeable iron or steel around an axially polarized cylindrical magnet. An inner pole piece transmits flux from one end of the magnet, say the north face to the outer south face, resulting in a radial flux field through the gap between these ferrous parts. In order to ensure that the coil moves axially, but is restrained from all other motions, it must be accurately centred in the narrow gap between the inner and outer poles. The coil is wound around the outer diameter of a stiff thin-walled tube, the coil form.

The larger electrodynamic shakers, though very similar in operation, use DC-excited electromagnets. These coils may require cooling jackets or blowers, especially those in exciters with high force ratings. Care should be taken to ensure that the shaker (i) is used within the specified maximum table acceleration velocity and stroke limits, (ii) delivers no more than the rated maximum force, (iii) consumes root mean square (RMS) current less than the continuous duty rating and (iv) avoids eccentricity of loading beyond the rated moment restraint of the machine. Generally, the larger the shaker, the greater is the force which is available for exciting the structure, but this is at the expense of a reduced upper frequency limit. The acceleration level which can be obtained depends on the maximum current and the load. At low frequencies, however, this acceleration level is low. Resonances in the moving element decide the upper frequency limit.

The shaker is usually attached to the structure using a stinger (long, stiff and slender rod) so that the shaker will impart a force to the structure only along the axis of the stinger, which is the axis of force measurement. A load cell is generally connected between the structure and the stinger to measure the excitation force.

Advantages

(i) These exciters are relatively inexpensive, easy to control, simple to interface and quite linear in their behaviour (if used within their specified force and motion limits).

(ii) The frequency and amplitude of excitation can be controlled independent of each other using a function generator, allowing a lot of operational flexibility and this is found to be useful especially around resonances.

Disadvantages

(i) The performance of the exciter is affected by stray extraneous magnetic fields.

(ii) It has a low specific force, i.e. output force to exciter weight ratio.

Typical specifications

Maximum dynamic force	:	1700 N
Maximum acceleration (peak)	:	140 g
Maximum velocity (peak)	:	120 cm/s
Maximum displacement (peak to peak)	:	18 mm
Mass of moving element	:	1 kg
Suspension stiffness	:	300 N/cm
Frequency range of operation	:	5–5000 Hz
Input current	:	0.1 A/N

Applications

This shaker is very popular for modal analysis (finding natural frequencies, mode shapes and damping ratios), product evaluation, stress screening, squeak and rattle testing, testing of aircraft, automobiles, etc.

4.2.1.1 Indirect measurement of excitation force   It is often required to measure the force imparted by the exciter for computation of the frequency response function of a vibrating structure. The force may be computed from the expression given in Eq. (4.1) where it is assumed that the magnetic flux density B, the length of conductor in the flux field l and the number of coil turns N in the field are constant throughout the full range of travel of the exciter. Thus instead of using a force transducer to measure the force input into a structure, the force can be calculated from the measured current fed to the coil. As mentioned earlier, a flexible push rod or stinger links the exciter to the structure directly. Though by definition, the force generated in the coil is directly proportional to the coil current and the two are in phase, it may be inappropriate to deduce the force by measuring the current passing through the shaker coil, since this does not measure the force applied to the structure itself, but to the assembly of the structure and the shaker's drive. The moving components of the exciter and the stinger have a dynamic modifying effect on the force. It may appear that the difference between the force applied to the structure and that generated within the shaker is small, but it should be remembered that in reality, especially around resonant frequencies, the applied excitation force becomes very small, even without alteration of the settings on the power amplifier or signal generator, nevertheless producing a large response and making the frequency response function vulnerable to noise or distortion in this frequency range.

The true force applied to the structure is actually the difference between the force generated in the shaker and the inertia force required to move the stinger and the shaker table, and is in fact, much smaller than either. Suppose that the mass of the moving part of the shaker and its connection to the structure is m_s, the force generated in the shaker is f_s and the force actually applied to the structure, i.e. the force to be measured is f_p, then we have the relation $f_s - f_p = m\ddot{x}$ where $\ddot{x}$ is the acceleration of the structure. With lightweight structures the errors can be significant, and corrections in the direction in which force is being sensed and in perpendicular directions will be required. Thus it is clear that there is a need for direct measurement of the force applied to a structure in order to obtain an accurate and reliable indication of the force and therefore the frequency response function properties.

4.2.1.2   Direct approach to force measurement   In this method, a load cell/force transducer is attached between the structure and the stinger to measure the excitation force. Any force transducer is sensitive to (i) bending moments and transverse movements which result in translational motion of the structure in any direction other than along the driven axis, (ii) rotational motion of the test structure at the location of the force transducer and (iii) rotational or translational motion of the exciter on its stand or suspension system. The function of the stinger is to minimize these effects. A conventional test set-up using a force transducer can be seen in Fig. 4.2(a). The force transducer is attached with its 'base' towards the test structure and the 'top' is connected to the shaker via the stinger. The 'live-side' (base-side) mass of the force transducer is kept small in an effort to minimize the modification to the structure. However, what is actually minimized is the mass modification to the structure only in the direction in which force is being sensed. The full mass of the force transducer modifies the structure in directions perpendicular to the direction of force measurement. Thus there is a large difference between the apparent mass of the force transducer, as seen by the structure, in the direction in which force is being sensed and perpendicular to it. For a typical commercial force transducer, the 'base-side' mass is only of the order of say, 3 gm, and this is the mass seen by the structure in the direction of measurement. The total transducer mass may however be 20 gm, say, and this is the mass seen by the test structure in directions perpendicular to the one in which the force is being sensed.

If the force transducer were to be mounted on the structure the 'wrong' way up (Fig. 4.2b), then there would be less difference in the apparent loading on the structure in the measurement direction and those directions perpendicular to it. However, the 'live-side' mass as seen by the structure in the sensing direction would now be 17 gm (for the force transducer mentioned earlier), causing mass loading. The mass seen by the structure in directions perpendicular to this would remain unchanged at 20 gm, i.e. the total mass of the force transducer.

A third way of fixing the force transducer is to mount it directly on the shaker platform, with the 'base' arranged towards the structure and the 'top' towards the shaker as shown in Fig. 4.2(c), with the stinger connecting the force transducer to the structure. The difference between the mass added in the sensing and perpendicular directions is thus reduced. Only the small 'base-side' mass of the transducer plus the stinger modifies the structure in the perpendicular directions. Quite often, since the stinger mass is smaller than the total mass of the force transducer, the overall modifying effect will be less than for the conventional arrangement. Besides, there will be a much smaller difference in the effect of modification between the measurement direction and perpendicular directions. An advantage of this arrangement is that the force transducer cable (which can be relatively stiff and massive with indeterminate effect) may also have less mass loading on a lightweight structure.
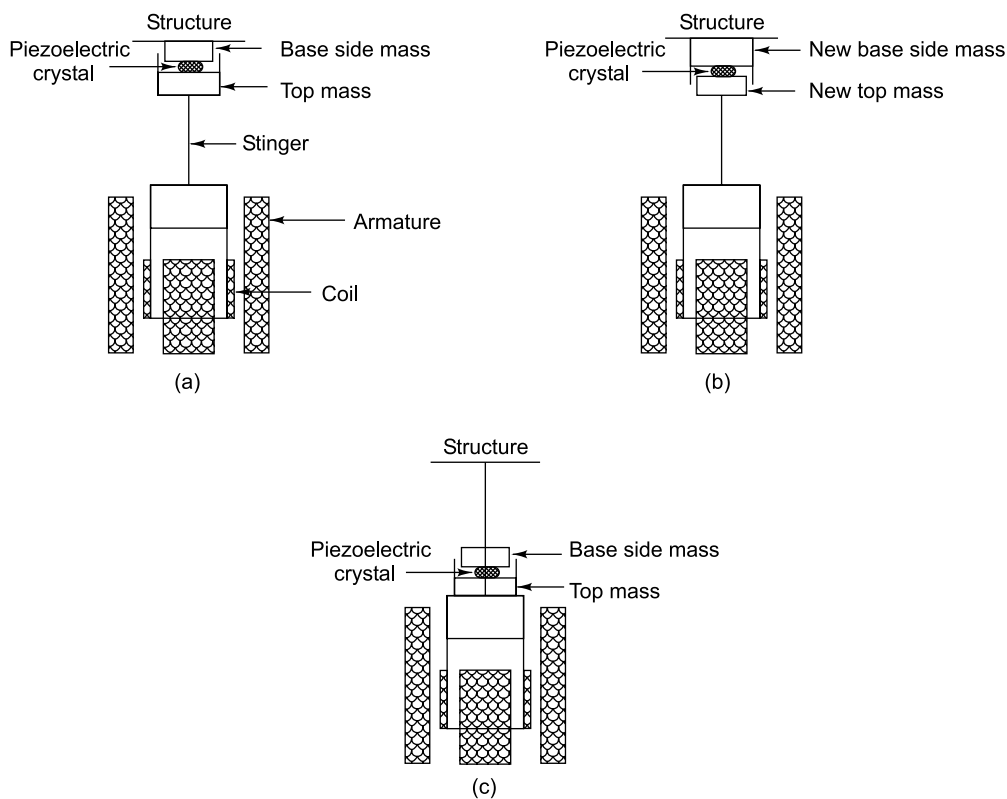
FIGURE 4.2   Force transducer mounting: (a) conventional mounting, (b) wrong mounting, (c) alternate correct mounting arrangement

Only the translational effects of the force transducer have been considered so far. It is to be noted that the force transducer will modify the structure in all six DOF and the effects of the rotational degrees of freedom should be considered for accurate measurements. In the case of multipoint excitation with relative phases between the different inputs, the phase shift between the drive signal and the force must also be taken into account. This can be done using the transfer function between each drive signal and its associated force. The simpler but more laborious approach is to control the phase iteratively.

4.2.1.3   *Slip tables*

As test objects become heavier and larger, electrodynamic generators can no longer be used. Slip table systems help simulate practical conditions and enable structures under test to be excited in their working position. A linear guide system guarantees high stiffness of the slip plate and minimizes cross-vibration if the test setup is not symmetrical. Systems with horizontal slip table and vertically guided load bearing platform (Fig. 4.3) make it possible to test extremely large and heavy loads in x-, y- and z-directions.

Guided oil film slip tables typically consist of a precision ground and lapped natural granite block with a magnesium plate that slides on an oil film to provide a highly damped horizontal slip surface and simultaneously eliminates overturning moments. The slip table generally comes with a

FIGURE 4.3   Photograph of slip table (Courtesy of Saraswati Dynamics Private Ltd., India)

precision granite surface, internal oil reservoir, filter, regulator, a magnesium slip plate and a magnesium drive bar. Depending upon the application, the thickness of the slip plate varies from 40 to 75 mm. Low pressure bearings designed for testing of heavy loads and voluminous objects help restrain pitch and roll moments. An oil film is automatically created through a closed loop system, when the oil pump coupled with the motor is started. Adequate oil supply is provided for table breakaway friction under heavy specimen loads. The oil supply returns to the reservoir through a flexible pipe under gravity and is adjustable for various table loads. The oil used in the table is generally a standard hydraulic fluid. The important factor for operation of this table is viscosity of the oil. When operating at higher temperatures, it may be necessary to use a larger supply or oil of higher viscosity, or both.

Commercially available models come in two versions: (i) stand alone models designed for retrofitting existing vibration generators along with the slip table frame on a base plate. (ii) slip table integrated in a steel frame, together with the vibration generator, allowing quick alignment and precise coupling of the generator to the sliding table. These are available with standard working areas of a maximum of 4 m^2.

Driver bars provide the link between the shaker and slip plate and are generally made of magnesium. Their geometric design enables perfect force transmission while minimizing the moving mass. Pneumatic isolation elements are also available and these make it possible to install slip tables without expensive foundations. Due to the low natural frequency of these isolators (3–5 Hz), a wide test spectrum may be applied. Hydrostatically guided slip tables are also available; these incorporate hydrostatic slide bearings, making it possible to apply the high yaw, roll and pitch moments as they appear with heavy test items or very large loads which may have a high centre of gravity above the slip plate. These tables use high-pressure bearings with a separate hydraulic supply unit.

Typical specifications of electrodynamic shaker with slip table

Frequency range	:	5–2500 Hz
Maximum dynamic force	:	80 kN (peak)
Shock (peak)	:	260 kN
Maximum acceleration	:	100 gm peak (for continuous operation)
Maximum velocity	:	200 cm/s peak (for continuous operation)
Maximum displacement	:	40 mm peak to peak (for continuous operation)

Table for vertical excitation

Diameter	:	430 mm
Mass of moving armature	:	59 kg
Mass of mounting plate	:	8 kg
Suspension stiffness	:	600 N/cm

Table for horizontal excitation (slide table)

Size	:	900 × 900 mm
Mass of moving elements	:	170 kg
Maximum payload	:	7000 kg

4.2.1.4 Vibration/climatic test systems Today industrial equipment are expected to work in all kinds of environment covering a wide range of temperature, vibration and high humidity. Combined vibration/climatic test systems can simulate the effect of ambient conditions on components. This helps identify weak points at an early stage, preventing expensive downtime. There are complete solutions available with software for controlling vibration testing with required test profiles in climatic chambers. Such multitest systems are very useful for quality assurance, research and development. Figure 4.4 shows such a climatic chamber.

FIGURE 4.4 Climatic chamber (Courtesy of Saraswati Dynamics Private Ltd., India)

Specifications

Test space	:	140 × 95 × 115 cm^3
Maximum table size	:	130 × 90 cm^2
Temperature range	:	–70 to +175 °C
Temperature change rate	:	5 °C/min
Temperature control	:	± 1 °C
Humidity range	:	20–95% R.H.
Controllable programmer	:	controls for temperature, humidity, etc.

4.2.2 Modal Thrusters

From the discussion on electrodynamic exciters and force measurement in Sections 4.2.1.1 and 4.2.1.2, it is evident that the measurement of input force in a modal test is very important for successful experimental modal analysis. Modal thrusters/exciters are specifically designed with the idea of obtaining consistent FRF measurements. They are also designed for long-time operation and to ensure the highest possible quality and accuracy with minimum setup time.

A modal exciter is essentially an electrodynamic generator which is supported in a rugged revolving frame permitting excitation in the vertical or horizontal direction (Fig. 4.5).

There are many test structures which require an accurately controlled static preload force in order to take up the slack in bearings, gears or joints and this can easily be set in a modal exciter. These exciters are available in the market with forces from 100 N up to 15 kN. The modal generator is precisely coupled to the test structure using a stinger which faithfully transmits axial forces, and also helps isolate and decouple the force transducer/impedance head from side loads and bending moments. Modal exciters sometimes have built-in electrical stinger pre-tensioning systems, enabling exciter positioning and orientation with complete flexibility. The axial stiffness can also be adjusted electronically in many systems. The exciters come with compatible power amplifiers which can be run in current or voltage mode. Vibration transfer to the floor can be reduced by a swivel-frame which allows a great variety of coupling options. The frame also has vibration isolators as a standard feature so that no extra foundation is needed.

FIGURE 4.5 Modal exciter (Courtesy of Brüel & Kjær, Denmark)

Present state-of-the-art stinger designs are all based on the tension wire concept. A tension wire test setup utilizing a thin metal (piano) wire with zero compression and negligible bending stiffness is arranged in such a way that there is a constant static tension in the wire, upon which the dynamic oscillatory force is superimposed. In practice, the tension wire technology is implemented by using a 'through-hole' design of the modal exciter and exciter fixture, whereby the exciter can be 'slid' along the wire. A complete range of dedicated accessories including force transducers, impedance heads, exciter stands for easy horizontal positioning, as well as a family of stingers is generally provided with these exciters. The generator is cooled by a fan/blower, with cooling air entering through a filter assembly. The blower can also be installed outdoors.

Modal exciters are characterized by:
- High excitation force per unit cost of exciter
- High force to weight ratio due to unique rare earth magnet technology
- High cross-axial stiffness
- Large displacements of the order of 20–100 mm (peak–peak) allowing good low-frequency performance
- Low mass and high rigidity of armature (moving element)
- Low weight of exciter allowing ease of positioning and orientation relative to test object
- Minimum preparation for attachment and pre-tensioning of stinger wire
- Wide frequency range
- Special 'through-hole' magnesium armature for tension wire stinger technology

Typical specifications

Rated force	:	100 N (peak sine)/70 N (RMS random)
Frequency range	:	2–5000 Hz
Maximum travel	:	25 mm
Rated current	:	5 A
Natural frequency	:	6 kHz
Spring stiffness	:	4 N/mm

4.2.3   Mechanical Exciters of the Direct Drive Type

The direct drive mechanical exciters consist of vibrating tables which are guided to have rectilinear motion and are driven by some appropriate mechanisms. They are generally driven by DC or AC motors with variable speed drives, typically from 100 to 6000 rpm (1.67–100 Hz), so as to be able to control the frequency of excitation. The displacement type of exciter is capable of producing excitation at a predetermined frequency with a predetermined magnitude of displacement. It produces a constant displacement irrespective of the speed. The displacement type of exciter consists of a rotating eccentric disc or cam driving a positive linkage mechanism between the base of the exciter and the table. A schematic drawing of two such exciters is shown in Fig. 4.6.
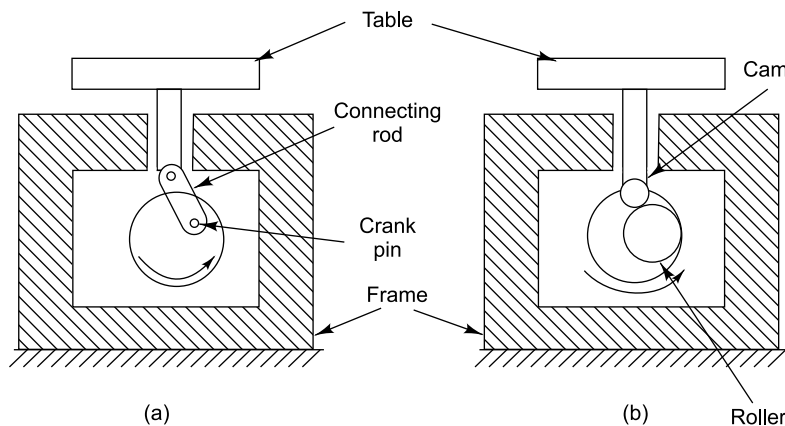
FIGURE 4.6   Direct drive mechanical vibration exciter: (a) eccentric and connecting link mechanism, (b) cam and follower arrangement

The mechanism may be (i) a connecting rod, (ii) a cam and roller follower or (iii) a cam and flat face follower. The amplitude of vibration is dictated purely by the crank length in the first case and by the eccentricity setting of the cam in the second case and is independent of the frequency of operation. The first mechanism will not give rise to pure simple harmonic motion, but will contain higher harmonics also. In the case of the cam mechanism, variable types of motion, dependent on the cam design employed, can be obtained. The flat face follower, on the other hand, ideally produces pure harmonic motion, but with wear over a period of time, performance may degrade. The acceleration obtained on the table is proportional to the square of the speed and hence for the same setting of the eccentric, maximum dynamic force is obtained for the highest speed of operation. With an eccentricity of around 5 mm, a maximum acceleration of 50 g is obtained at a speed of 3000 rpm.

Advantage

It can produce a constant vibration displacement irrespective of frequency of operation.

Disadvantages

It is generally meant for small test loads of the order of a few hundreds of newtons.

Typical specifications

Amplitude   :   ± 0.1–1 mm vertical and horizontal
Frequency range   :   0–200 Hz
Test mass   :   8 kg

Applications

These exciters are used for evaluating the physical properties of structures, for studying the effects of vibration on components and for fatigue tests on specimens with masses of the order of 20–40 kg.

4.2.4   Reaction Type Exciters

Reaction types of exciters produce excitation with a predetermined magnitude of dynamic force at any chosen frequency. A reaction exciter develops its excitation force through an inertial loading which is caused by the acceleration of a reaction mass. In operation the exciter's base is attached to the structure under test. A double mass reaction exciter consists of two rotating unbalanced masses of equal eccentricity rotating in opposite directions and phased such that the unbalanced forces add up to a sinusoidal force acting in a plane at right angles to the two axes of rotation. The total force produced by the exciter is

$$F = 2\, me\omega^2 \sin \omega t \tag{4.2}$$

where m is the mass of one rotor, e is the eccentricity and ω is the angular speed of the rotor in rad/s.
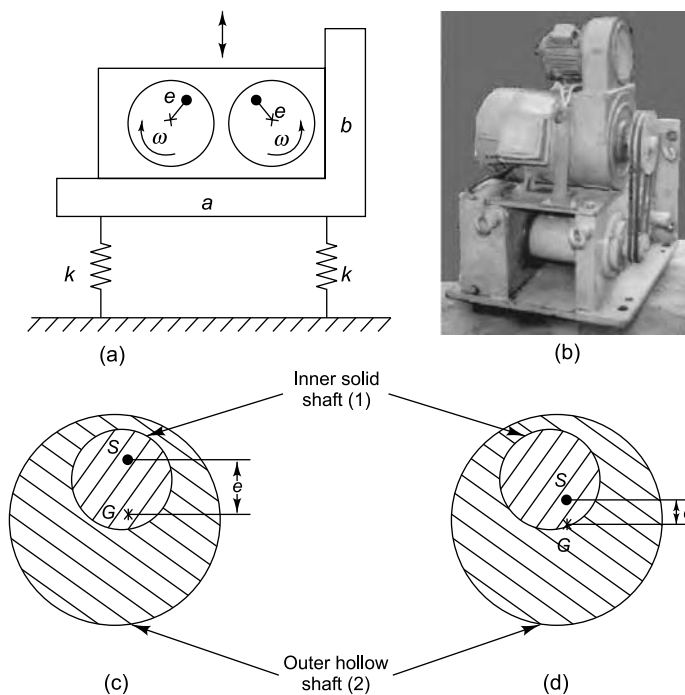
FIGURE 4.7   Double mass mechanical exciter: (a) schematic of exciter, (b) photograph (Courtesy of IIT Madras), (c) maximum eccentricity setting, (d) minimum eccentricity setting. S—centre of rotation, G—centre of mass, e—eccentricity.

Thus, it is seen that the reaction type of exciter produces an exciting force proportional to the square of the speed for a given setting. Figure 4.7(a) shows schematically the principle of operation of this exciter. For the arrangement shown in this figure, vibratory forces cancel out in the horizontal direction. Vibration in the vertical direction is facilitated by fixing the exciter base a and the test structure on a test platform resting on vertical springs as shown in Fig. 4.7. By mounting the exciter on face b on a

platform resting on rollers, excitation in the horizontal direction can be obtained. The photograph of a heavy duty reaction exciter is shown in Fig. 4.7(b). Figures 4.7(c) and (d) indicate the procedure for changing the eccentricity in steps using the principle of double eccentric. The eccentric mass consists of an outer hollow shaft 2 and an inner solid shaft 1. When the unit is stationary, eccentric 2 can be rotated relative to eccentric 1 as shown in Fig. 4.7 to get different settings of eccentricity.

Advantages

  (i) This exciter has a very high force to exciter weight ratio and is capable of generating large amounts of forces at higher frequencies.

 (ii) The magnitude of excitation force can be calculated fairly accurately knowing m and e and no further measurement of force may be required.

Disadvantages

  (i) This exciter produces very small forces at low frequencies because of speed-squared dependence.

 (ii) There is relatively little flexibility or control over the generated forces as compared to other exciters like electrodynamic or electrohydraulic since the magnitude of force is dictated by eccentricity e, which can be changed only in discrete steps and not continuously.

(iii) In practice it is not possible to change the magnitude of force at a particular speed while the exciter is running, since this involves changing e and this can be done only when the exciter is stationary.

(iv) Since the exciter is driven by a motor, its upper frequency limit of operation is typically restricted to around 100 Hz.

 (v) This exciter can be used only as a reaction exciter.

Applications

  (i) These exciters can be used to determine the natural frequencies of heavy structures like foundations, bridges, masts, towers, etc. which require large forces and the shaker itself offers negligible weight compared to the test item.

 (ii) It can also be used to perform simulated seismic tests on specimens which are required to undergo seismic qualification tests like bus ducts, transformers, panels and other parts of power plants which are to be erected in earthquake prone zones.

Typical specifications

Speed range: 100–4200 rpm
Drive: 2.5 kW DC motor with speed control unit
Exciting force: 30,000 N at 3000 rpm

4.2.5 Inertial Systems

An interesting development in exciters is that of the smaller inertial shakers (reaction mass shakers) used extensively in modal analysis. Figure 4.8 shows the photograph of an inertial exciter.

   Inertial exciters can efficiently apply dynamic forces to large structures and have found their applications in common production, aerospace and civil engineering,

FIGURE 4.8 Inertial exciter (Courtesy of http://www.xcitesystems.com)

and shipbuilding. The inertial mass modal exciter allows the testing of structures where backup fixturing is not available or possible. The inertial forces are generated by accelerating an inertial mass with a closed loop hydraulic excitation system. In some systems excitation is made by permanent magnets, and a special spring system provides optimal guidance so that the full body mass of the exciter can impact the structure. The controller helps to maintain a constant force over the required frequency range by continuously altering the drive signal to maintain the required dynamic force level. Inertial systems are available in the force range from 100 to 5000 N. These systems can be bolted directly to the structure and aligned at any angle within 360° on the test structure. The generators have an excellent lateral and axial stiffness. They are fan-cooled, with cooling air entering through a filter assembly. They often come with load cells and LVDTs for force and displacement measurements.

Advantages

The ease with which this exciter may be positioned and oriented on the test structure makes it highly suitable for field applications.

Typical specifications

Dynamic force	:	4450 N
Stroke	:	25 mm
Moving mass	:	25 kg
Weight	:	61 kg
Frequency range	:	30–300 Hz

Applications

The inertial shaker is ideal for extracting the response of a ship hull or car body which requires fixing the exciter in an arbitrary fashion; this type of excitation requires the least test preparation. This exciter is useful in testing electrical transmission towers, turbine rotors, stators and bearing housings, generator armatures, stators and windings, diesel engines and motor-generator sets. The ease with which these systems can be fixed allows for testing ship bulkheads, propeller shafts and propellers, as well as structure borne noise isolation systems in submarines.

4.2.6 Electrohydraulic/Hydraulic Exciters

Electrohydraulic exciters are highly suitable for applications requiring large amounts of forces, typically of the order of tonnes. These exciters can provide relatively long strokes, allowing excitation of structures at large amplitudes, a feature not available in electrodynamic exciters of comparable sizes. The power amplification required for the generation of large forces is achieved through the use of hydraulics. These exciters are, in general, more complex and expensive than electrodynamic and mechanical exciters. One main advantage of these exciters is their ability to apply simultaneously a static load with a superimposed dynamic load. This is useful while testing structures with a major static load, which may change the dynamic properties, or even geometry of the structure. They exhibit high force capability, high force per unit weight and volume, good mechanical stiffness and high dynamic response. There are also commercially available units which produce triaxial forces. These exciters are suitable for low frequency operation; at higher frequencies, the stroke gets reduced considerably. The frequency range of these exciters is typically 0–100 Hz, the exact frequency range depending on a number of factors such as pump and flow rate capacity. The upper usable frequency range of the electrohydraulic exciters available today is considerably higher than what it was earlier; very few operate above 1 kHz though.

The hydraulic vibration machine, the schematic of which is shown in Fig. 4.9(a), transforms power in the form of high pressure fluid flow to the reciprocating motion of the table attached to the stem of an actuator. The major components of this exciter are the system controller, the servo valve, the hydraulic pump (power pack), a four-way control valve, the hydraulic actuator and the test table. The power pack typically consists of a high-performance hydraulic pump with appropriate pressure relief valves, oil coolers, etc. and produces the required high pressure of the oil. The four-way control valve regulates the flow of oil to either side of the working actuator and is in many systems actuated by a double eccentric and connecting rod mechanism. The double eccentric is driven by a variable speed motor.

In the figure, the four-way valve is shown in its mid-position blocking both high pressure line *P* and return low pressure line *R*. The position of the servo valve controls the oil flow from the pump's high pressure line *P* to the actuator piston and from the piston low pressure side to the pump's reservoir line *R*. If the servo valve position causes the spool to be displaced to the right, high pressure oil enters the right side of the actuator piston and the low pressure oil is drained to the reservoir on the left hand side of the piston. This causes the piston to move to the left so as to develop a force on the test table. As there is no feedback arrangement in this setup, the motion of the valve spool should be symmetric with reference to the mid-position closing the control ports. Hence a reciprocating motion which can be approximated to simple harmonic motion is provided by the double eccentric and connecting rod mechanism, by making the connecting rod sufficiently longer than the crank.

In some simpler designs of the piston-type hydraulic actuator shown in Fig. 4.9(b), one side of the piston contains a spring. The vibration of the test table is produced by allowing alternating flows of the high pressure hydraulic fluid by means of the control valve into the actuator cylinder. The actuator has a test table attached to its piston rod. Generally there is provision for clamping the cylinder so that the table vibrates either in a horizontal or vertical plane. The stem transmits the motion of the piston to a valve. By regulating the amount of oil supplied to or drained from the actuator cylinder, the valve can be positioned between fully open and fully closed conditions.

Figure 4.9(c) shows a block diagram of a typical closed loop hydraulic vibration exciter. The test item is fixed on the table driven by the hydraulic actuator. The valve driver which is an electrohydraulic device is driven by the error signal between the required output motion and the input force and/or motion of the hydraulic actuator, which is used as a feedback signal. This error signal actuates the larger servo valve, which in turn controls the amount of oil that is forced into the hydraulic actuator to obtain the desired motion.

The flow rate performance of such a system is frequency dependent, the high frequency behaviour being essentially controlled by the servovalve, hydraulic actuators and test system dynamics. Important specifications for electrohydraulic and hydraulic valve actuators include actuation time and hydraulic fluid supply pressure range. Devices that have linear motion valves vary in terms of valve stem stroke length and actuator force. For both electrohydraulic and hydraulic valve actuators, acting type is an additional specification. With single-acting devices, fluid pressure actuates the valve in one direction, while a compressed spring actuates the valve in the other. With double-acting devices, fluid pressure actuates the valve in both directions. Most of these exciters come with the capability of sine, random, sine on random, random on random, or resonant search and dwell testing features.
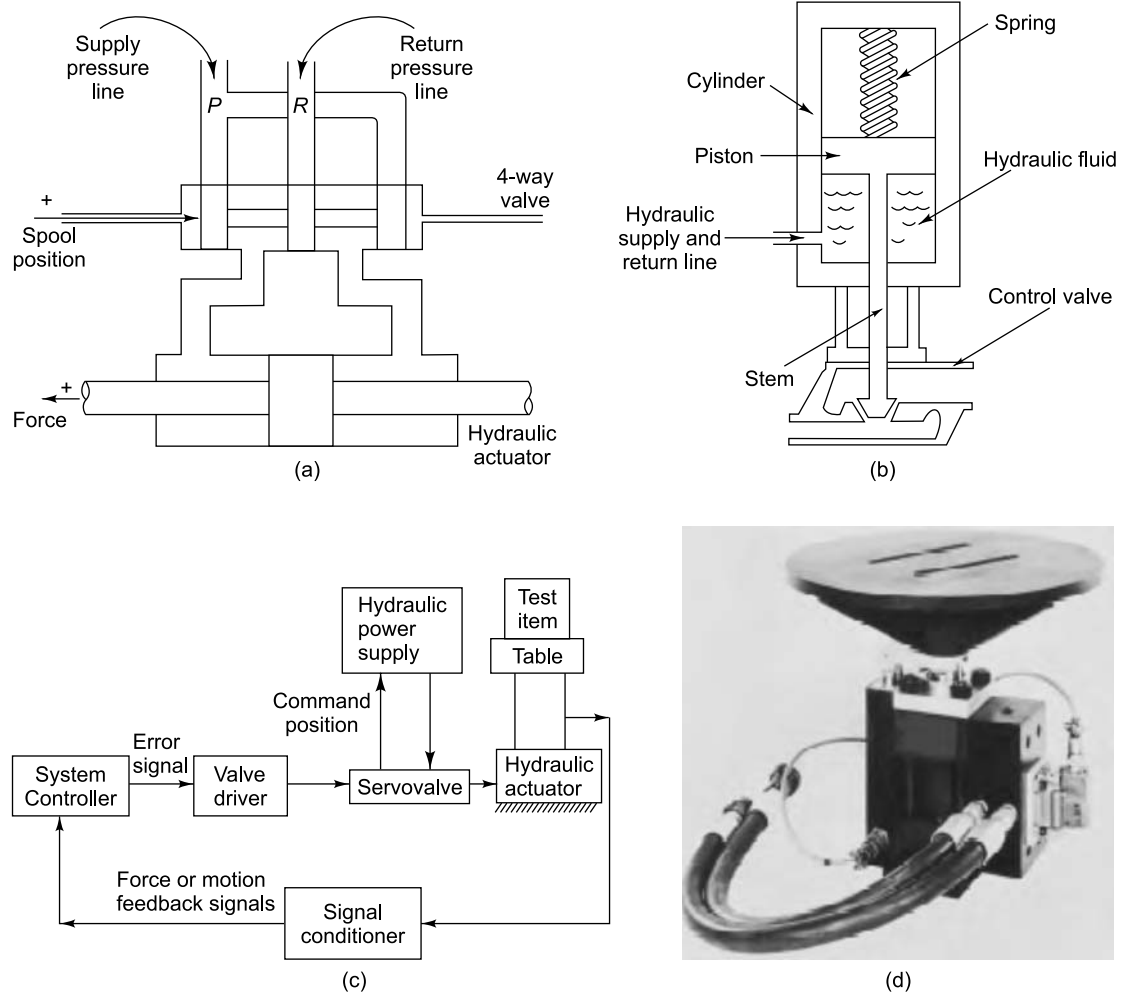
FIGURE 4.9   Hydraulic shaker: (a) schematic, (b) piston type hydraulic actuator, (c) closed loop hydraulic vibration exciter, (d) photograph (Courtesy of http://www.mbdynamics.com)

Advantages

  (i) They provide relatively long strokes, allowing excitation at large amplitudes.

 (ii) They can apply simultaneously a static load with a superimposed dynamic load.

(iii) They have high force capability, high force per unit weight and volume, good mechanical stiffness and high dynamic response.

(iv) Triaxial forces can be produced by these exciters.

 (v) They are highly suitable for low frequency operation.

Disadvantages

These exciters are in general, more complex and expensive than electrodynamic and mechanical exciters.

Typical specifications

Maximum stroke (peak to peak)   :   50 mm

Maximum dynamic force (peak)   :   15 kN

Frequency range	:	0–800 Hz
Oil-hydraulic pump system	:	
Maximum flow capacity	:	180 L/min
Maximum pressure	:	25 MPa (250 bar)

Applications

These exciters were originally developed for materials testing applications. Present-day applications however include fatigue testing, fluid structure interaction studies and earthquake simulation studies. They find wide use in the aeronautical and automobile industry, where they are placed under the wheels of a vehicle to simulate the condition of a road or runway to test the suspension system of automobiles or aircrafts or to conduct seat vibration tests on a human being.

4.2.7 Lead Zirconate Titanate (PZT) Actuators

Devices such as electromagnetic shakers, piezoceramics and films, magnetostrictive and electrohydraulic devices work as fully active actuators supplying mechanical power to a system. They generate secondary response in a linear mechanical system, which interferes with the original response of the system caused by the primary disturbance and reduces overall response. Semi-active actuators behave basically as passive elements, capable of storing or dissipating energy. These actuators are adaptive since their passive mechanical properties can be adjusted by applying control signals. Electrorheological fluids, magnetostrictive actuators, electrostatic devices or shape memory alloys can be used in the fabrication of these semi-active actuators. Amongst these different options, piezoelectric transducers offer several distinctive advantages in active control applications because they are lightweight, can be of any shape and can even be microsized and are easily driven by a voltage. These actuators are flat, thin piezoceramic devices bonded to the surface of a structure. When used as strain actuators, they create linear motion and extend with applied voltage. When used as bimorph actuators, they bend with applied voltage while clamped firmly at one end. They have a wide frequency response, are capable of delivering considerable amount of forces with very little power input and are well suited for high precision applications. They can be directly embedded or bonded to the structure.

The best known and most commonly used piezoceramic is lead zirconate titanate (PZT). Many of the newer devices do not have the problems traditionally associated with using piezoelectric materials, i.e. no wires are required to be soldered. They come in a protective skin which provides electrical insulation and defence against humidity and harsh contaminants with pre-attached electrical leads, producing a highly reliable component. The new piezoelectrics are thus easy to integrate into volume manufactured products. Figure 4.10 shows a photograph of a PZT patch.
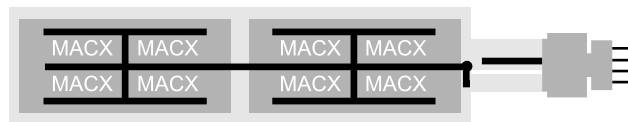
FIGURE 4.10   Photograph of PZT patch (Courtesy of http//www.mide.com)

When piezoelectric materials are subjected to stress, a charge proportional to the applied stress is obtained and hence gives rise to an electric field. Likewise, when these materials are exposed to an electric field, the intermolecular distances change. This gives rise to a strain field that is proportional to

the strength and polarity of the electric field. These two effects are represented in tensor form as

$$S = s^E T + dE$$

$$D = dT + \varepsilon^T E \tag{4.3}$$

where S is the time-invarianthe compliance with unit 1/Pa, T is the stress in Pa, d is the piezoelectric coefficient in C/N, E is the electric field in V/m, D is the dielectric displacement in C/m^2 and ε is the permittivity in C/Vm. The superscript E on compliance matrix s^E indicates that the compliance data was measured under at least a constant and preferably a zero electric field. Likewise superscript T on the permittivity matrix ε^T means that the permittivity data was measured under at least a constant, and preferably a zero stress field.

Advantages
  (i)   They are light and extremely small and are well suited for high precision applications.
  (ii) They are easily driven by a voltage and can deliver large amounts of forces with very little power input.
  (iii) They have a wide frequency response.
  (iv) Besides, they can be used as both sensors and actuators.

Disadvantages
  (i)   They are consumables, in the sense that once bonded to a particular location, they cannot be removed and reused.
  (ii) They are not as rugged as other exciters.
  (iii) The force obtained, as compared to electrodynamic or mechanical exciters is very small.

Typical specifications of actuator patches

Application type	:	Strain actuator
Device size (cm)	:	$10 \times 2.5 \times 0.08$
Device mass (gm)	:	10
Active elements	:	2 stacks of 2 piezos
Piezo wafer size (cm)	:	$4.6 \times 2.1 \times 0.025$
Device capacitance (μF)	:	0.26
Full-scale voltage range (V)	:	± 200
Power output (W)	:	20

4.2.8   Summary of Intrusive Excitation Techniques

Table 4.1 summarizes the features of various contact type of exciters in terms of force ratings, frequency ranges, applications, etc. These are indicative values only and will vary even within the same class of exciters.

TABLE 4.1   Summary of intrusive exciters

Type	Force rating	Frequency range	Application	Nature of force signal	Maximum acceleration (g)
Electrodynamic shaker	2 kN	DC–5 kHz	Product testing/resonance search/fatigue testing/ modal testing	Sine/random/ sine sweep/shock	150

<div align="right">(Contd)</div>

Table 4.1 (*Contd*)

Type	Force rating	Frequency range	Application	Nature of force signal	Maximum acceleration (g)
Electrodynamic modal exciter	0.25 kN	DC–2 kHz	Modal testing	Sine/random/ sweep burst	150
Mechanical direct drive exciter	0.2 kN	5–100 Hz	Fatigue testing/calibration of transducers	variable speed	70
Unbalance Mass exciters	20 kN	1–100 Hz	Modal testing of large structures/seismic testing	Sinusoidal, variable speed	10
Electrohydraulic shakers	250 kN	0.1–100 Hz	Automobile testing, human body response, seismic simulation	Sine/random	5
Piezoactuator	10s of N	0–5 kHz	Active vibration control	Sine/random/ sine sweep	0.1

4.3   NON-INTRUSIVE (NON-CONTACT) EXCITATION TECHNIQUES

Frequency response measurements of medium and heavyweight structures are possible through the use of either intrusive or non-intrusive (non-contact) forms of excitation and vibration response transducers, since they do not alter system dynamics significantly by causing mass loading effects for the extraction of modal parameters. However in the case of rotating as well as lightweight delicate mechanical structures, the use of intrusive excitation techniques, as well as contact type of transducers, causes considerable mass loading so as to significantly alter the system dynamics, preventing accurate modal parameter extraction. Similarly, the structural impedance matching and mass loading effects between a modal shaker and a lightweight test structure may also alter the dynamics. The major requirements of an ideal non-intrusive structural excitation system are as follows:

- The system should apply a force of sufficient magnitude to excite all the structural modes
- Direct force measurement should be possible to obtain calibrated input/output data, and hence FRFs and modal parameters
- It should be possible to position the exciter and the transducer at various locations since existing modal analysis procedures require the estimation of input/output data in the form of point FRFs and transfer FRFs (discussed in Section 10.3.2.2)
- It is desirable to have the ability to control the force amplitude and time waveform characteristics to make the system compatible with standard testing procedures
- The exciter must be able to deliver excitation forces in a broad frequency range, being equally effective at very low and high frequencies so that the system is applicable to different types of structures

4.3.1   Near-Non-intrusive Excitation Using Electrodynamic Shaker

Structural excitation with a shaker system is highly dependent on the test setup and is essentially intrusive in nature as we have seen in the earlier sections. For effective shaker excitation of a lightweight structure, the shaker should be attached to the structure in such a way that the dynamics of the structure is not altered. One way of doing this is by mounting the entire test structure directly on the shaker system.

This technique is attractive since it provides a means to use the shaker and its inherent control of the excitation signal. The drawback of this approach arises from the difficulty of separating the dynamics of the exciter and test stand from that of the structure under study. The main body of the shaker must be isolated from the structure to prevent any reaction forces from being transmitted back to the structure and this is done using a stinger as has been discussed. Besides, mounting the entire test stand on a shaker and exciting it to acceptable measurement levels may be expensive/impossible for some applications. Another potential problem that arises with electrodynamic shakers is the impedance mismatch that can exist between the structure and the shaker coil. The electrical impedance of the shaker varies with the amplitude of motion of the coil. As discussed earlier in Section 4.2.1.1, at resonance with a small effective mass, very little force is required to produce a response. This can result in a drop in the force spectrum in the vicinity of the resonance, causing the force measurement to be susceptible to noise. The problem can be corrected by moving to a point with a larger effective mass, by using shakers with coils of different sizes or driving the shaker with a constant current type amplifier.

4.3.2 Operational Excitation/Ambient Testing

For very large structures, such as dams, large bridges and buildings, ambient testing is used. Here it is important not to interfere with the functioning of the structure. For example, in the testing of bridges, interruptions of traffic flow are undesirable. Any ambient testing must be done in the presence of the inherent operating forcing functions. Also, the applied force functions are often immeasurable and of distributed nature. Since the actual force cannot be measured, only response and not FRF data can be got and the point-to-point structural dynamics are often estimated through the use of a reference and a moving transducer. This requires special procedures and prevents the use of reliable forced vibration studies with data of this type.

Operational excitation provides a viable technique for lightweight structures also. Since no external connections are required, external loading effects are eliminated. However, since operational excitation may typically be harmonic or periodic in nature, all the structural modes may not be excited and hence may not be present in the vibration data base. In addition, this method is applicable so long as the test structure does not undergo any gross motion during its work cycle, necessitating delicate positioning of the non-contact transducers.

4.3.3 Step Relaxation

Dynamic testing of large structure such as bridges and windmills, presents a variety of measurement challenges. Foremost among these challenges is a means of supplying an impact with sufficient energy to excite the test structure uniformly. Since a majority of the dynamic characteristics of interest in large structures are of low frequency, a means of restricting the input spectrum of the force is an important consideration. Step relaxation/free decay testing is one of the most common forms of structural excitation in which the system is given an initial condition (displacement, velocity, etc.) and the free decay response of the system is measured. The technique entails preloading the structure at a point by a flexible tendon. When either the tensile load in the tendon or the displacement reaches a specified value, the load is suddenly released. This imparts an impulsive step function to the structure. Typically a load cell is attached to the tendon anchor to measure the force time history. The resulting free decay can be processed using any of the parameter estimation algorithms available for processing frequency response or unit

impulse response measurements. For the dynamic testing of bridges, the step-relaxation technique has been used very successfully with the load provided by hanging a large mass from the structure and then suddenly releasing it. The technique is especially useful for low frequencies.

The technique is easy to use and requires minimal equipment. The difficulties in this technique are related to the tedious time consuming setup procedures required to produce repeatable impulses and the application of the tendon fracture mechanism. Also, because the technique produces an impulsive excitation, it generates a poor signal-to-noise ratio when the time window is much larger than the actual impulsive signal, as is the case with FFT zoom analysis. Finally, a potentially time-consuming trial and error process may be required to obtain an appropriate input, since the technique has little direct control over the excitation levels.

4.3.4   Impact Testing using Modal Hammers

Hammer impact testing is one of the most widely used and popular excitation techniques for modal analysis due to its ease of use and ability to adapt to difficult testing environments. The impact/modal hammer is simply a hammer with an integral force transducer built into its head. This transducer detects the magnitude of force felt by the impactor, which is equal and opposite to that experienced by the structure. It is generally used in conjunction with an accelerometer on the component being tested to obtain FRFs by processing impact force and acceleration response data simultaneously. When the hammer tip impacts the specimen, the latter experiences a force which is that of a half sine pulse. Although the hammer contacts the test structure, the contact time is usually quite short and does not affect the structure's vibration response. The short impulsive force produces desirable broadband excitation. Since the excitation is impulsive, it suffers from the same poor signal-to-noise ratio and poor load control limitations as the step relaxation technique, making it difficult to obtain consistent results. However, its low cost and ease of operation make it an attractive alternative for some applications. This technique can be used with lightweight structures, provided the structure is sufficiently stiff to produce a useful force level. Testing on miniature structures requires a precise positioning mechanism to ensure that the impact is applied consistently at the same point. Impact hammers can be controlled either manually or remotely. That is, the device could be an actual hand-held hammer to be used by manually swinging the impactor or in a cradle arrangement to enhance the repeatability of the impulse as shown in Fig. 4.11.
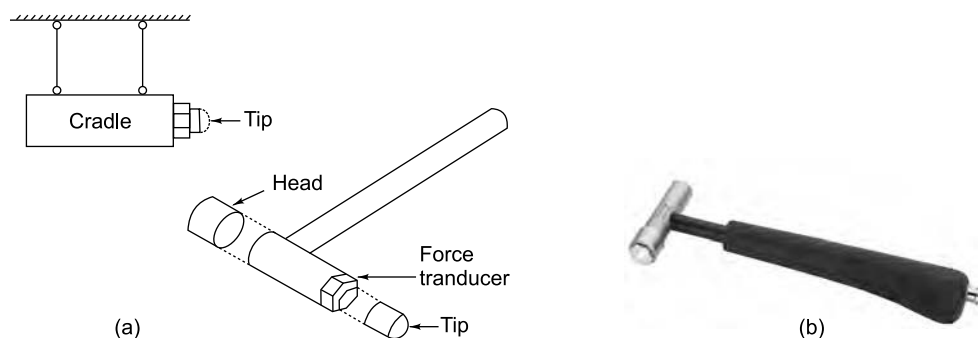
FIGURE 4.11   Impact hammer: (a) different arrangements, (b) photograph (Courtesy of Brüel & Kjær, Denmark)

Selection of the hammer involves choosing the appropriate size and mass of the hammer, which will impart the required force amplitude and spectral content for excitation of the test structure. Each hammer comes with a number of tips and an extender mass which allows further tuning by concentrating more energy at lower frequencies. The striking tips are usually of various materials and hardness (typically soft plastic, rubber, aluminum or steel stud), allowing for different force level ranges (energies for excitation) and different impact durations (measured in milliseconds) which control the frequency range of excitation. In general, harder tips will deform less than softer tips during impact and will therefore have shorter pulse durations. The frequency range effectively excited is controlled by the stiffness (not hardness) of the contacting surfaces and the mass of the impactor head and there is a system resonance at a frequency ω_c equal to $\sqrt{\text{contact stiffness/impactor mass}}$, above which it is difficult to deliver energy into the test structure. Typically the stiffest tips are used to measure response at the highest frequencies. Figure 4.12 illustrates the typical impulse response and force spectra for different hammer heads. Table 4.2 gives typical force values and frequency ranges for the different tips.
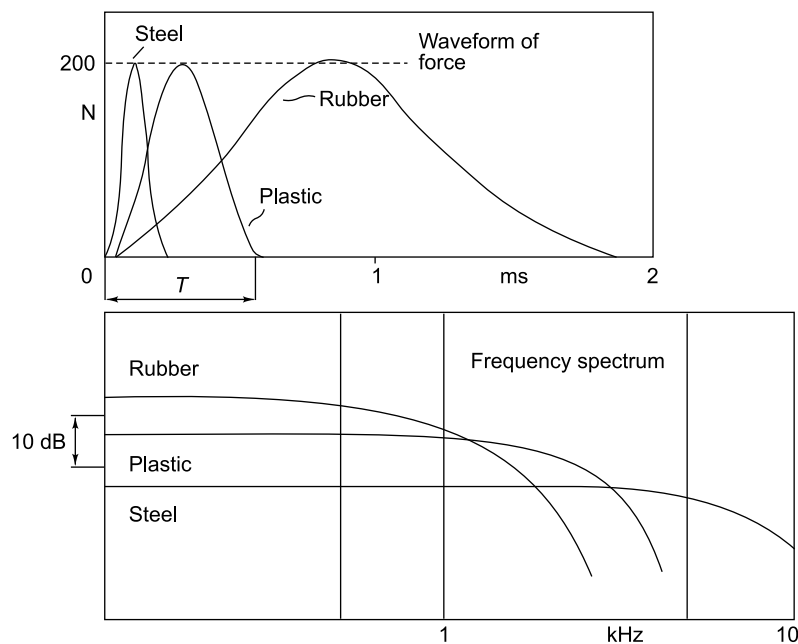
FIGURE 4.12 Typical impact time history and force spectrum

TABLE 4.2 Force and frequency characteristics of different tips

Tip	Obtainable force (N)	Frequency range (Hz)	Duration of impulse (ms)
Rubber	500	~400	1.3–1.8
Plastic	1500	~2000	0.3–0.5
Steel	5000	~6000	0.1–0.2

A number of hammer kits are commercially available; they consist of matched components which are tuned for testing structures of certain size and weight combination. The manufacturer's data sheets

indicate the frequency content of the force impulse that can be obtained using the various tips. Low mass hammers with hard tips provide short duration impacts, while softer tips and mass extenders are used to broaden the impulse. The kits also sometimes include response accelerometers, signal conditioners, cables and all other accessories. PCB's Modally Tuned® Impact Hammer eliminates hammer resonances from the test results, reducing double impacts and resulting in more accurate test results. Modally tuned hammers have been refined, through the selection of their materials of construction, to deliver consistent, accurate results. The frequency spectrum obtained is nearly flat over a broad, yet specified frequency range and is similar to broadband random excitation. Thus a simple impact can excite all the structural resonances within the frequency range. These hammers have found use in such applications as automotive design, bridge health assessment and aerospace vehicle development.

Remote operation impact hammers are also available and these are more like an automated piece of machinery akin to a handheld power tool that is operated by actuating a switch; here the operator can control the force to be applied. These electric hammers minimize errors due to overloads, mis-hits and multiple hits, in addition to improving the consistency of impulse excitations. They also avoid difficulties of a hand controlled hammer due to multiple impacts and poor repeatability of impact conditions, namely the force amplitude and angle of force detection.

Structures often exhibit geometries which make the impact excitation of certain locations difficult or impossible with a standard impact hammer. The impacting force provided by the hammer is not equal to the exciting force transmitted to the structure due to the inertial mass of the punch. A device called the modal punch or instrumented punch or modal sledge hammer consisting of a force transducer attached to the end of a variable length shaft, was designed to eliminate the inertial mass of the shaft by positioning the force transducer to be in contact with the test object. The primary use of a punch impactor is to impact locations that are inaccessible to a conventional impact hammer. It can also be used to precisely locate impacts to minimize the variation in the impact location and to impact in a skewed direction (non-orthogonal to the global coordinates) on the edge of a structure. The punch impactor should be fitted with a hard tip. A tip of metal or hard plastic tip is recommended. The punch impactor should be calibrated in the configuration in which it is to be tested since the sensitivity is dependent on the tip and specimen combination. It can be swung manually or used in a cradle arrangement to enhance the repeatability of the impulse.

4.3.4.1 *Problems in impact testing*

Impact testing has its own disadvantages, limitations and potential problems like double hits, local deformation of a structure, skewed inputs, etc. It also poses problems when used on non-linear structures.

Double hits

Multiple impacts can happen if the surface is not impacted sharply or if the amplitude of vibration is large enough to cause the rebounding surface to contact the hammer tip before the hammer rebounds away from the surface. This condition is most common with very lightly damped systems, in which case a light hammer should be used to reduce the possibility of multiple impacts. Multiple impacts should be avoided whenever possible, because they can produce substantial measurement errors, especially in FRF measurements.

Local deformation of structure

Impacting very flexible areas of a structure, such as a panel, can cause large, local deformation which effectively lengthens the duration of the impact, thereby reducing the usable frequency range of the

input spectrum. In this case, it is not possible to impart energy at the higher frequencies, leading to bad estimates of the frequency response function. In addition, care must be taken not to dent the surface being impacted.

Use on non-linear structures

An impact force is a poor input for exciting a system because it is impulsive and non-deterministic, both of which are often undesirable for measuring the frequency response of a non-linear system when modal parameters are to be estimated. The high peak levels can overdrive the system and exaggerate its non-linear response. To reduce the effects of non-linearities which are inherent to all physical structures, it is desirable to apply a series of impacts to the test structure and take the average of the responses over a large number of records. Because of inconsistencies in the duration and force of the blows, the level of energy in impact excitation applied with a conventional hammer can be difficult to control. To ensure repeatability in averaged impact measurement, an electric hammer may be used.

Skewed inputs and immeasurable DOFs

Obtaining impacts that are consistent in magnitude and alignment is an important concern. Consistent alignment refers to impacting at the same point on the surface and normal to the surface, for each average at an impact location giving rise to FRF measurements showing accurate reciprocity between the input and output DOFs.

Advantages

(i) Low cost and ease of operation make it an attractive technique for many applications.
(ii) No special fixtures are required even in difficult testing environments.
(iii) Self windowing potential for minimal signal processing errors makes it attractive.

Typical specifications	For medium structures	For very heavy structures
Frequency range	8 kHz	0.5 kHz
Maximum force output	2200 N	22,000 N
Sensitivity	2 mV/N	0.2 mV/N
Resonant frequency	30 kHz	3 kHz
Hammer mass	150 gm	5 kg
Tip diameter	0.5 cm	8 cm
Handle length	20 cm	100 cm

4.3.5 Acoustic Excitation

Acoustic excitation of a structure is typically used for qualification tests on space vehicles and requires placing the structure in an artificially generated sound field. These vehicles and their accessories are subjected to high noise levels of the order of 140–150 deciBel (dB) transmitted through the surrounding air during the launch phase. This acoustic loading is particularly critical for lightweight structures with large surfaces such as solar panels. It is preferable to have these structures tested in reverberant chambers with a diffuse sound field of comparable power distribution over the frequency range generated under actual conditions. Acoustic excitation could be generated either by a single loudspeaker or an array of loudspeakers or by using an electropneumatic transducer (Section 12.3.4). The sound wave from the sound source is coupled to the object under test by means of an acoustic horn which is typically made of titanium or steel and is coupled to the object through a softer buffer material that serves, both to transmit the sound and to prevent direct contact between the horn and the structure under test. Although

in some cases the horn is directly coupled to the part with no intermediate buffer, this might produce damage to the structure at the point of contact. It is possible to input a wide variety of excitation signals (random, swept sine, pseudo-random, etc.) which can greatly enhance the quality of the acquired data. The most important feature of this technique is that it is truly a non-intrusive excitation technique. The major drawback with this technique lies in the fact that it is a distributed and not point excitation technique. This fact, combined with the problem of obtaining calibrated structural force data makes it difficult to apply structural forced response studies to modal models developed with acoustic excitation.

4.3.6   Electomagnetic Exciter

The electromagnetic exciter is a non-contact exciter and is extremely suitable for applications, where exciter and transducer loading are of particular importance. This exciter is also useful for transmitting forces to rotating objects, enabling analysis and control of vibrations of rotating machines and calculation of stiffness and damping of bearings. The exciter is also useful for quantitative evaluation of the dynamic response of impellers, circular saws, rotor discs, etc. This exciter essentially works in the reverse direction as the non-contacting electromagnetic pickup discussed in Section 3.6.1. Figure 4.13(a) shows a generalized block diagram of the test setup for electromagnetic excitation and Fig. 4.13(b) a photograph of the exciter.

(a)

(b)
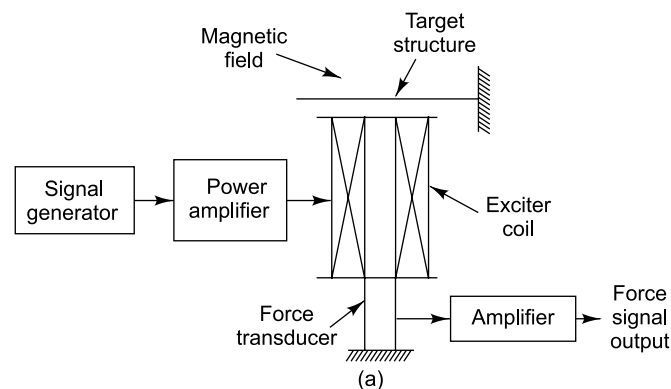
FIGURE 4.13   Electromagnetic exciter: (a) schematic, (b) photograph (Courtesy of IIT Madras)

The current through the coil produces electrical energy which is stored in the electromagnetic field. This energy is transferred in a relatively complex way to the target structure. This process involves the interaction of the electrical inputs with not only the electrical parameters of the exciter coil, but also with the properties of the exciter coil, core material and the target itself. The driving source input current produces a magnetic field in and around the exciter coil, which interacts with the ferromagnetic target structure to produce a mechanical force input for structural excitation. For non-ferromagnetic materials, it is necessary to attach a small ferromagnetic target or magnet to the target surface. The induced force from the magnet is given by the equation

$$F = \frac{PA}{2\mu_0} \left[B \sin \omega t \right]^2 = \frac{PA}{2\mu_0} \left[B^2 \left(\frac{1 - \cos 2\omega t}{2} \right) \right] = \frac{PAB^2}{4\mu_0} (1 - \cos 2\omega t) \ N \tag{4.4}$$

where F is the total induced force of attraction at the air gap (N), P is the number of air gaps assumed to have the same length and cross-section, A is the cross-sectional area of magnetic core at air gap (m^2), B is the magnetic flux density through the core cross-section (weber/m^2 or tesla), μ_0 is the permeability of free space (henry/m) and ω is the angular frequency of the exciter driving current (rad/s).

Equation (4.4) shows that the input force from an electromagnetic exciter is a pull-release force with a DC effect, and not a typical push–pull force like that produced by a standard modal shaker. This effectively produces a preload on the target structure, which may not affect the modal parameters for a linear structure, but may cause some problems with the response transducer positioning or with the analysis of non-linear structures.

When the current in a ferromagnetic material is increased, the magnetic flux of the material increases along the idealized magnetization curve as shown in Fig. 4.14. A sinusoidal input current is distorted by the magnetization curve, producing a distorted magnetic field in the exciter, leading to significant harmonics in the force. Thus when a pure sinusoidal excitation is required, the experimenter has to use lower amplitude of input current so as to operate along only a small part of the magnetization curve, is
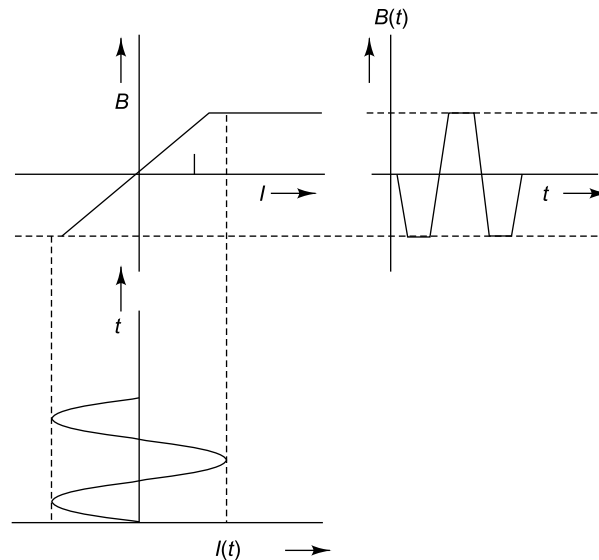
Figure 4.14   Idealized magnetization curve

to generate the magnetization curve thus linearizing the system by reducing the overall distortion caused by the magnetization curve. Another option is to generate the magnetization curve for the exciter system and to use this data to develop the required exciter coil input current signal to produce a pure sinusoidal magnetic field. The design of these systems is relatively straightforward. However, the key component in calibrating electromagnetic shakers is the force transducer, which directly measures the force transmitted to the target structure.

This exciter can be fed with a variety of force input time histories, in addition to the pure sinusoidal forcing function. A force transducer may be mounted between the exciter coil and the ground to measure the dynamic force on the target structure. The magnetic material properties of the core and the target material have a distorting effect on the magnetic field setup due to the current flowing in the exciter coil windings. Typically ferrite is used as the exciter core and target material to overcome the hysteresis phenomenon, resulting in electrical energy losses within the system. This is especially important at the low flux densities required to excite small, light and ultralightweight structures.

Typical specifications

Dynamic force   :   ± 300 N
Force constant   :   100 N/A
Frequency range   :   20–300 Hz
Output impedance   :   2.5 Ω with 20 mH in series
Weight   :   50 N

Advantages

This excitation technique possesses many of the desirable capabilities required for modal testing such as non-contact excitation, easily controllable excitation level and small target area on the structure.

Disadvantages

(i) This exciter is affected by stray magnetic fields.
(ii) Excitation current versus magnetic field is not linear, posing calibration problems.

4.3.7   Assessment of Non-intrusive Excitation Techniques

To examine the relative merits and demerits of the various non-contact excitation techniques discussed, five characteristics of the excitation systems such as frequency range, control of signal or waveform type, signal level or amplitude control, application footprint or size of target area and special test fixture requirements are assessed. All the ratings of these techniques have been made on a qualitative basis in Table 4.3.

The usable frequency range is highly dependent on the test setup. Only the acoustic and electromagnetic systems inherently possess the ability to impart an excitation over a specific frequency range independent of the experimental setup, since they are fully of the non-contact type. In the case of hammer excitation, the frequency range depends on the contact stiffness which in turn depends on the tip. As far as the ability to impart a specific excitation signal/waveform/forcing function to the test structure is concerned, the shaker and electromagnetic systems have the best ability. The acoustic approach may be limited by the ability of the structure to respond to the applied acoustic field in a specific fashion. The remaining approaches have no control over the waveform imparted to the structure. This may be a serious limitation for multiple-input multiple-output (MIMO) testing, where specific forcing functions may be required for excitation.

Considering the ability of each technique to control the forcing function level/amplitude imparted to the structure, the shaker and electromagnetic systems have excellent control over the input levels. The

acoustic system has a reasonable degree of control, while the other techniques have very little control and hence repeatability. As far as target footprint area is concerned, the shaker, step relaxation, impact and electromagnetic systems are capable of imparting essentially point excitation to the structure. This feature is important since most modal parameter extraction procedures require the estimation of frequency response functions at distinct points on a structure. The operational excitation technique has the disadvantage that the forcing function is not measurable and is often distributed over the entire test system. Acoustic excitation is also distributed over the entire structure.

Regarding specific laboratory procedures to be implemented to utilize the excitation procedures, the shaker approach may require significant amount of special hardware to be used, while the operational technique requires no special testing hardware. The step relaxation technique may require some hardware for repeated application of the impulse. The impact technique necessitates a positioning scheme for precise alignment of the hammer for multiple impacts. The acoustic system requires very little special hardware to be implemented. The electromagnetic system may require a moderate amount of hardware since positioning and mounting of the exciter are dependent on the test stand.

TABLE 4.3   Summary of non-contact excitation techniques

Excitation technique	Frequency response	Excitation signal control	Signal amplitude control and level	Specimen size	Target area	Special test-fixture requirement
Non-intrusive shaker	Wide; System dependent	Excellent	Excellent; large	Large	Point	Moderate
Operational	System dependent	None	None; large	Very large	Distributed	None
Step relaxation	System dependent	None	Poor; large	Very large	Point	Moderate
Hammer impact	System and tip dependent	None	Poor; very small	Small	Point	A little
Acoustic	Excellent; wide	Good	Good; not very large	Large	Distributed	Moderate
Electromag-netic	Excellent; wide	Excellent	Excellent; not very large	Large	Point	Moderate

A detailed description of the various excitation signals that may be used in vibration/modal testing is given in Section 10.2.5 as it is felt that it will be easier to understand this topic after signal processing concepts have been dealt with.

4.4   SHOCK TESTING

Closely associated with vibration testing is shock testing and hence some description of this is given here. A shock is defined by transmission of kinetic energy into a system with duration very short as compared to the system's natural period. This causes a non-periodic excitation of the body, resulting in

very large magnitudes of displacement. These shocks can result from suddenly applied forces or abrupt changes in motion encountered from mechanical phenomena such as rough transportation and mishandling, or natural phenomena such as air turbulence or earthquake. When a body is dropped and suddenly hits the floor, bringing it to rest, decelerations of the order of, say, 500 g, in a very short duration, of the order of 5 ms, can result. Most structures encounter some form of shock during their life time. This can cause them to degrade in performance, or to even undergo permanent damage. Failures that can occur from shocks during the service life of an electronic component are broken glass/crystals, wires, leads and solder joints, relay chatter, locked computer hard drives, blown fuses, etc. In mechanical equipment, shocks can lead to broken material lodged in valve, seal failure and structural failures such as cracked housings, fatigue cracks and failures, deformation of structural elements, etc. The main sources of shocks may be classified as pyrotechnic excitation, mechanical excitation and excitation from natural phenomena.

(i) *Pyrotechnic excitation*: These result from sources such as explosive bolts or detonating chords.

(ii) *Mechanical excitation*: Phenomena such as collision impact, dropping, gust loading in aircraft or missiles, ballistic impact, aircraft landing, braking, missile/rocket launching, gunfire, high-speed fluid entry, transportation over uneven surfaces, rough terrain give rise to this type of excitation.

(iii) *Natural phenomena*: Phenomena such as earthquakes, wind gusts, air blasts, ocean waves, ice impact, etc. also cause shocks.

4.4.1   Types of Shock Signals Employed for Testing

There are five different types of shocks that are used and are described in the following subsections.

4.4.1.1   Classical shocks   These are the simplest and can be reproduced fairly easily under laboratory conditions. The various shapes of the pulses used for classical shock testing are shown in Fig. 4.15. The half-sine (Haversine) pulse depicted in Fig. 4.15(a) is the most commonly used signal. It resembles a half-sine with rounded ends to make it physically reproducible. Figure 4.15(b) shows the sawtooth shock, which has an increasing slope followed by a sharp drop and resembles a tooth on a saw blade. This shock is also specified in MIL-STD-810F and may be obtained by dropping a shock table on cylindrical lead pellets with conical top. Another classical shock is the triangular pulse as shown in Fig. 4.15(c) which is not used very frequently. It is similar to a sawtooth; however the drop is not quite as severe. The trapezoidal wave depicted in Fig. 4.15(d) is generally considered the most damaging of waveforms; it conservatively estimates product fragility. All these pulses have some drawbacks in that they are arbitrary, non-realistic and classical pulses. They are still used today because they are easy to produce in the laboratory, are fairly repeatable and mathematically tractable.

4.4.1.2   Spectrum shock/shock response spectrum   A shock response spectrum (SRS) depicts the maximum response of an SDOF system to an applied shock input (Fig. 4.16). Complex structures are modelled as a set of parallel, SDOF systems. SRS plots may be expressed in terms of acceleration, velocity or displacement, in terms of absolute response, relative response or as maxi-max curves. The aerospace industry uses absolute acceleration, while for naval and seismic applications, absolute pseudo-velocity and absolute velocity respectively, plotted on tripartite paper are used. The maxi-max condition is the envelope of all conditions through the environment, with specifications given in terms of nominal test levels in *±dB*. An important feature of an SRS is that it does not uniquely define a shock input, since two different shock

pulses can result in the same maximum peak response. The assumption behind its use is that different shock events, which have the same response spectra, also have equal damage potential. The SRS concept has the advantage that it provides a consistent way of comparing different shock time histories.
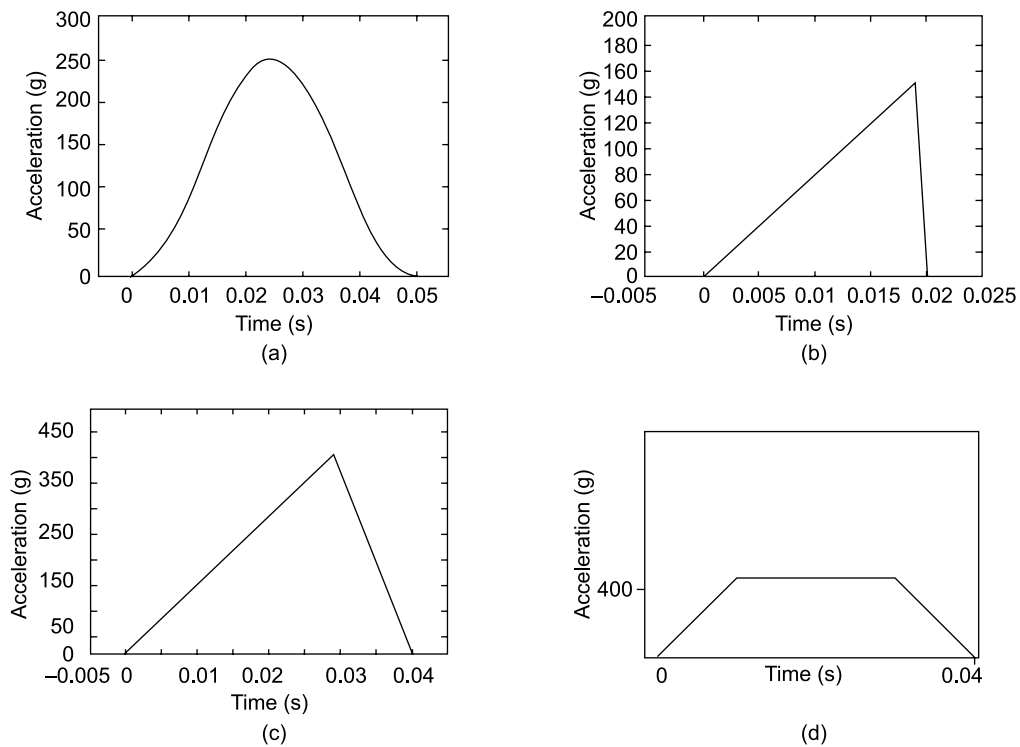
FIGURE 4.15 Classical uni-axial shock pulses: (a) half sine, (b) sawtooth, (c) triangle, (d) trapezoidal
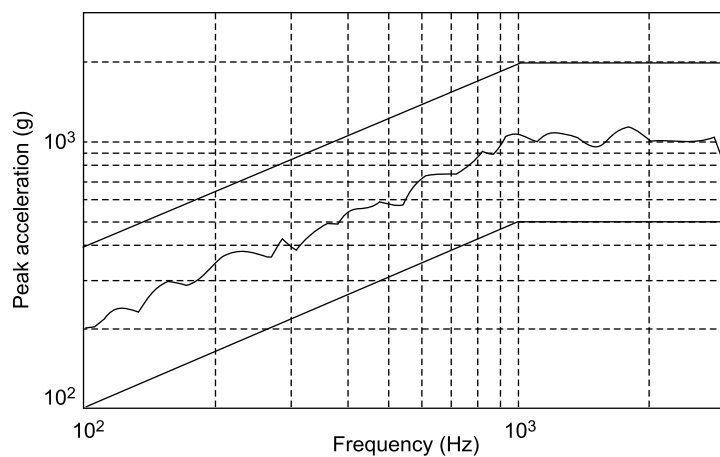
FIGURE 4.16 Shock response spectrum plot

4.4.1.3   *Pyroshock*   Pyroshock is characterized by a high acceleration, short duration shock pulse. The time history of such a shock pulse (Fig. 4.17a) has a zero mean value, and typically decays in a very short duration, less than 20 ms. These shocks may also be measured using the SRS (Fig. 4.17b).
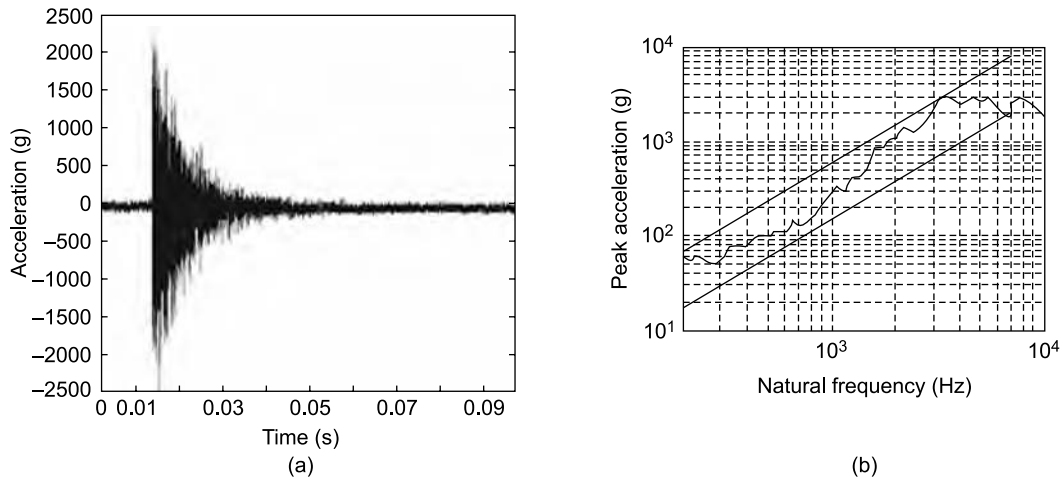
FIGURE 4.17   Pyroshock: (a) time history, (b) response spectrum

4.4.1.4   *Seismic shock*   Seismic shocks (Fig. 4.18) are due to earthquakes and result in low acceleration, high displacement, and long duration time records.
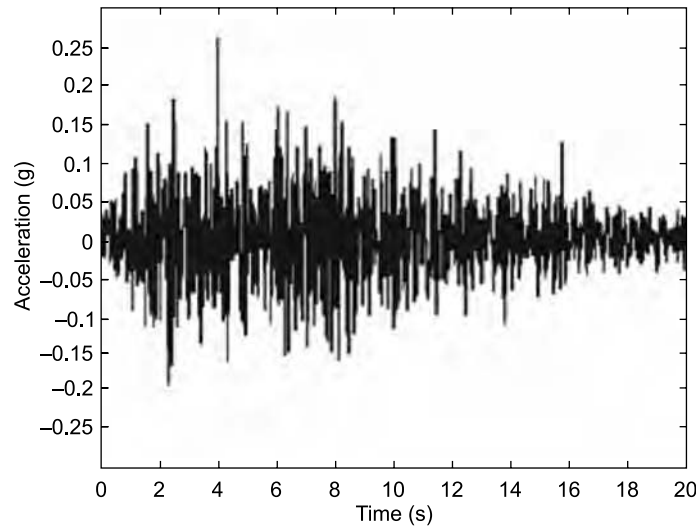
FIGURE 4.18   Seismic shock time history

4.4.1.5   *Air blast*   The shock pulse from an air blast (Fig. 4.19) is characterized by a very sharp rise and long decay (longer than for a pyroshock), typically a couple of seconds.

4.4.2 Shock Testing Methods

Shock testing is commonly performed by any of the following methods (The differences between the methods are in the methodologies used to impart kinetic energy to the system):

- Drop
- Hammer/impact
- Shaker (electrodynamic/hydraulic)
- Pyroshock
- Hopkinson bar
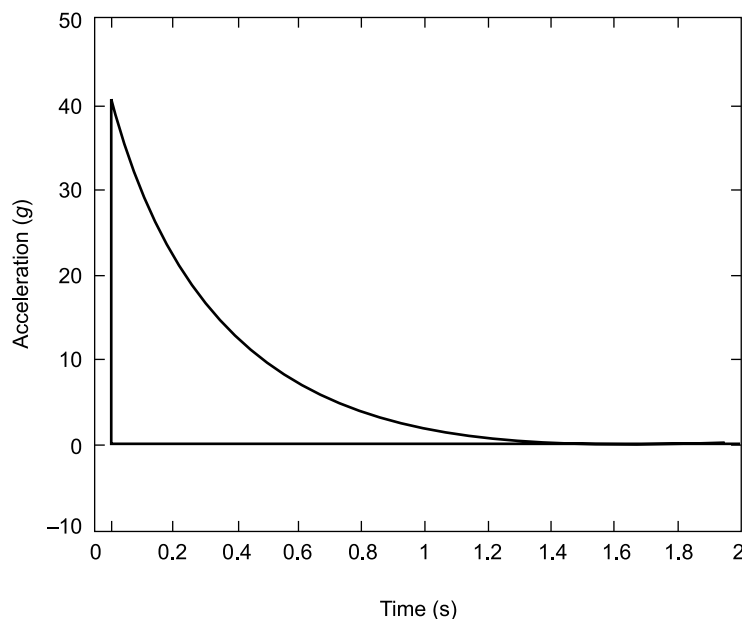
FIGURE 4.19   Shock from air blast

4.4.2.1 Drop shock   The mechanical shock test is a test performed to determine the ability of specimens to withstand shocks. This test requires a machine capable of imparting to the specimen classical shock pulses of amplitude 500–30,000 g (peak), with a pulse width (measured between points on the leading and trailing edges with acceleration 10% of the peak value) ranging from 0.1 to 1 ms. These machines are capable of producing large changes in velocity and displacement. There are shock testing machines of the 'moving carriage' type (Fig. 4.20). The machine itself must be mounted on a sturdy and levelled surface and the device under test should be rigidly fixed to the carriage (with ample protection for leads, if any). The carriage should be hoisted, then released suddenly, causing it to accelerate towards a massive impact surface. The fall may be a free fall or an assisted fall (pulled by a

FIGURE 4.20   Drop shock machine (Courtesy of http://www.mpmtechnologies.com)

bungee cord). The drop height determines the impact velocity. The material of the arresting bed at the bottom determines the shock pulse shape, duration and peak acceleration. Typically, sand is used. Rubber is used for half-sine pulses and pointed lead cylinders for saw tooth pulses. Many procedures such as ASTM D3332 are based on this type of machine. The more sophisticated machines may be equipped with programmers that allow adjustments of the severity and the duration of the arrest by providing different arresting surfaces for the falling carriage. Shock tables are generally expensive and occupy large amounts of laboratory space.

There are also *free fall drop* testers. The principle of operation of such a tester is simple; it pulls the restraining support away from the specimen, allowing it to fall freely. It is a much simpler and less expensive technique for creating shocks than the carriage type machine. This tester gives the same, if not better, test results at a fraction of the cost. Besides it has the advantage that it also allows rotational movement during the drop; this is a plus point since some products like disk drives react more to torsional forces than they do to translational forces. However, free fall drop testers, like shock tables, suffer from poor repeatability.

For any such test, the number of test shocks must also be specified. Some of the organizations that offer shock test standards are American National Standards Institute (ANSI), American Society for Testing and Materials (ASTM), Deutsche Institute fur Normung (DIN) and United States army with its Military Standard MIL-STD-810/883. The last named is a good reference for shock testing and definition for both classical and SRS shocks. Typically for qualification testing of a product, there are 3 (or 5) shocks required per direction (+/–) per axis (*x* /*y* /*z*) with a total of 18 (or 30) shocks, with the peak intensity and duration of the pulses complying with those defined by the specified test condition. For acceptance testing, it may be reduced to one shock per direction per axis with a total of only six shocks. Table 4.4 shows the various test conditions for mechanical shock testing as defined by MIL-STD-883 Method 2002.

TABLE 4.4   Test conditions for mechanical shock testing (MIL-Std-883 Method 2002)

Test condition	g level (peak)	Pulse duration (ms)
A	500	1
B*	1500	0.5
C	3000	0.3
D	5000	0.3
E	10,000	0.2
F	20,000	0.2
G	30,000	0.12

* Default condition, unless otherwise stated.

It is to be noted that in drop testing, change in velocity and change in displacement are not zero, while other methods maintain zero overall change in velocity and displacement. After the mechanical shock test has been completed, external visual inspection of the specimen is generally performed with a magnification of 10 or 20. For analysis, shock signals have to be captured using an appropriate trigger and saved on a storage oscilloscope/pc based recorder/instrumentation tape recorder for future analysis since real-time analysis would be very difficult.

4.4.2.2 Hammer/impact testing The impact testing machine used in naval applications is a pendulum hammer (similar to that used in vibration testing) which strikes an anvil plate to which the specimen is attached. For exciting heavier specimens, the pendulum hammer is made to strike a group of steel channels that are attached to an anvil plate. The plate is allowed to move up to 7.5 cm before being restrained by a ring. There are also heavy-weight shock machines which are essentially floating barges to which the specimen is attached. Explosive charges are set off underwater at prescribed distances away from the barge to excite the specimen.

4.4.2.3 Shaker testing An electrodynamic shaker may also be used for shock testing, for shock pulses less than 3000 Hz, typically. These shakers can impart both classical and SRS compatible shocks. These shock pulses can be controlled within test tolerances better than in other methods due to the control systems available. However, there are issues related to shaker velocity, force and displacement limitations.

4.4.2.4 Pyroshocks These are caused by propellant-activated devices or explosive devices for releasing stored strain energy. Being very lightly damped shocks, they are characterized by highly localized inputs with very short time duration (less than 20 ms), large acceleration levels (100–200,000 g) and high frequency content (100 Hz to 1–MHz). During pyroshock testing, energy is imparted in all directions and there is very little rigid body motion of the structure. The analysis of the shock is typically measured using an SRS, with 5% damping. Peak accelerations and frequency content of pyroshocks vary greatly with distance from the shock event (near-field, mid-field and far-field).

4.4.2.5 Hopkinson bar method This is a method used in the calibration of accelerometers and provides extremely high g levels. The Hopkinson bar uses controlled projectile impacts at the end of a metallic bar, causing a stress wave of known magnitude to propagate through the bar. The unit under test is at the end of the bar and experiences high accelerations and stress with rapid rise time when the shock wave arrives.

FURTHER READINGS

1. Buzdugan, Gh., Mihailescu, E. and Rades M., *Vibration Measurement (Mechanics: Dynamical Systems)*, Springer, Netherland, 1986.
2. Ewins, D.J. and Inman, D.J., *Structural Dynamics @ 2000*: *Current Status and Future Directions,* Research Studies Press Ltd., England, 2001.
3. Inman, D.J., *Engineering Vibration*, Prentice-Hall, Upper Saddle River, New Jersey, 2001.
4. Lalanne, C., *Mechanical Vibration and Shock*, Taylor & Francis, North America, 2002.
5. McConnell, K.G., *Vibration Testing, Theory and Practice*, John Wiley & Sons, Inc., New York, 1995.
6. Piersol, A.G., *Harris' Shock and Vibration Handbook*, McGraw-Hill Professional, New York, 2002.
7. Ramamurti, V., *Mechanical Vibration Practice with Basic Theory*, Narosa Publishing House, New Delhi, 2000.
8. Rangan, C.S., Sarma, G.R. and Mani, V.S.V., *Instrumentation Devices and Systems*, Tata McGraw-Hill Publishing Company Ltd., New Delhi, 1983.

9. Reza Moheimani, S.O. and Fleming, A.J., *Piezoelectric Transducers for Vibration Control and Damping (Advances in Industrial Control)*, 1st ed., Springer, eBook, 2006.
10. Thomson, W.T. and Dahleh, M.D., *Theory of Vibration with Applications*, Pearson Education India, India, 2003.
11. http://www.bksv.com.
12. http://www.engineersedge.com.
13. http://www.globalspec.com.
14. http://www.mbdynamics.com.
15. http://www.mide.com.
16. http://www.mpmtechnologies.com.
17. http://www.sensors-transducers.globalspec.com.
18. http;//www.saraswatidynamics.com.
19. http://www.xcitesystems.com.

Other Transducers and Equipment Related to Vibration Measurement

<div align="right">Chapter
5</div>

5.1 INTRODUCTION

This chapter describes transducers used for the measurement of physical quantities that often complement vibration such as strain, force and pressure. Besides, other measurement devices and equipment such as stroboscopes, slip ring assemblies and telemetry units which facilitate vibration measurements on rotating machines are discussed. Details of operation of signal conditioning amplifiers like carrier frequency and charge amplifiers which are to be used with pickups and power amplifiers to be used with exciters are also mentioned.

5.2 STRAIN GAUGES

Mechanical strain is the ever-present companion of physical deformation. Stress analysis and vibration analysis often go hand in hand in the study of dynamics. In engineering design, anything built from available materials must be tested to see if, regardless of its intended function, its form distributes the forces in acceptable concentrations of strains. Resistance strain gauges are very often used for measuring displacement, force, torque, pressure, acceleration, etc. The effect of elongation on the resistance of the metallic wire was first found out by Lord Kelvin in 1856. The strain gauge is essentially a simple electrical resistor, specially designed to be easily bonded to the surface of a solid object and to undergo a change in resistance when a strain is present in the direction of its sensing grids. For uniaxial loading, there is a single non-zero principal stress in the direction of application of load and zero principal stress in all other directions perpendicular to the axis of load. The resistance R of the wire is given by

$$R = \frac{\rho L}{A} \tag{5.1}$$

where ρ is the specific resistance or resistivity of the conductor material, L is the length of the conductor and A is its (uniform) cross-sectional area. Taking logarithm on both sides of the equation yields

$$\log R = \log \rho + \log L - \log A \tag{5.2a}$$

Differentiating the above equation,

$$\frac{dR}{R} = \frac{dL}{L} + \frac{d\rho}{\rho} - \frac{dA}{A} \tag{5.2b}$$

Since

$$A = \frac{\pi D^2}{4} \text{ and } dA = \frac{\pi D dD}{2}$$

$$\frac{dA}{A} = \frac{2dD}{D} \tag{5.3a}$$

Here D is the diameter of the conductor.

Also

$$\frac{dD}{D} = -v\frac{dL}{L} \tag{5.3b}$$

where v is Poisson's ratio.

Hence

$$\frac{dA}{A} = -2v\frac{dL}{L} \tag{5.4}$$

Substituting for dA/A from Eq. (5.4) into Eq. (5.2), the following equation is obtained.

$$\frac{dR}{R} = \frac{dL}{L} + \frac{d\rho}{\rho} + 2v\frac{dL}{L} \tag{5.5}$$

Dividing the above equation throughout by dL/L

$$\frac{dR/R}{dL/L} = 1 + 2v + \frac{d\rho/\rho}{dL/L} \tag{5.6}$$

The term $(dR/R)/(dL/L)$ often denoted by G is called gauge factor and the term $(d\rho/\rho)/(dL/L)$ is called the piezoresistance effect of the material.

Requirements of a good strain gauge are as follows:

- High sensitivity
- Good stability, repeatability and linearity over a large range of strains
- Freedom from (or ability to compensate for) temperature effects and other environmental conditions
- Low 'drift' or 'null displacement' or 'zero shift' which is a change in the transducer's no-load output signal over time
- Small size and mass
- Ease of fixing to the test specimen
- Ease of production over a range of sizes
- Robustness
- Suitability for static and dynamic measurements and remote recording
- Low cost

5.2.1 Electrical Wire Gauge

In an electrical wire gauge, a single length of wire is used as the sensing element. Strain gauge circuits used for measuring resistance changes impose certain restrictions on the minimum resistance that the strain gauge should have. Absolute resistance of a strain gauge relates to the need to initially balance the Wheatstone bridge. Large resistance imbalances in one or more arms of the bridge may exceed the bridge balance capability of the strain indicator being used. The larger the gauge resistance, larger is the possible change in the resistance for a given gauge factor, lower is the current and less the heat dissipation at a given bridge excitation or in other words higher is the signal-to-noise ratio (SNR) at a given power level. A resistance of the order of $60-1000 \ \Omega$ is normally chosen for optimum performance and gauges are

available with nominal resistances of 120, 350 and 1000 Ω. To achieve these values, a grid pattern is formed, thereby increasing the length of the wire, and at the same time keeping the gauge length and width minimal.

The materials used for fabrication of strain gauges should have a few basic qualities in order to achieve high accuracy, good sensitivity, stability, insensitivity to temperature, long life and the ability to operate under the required environmental conditions. Unfortunately, the most desirable strain gauge materials are also temperature sensitive and their resistance also tends to change with age. Though this may not be a matter of concern for tests of short duration, temperature and drift compensation must be provided for continuous industrial measurement. Materials with high specific resistance, low temperature coefficient of resistance, constant and preferably high gauge factor and constant strain sensitivity over a wide range of strains are preferred. The most popular and widely used alloys are copper–nickel alloys and nickel–chromium alloys. Typical materials for strain gauge applications include Constantan (copper–nickel alloy), Nichrome V (nickel–chrome alloy), platinum alloys (usually tungsten), Isoelastic (nickel–iron alloy) or Karma (nickel–chrome) alloy wires, foils or semiconductor materials. The gauge factor typically varies between 2 and 4 for metallic wire strain gauges. Wire gauges can be classified into bonded and unbonded gauges and are fabricated in four basic varieties, viz. grid, wrap around, single wire and woven.

5.2.1.1 Bonded wire gauge

In this type of gauge, the strain gauge is bonded directly to the surface of the specimen being tested with a thin layer of adhesive cement, which serves to transmit the strain from the specimen to the gauge wires. The stress analyst has to ensure that the bonding is perfect so that the strains are fully transmitted from the surface of the object to the gauges, resulting in $\mu\Omega$ resistance

changes. Such strain gauges are consumables; once bonded, they cannot be removed and re-bonded. The material used for backing and the bonding cement should offer a high insulation resistance and excellent transmissibility of strain from the base to the wires and must be immune to moisture effects. The most common materials used for wire strain gauges are Advance or Constantan alloys. Epoxy paper or bakelite is often used as the backing material to which the grid structure is bonded. The most popular wire gauge is the flat grid type where the wire is wound back and forth as a grid, as illustrated in Fig. 5.1.
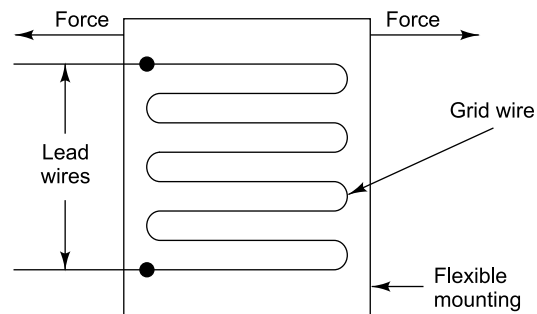
FIGURE 5.1   Wire strain gauge

The grids of these gauges with linear patterns are meant to measure strains in a single (longitudinal) direction. However, since the ends of each section of the wire are looped around, grids are also sensitive to the component of strain acting perpendicular to their length. Thus transverse strains also cause changes in resistance. Transverse sensitivity K_t is the ratio of the resistance change of the gauge produced by uniaxial transverse strain to the resistance change produced by uniaxial longitudinal strain of the same magnitude and is expressed as a percentage. In order to reduce transverse sensitivity, loop lengths in the transverse sections of the wire should be minimized or joined through a material having a lower sensitivity to strain. Fortunately this sensitivity to transverse strain has been greatly reduced in foil gauges through the design of the grids. It is preferable not to have transverse sensitivity greater than 2% of the sensitivity

along the major axis. The wire grid plane should be kept as close to the specimen surface as possible to achieve best transfer of strain.

5.2.1.2 *Unbonded strain gauge* The unbonded strain gauge (Fig. 5.2) consists of a free filament-sensing element and strain is transferred to the resistance wire directly without any backing. The sensing element is made by stretching relatively long strands of high tensile strength wire (typically 25 μm in diameter) around an array of posts that are linked to a diaphragm, such that it experiences relative motion when the diaphragm deforms under pressure. One of these posts is generally attached to a stationary frame and the other to the movable frame of a linkage, which is designed such that when pressure increases, one half of the wire is stretched more than the other half.
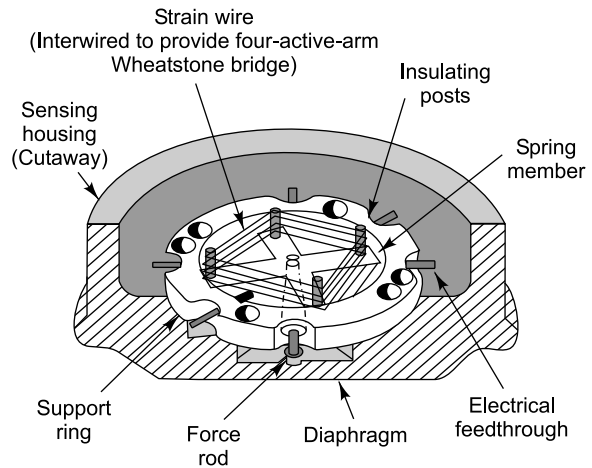
FIGURE 5.2   Unbonded gauge

The main applications of such strain gauges are in displacement, pressure and acceleration transducers. The primary advantage of unbonded strain gauges over bonded gauges is that they have a higher gauge factor, of the order of 3. Because no adhesives are required, they can also be designed and fabricated for use at higher temperatures. Owing to the radial symmetry of the transducer, spurious signals from transverse forces are cancelled. Besides, they have very low hysteresis and creep (due to the absence of backing and bonding). Unbonded strain gauge transducers, however, tend to be large.

5.2.2 Foil Gauge

The foil gauge was born out of a need to minimize the transverse sensitivity of a wire gauge. This gauge is basically an extension of the wire gauge, the main difference being in its construction. In a foil gauge, the required grid pattern is produced by printed circuit technique using a very thin foil of the same material as that used for the wire gauges, with a plastic backing. Owing to the larger surface area, this gauge has higher heat dissipation capability and better thermal stability. Besides, transverse sensitivity is reduced in these gauges by making the perpendicular sections of the foil wide. For the strains to be faithfully transmitted from the specimen to the gauge, the foil diaphragm and the adhesive bonding agent must work together; the adhesive should also electrically insulate the foil grid from the surface of the specimen. The construction of these gauges ensures that there is no stress concentration at the terminals, thereby extending their life. Since photochemical etching processes are used for manufacturing these gauges, they can be made in any complex pattern or size, such as circular gauge or diaphragm gauge, as well as rosettes for two-dimensional stress analysis. Three element rosettes are also available and can be used to determine the biaxial state of strain. Figure 5.3(a) shows a linear foil gauge and Fig. 5.3(b) various types of rosettes, and a circular gauge.

Gauges are available in lengths of 0.25–50 mm, the most commonly used lengths being 5–10 mm. Smaller gauges tend to have degraded performance, since the strain is not uniform along the length; besides it is difficult to handle these gauges. Gauge factors of most metal foil strain gauges are well

behaved: typically 95% (2σ confidence limits) of all gauges constructed have a gauge factor within the limits of the tolerance given. A gauge factor tolerance of ±5% indicates that if the indicated strain is 1000 μm/m, the actual strain lies between 995 and 1005 μm/m.

(a)

(b)
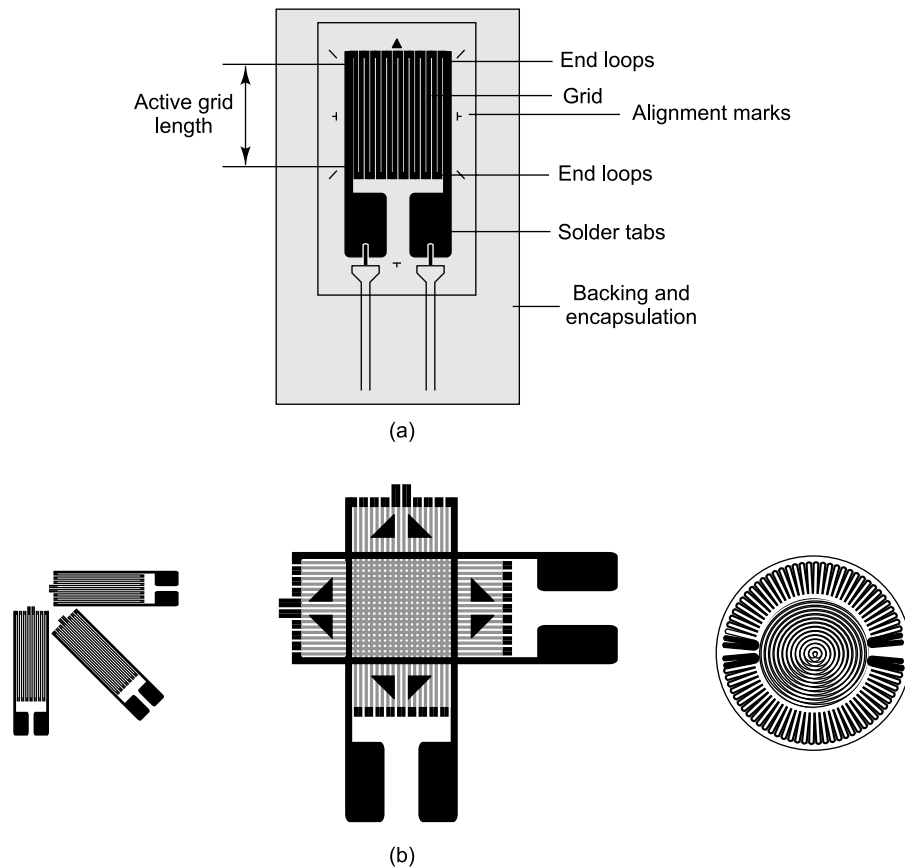
FIGURE 5.3   Foil gauges: (a) linear foil gauge, (b) various types of strain gauges (Courtesy of http://www.omega.com)

Strain gauge specifications

Nominal resistance	:	120 ± 0.4 Ω
Gauge factor	:	2.0 ± 5%
Operating temperature range	:	−20 to +100 °C
Maximum strain	:	5% or 50,000 μm/m
Active length	:	6 mm

5.2.3   Thin Film Gauges

Evaporated, thin film gauges have been receiving a lot of attention of late, especially for applications such as in diaphragm type pressure gauges. The stressed force sensing element in these gauges is generally in the form of a metallic diaphragm on a ceramic layer. These gauges do away with the need for adhesive

bonding of the grid on to the backing. Thin films of metals such as aluminium, gold, nickel, platinum or palladium are made in desired patterns and molecularly bonded directly to a ceramic layer (substrate) by vacuum deposition (thermal evaporation of the desired material in vacuum) or sputtering techniques. The substrate is bonded to the test specimen in the same manner as that used for other gauges. The major contribution to change in the gauge factor in these gauges is due to resistivity changes, rather than geometrical changes. These gauges can operate over a very wide temperature range, say from –200 to +400 °C with good stability. The sensitivity is also high because of a high gauge factor and high gauge resistance. Because the thin-film gauge is molecularly bonded to the specimen, it is highly stable and has very little drift in resistance. These gauges are rugged and have very low hysteresis and creep.

5.2.4 Semiconductor Strain Gauges

Important advances in strain gauge technology have been made through studies of the piezoresistive properties of silicon and germanium. These materials were studied in the mid-1950s by researchers at Bell Laboratories, U.S.A. and were found to have gauge factors more than fifty times, and sensitivity more than a hundred times that of metallic wire or foil strain gages. Silicon wafers were also found to be more elastic than metallic ones, returning quite readily to their original shapes after being strained. They have a major drawback though, in terms of material non-linearity (resistance-to-strain relationship is non-linear varying 10–20% from a straight line equation) and temperature sensitivity. The first silicon strain gauges were developed for the automotive industry around 1970.

In the metal alloy strain gauges described earlier, the strain sensitivity is mainly due to the dimensional change, with a smaller contribution from the resistivity changes. Semiconductor strain gauges, on the other hand, exploit the piezoresistive effects of silicon or germanium and measure the change in resistance with stress as opposed to strain. The construction of a semiconductor strain gauge is different from that of the metallic strain gauge. Figure 5.4 illustrates the general construction of a semiconductor gauge.
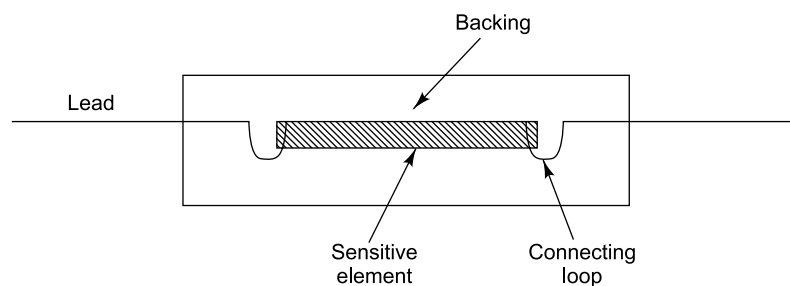
FIGURE 5.4 Typical semiconductor gauge construction

The semiconductor gauge is essentially a wafer with the resistive element diffused into a substrate of silicon. It is fabricated from a single crystal of silicon and germanium using integrated circuit technology. It is made in the form of a single rectangular strip with lead wires and is held by a thin paperback with cement. A thin layer of silicon, typically 0.125×0.0125 mm in cross-section is the common transducer element for the semiconductor strain gauge. Effective lengths may range roughly from 1.25 to 12 mm. Essentially, the same type of backing, bonding epoxies and mounting techniques as those used for metallic gauges, are used for semiconductor gauges also. The main characteristics specified for a semiconductor

gauge are filament material (*p* or *n* type silicon) and gauge factor (positive or negative), gauge length, gauge resistance, temperature coefficient, backing or encapsulation, bonding and lead geometry.

The piezoresistive effect of semiconductor gauges is due to the change of the sensor geometry resulting from an applied mechanical stress. The following relation describes the geometrical piezoresistive effect on the gauge factor of a semiconductor gauge.

$$G = \frac{dR/R}{dL/L} = 1 + 2\nu + \frac{d\rho/\rho}{dL/L}$$

$$= \frac{dR/R}{\varepsilon} = 1 + 2\nu + \pi E \tag{5.7}$$

where *G* is the gauge factor, *dR* is the change in gauge resistance in Ω, *R* is the resistance of the unstrained gauge element, ε denotes the strain, *dL* denotes the relative increase of the length of the element, *L* is the length of the unstrained gauge element, ν denotes the material dependent Poisson's ratio, *E* is the Young's modulus of the gauge along its length and $\pi = \dfrac{d\rho/\rho}{dL/L}$ represents the coefficient of piezoresistance of the material along the axis of the gauge.

Typical specifications

Gauge length	:	5 mm
Resistance	:	200 Ω
Gauge factor	:	120 ± 5%
Strain limit	:	6000 µm/m
Working current	:	<25 mA
Working temperature	:	<80 °C
Temperature coefficient of resistance	:	@ 32 °C <0.15%
Temperature coefficient of gauge factor	:	@ 32 °C <0.15%

Advantages

The major advantage of the semiconductor gauge is the high gauge factor, currently approximately 150, but which may be increased through further development. This represents a marked improvement in sensitivity compared to the 2–4 range exhibited by the ordinary metallic element. It is approximately 100 times more sensitive to strain than the metallic gauge. Other advantages include chemical inertness, freedom from hysteresis and creep effects, good fatigue life and a low transverse sensitivity. Besides, these gauges are smaller in size and are suitable for both static and dynamic measurements.

Disadvantages

The increased sensitivity and higher unit resistance, however, are also accompanied by comparative disadvantages such as the following.
  (i) The output of the semiconductor is inherently non-linear with strain.
 (ii) Strain sensitivity is markedly temperature dependent.
(iii) The output of the semiconductor gauge tends to drift with time.
 (iv) The gauge is somewhat more brittle and fragile than the corresponding wire or foil element, though it can be bent to a radius as small as 3 mm.

(v) The strain range of the semiconductor gauge is roughly limited from 3000 to 10,000 μm/m (depending on the specific gauge type) as compared to an upper limit of 100,000 μm/m for some metallic resistance gauges.

(vi) The semiconductor gauge is considerably more expensive than the ordinary metallic gauges.

(vii) Because of the high sensitivity of the semiconductor element, the non-linearity of the simple Wheatstone bridge cannot always be ignored, as is normally done when conventional metallic element gauges are used. This may necessitate special computer-controlled instrumentation in which these drawbacks can be overcome through software compensation.

5.2.5   Error in Strain Gauge Measurements

- The potentially most serious source of error in strain gauge measurements is thermally induced apparent strain due to resistance changes. Data sheets normally provide plots of apparent μstrain versus temperature. Another source of error is gauge factor variation with temperature. For Constantan alloy gauges, gauge factor typically increases by about 0.01% per °C.

- Another problem is the hysteresis effect. The bonded strain gauge does not sense exactly the same strain as that experienced by the surface to which it is bonded. This gives rise to a hysteresis loop in a plot of relative resistance variation versus strain for cyclic loading. Hysteresis effect can be reduced by repeated cycling to a strain level higher than that required. Faulty bonding can also give rise to large hysteresis, which however cannot be eliminated by cycling.

- Strain gauges are susceptible to creep. If the backing material and the bonding cement are not stronger than the resistive element, the transmission of the strain to the gauge is not faithful, and the indicated strain is less than the true value, becoming increasingly non-linear as strain increases. Creep may be due to a faulty structure in the gauge itself, imperfect bonding or a rise in the operating temperature. It is generally large immediately after gauge installation. Unlike hysteresis, creep is time dependent and has a greater effect on static tests, as well as on small gauges with a small bonding area.

- Fatigue failure is another problem arising in dynamic conditions due to stress reversals. The weakest points in a gauge are the locations where the connecting leads are attached to the gauge. Fatigue failure may be reduced by ensuring that there is no abrupt change in area at the joints.

- Moisture can also have a disastrous effect due to a change in the volume of the backing or bonding adhesives. It leads to a sharp fall in the insulation resistance between the gauge and the specimen.

5.3   TRANSDUCERS FOR FORCE MEASUREMENT

Force transducers are used for static as well as dynamic measurements of forces. They are commonly used to obtain static force versus displacement characteristics. In the context of vibration measurements, they are often used to measure the dynamic force produced by a shaker or impact hammer or for obtaining FRFs in modal analysis. Force may be measured either directly, by using an elastic device positioned in the line of action of the force, or indirectly in which method the elastic deformation of any of the force bearing members is measured to deduce the force. Due to the property of elasticity, the force results in a proportional reaction which can be measured either as displacement or as strain. The accuracy of measurement would depend on the accuracy of the sensor used. A load cell is a transducer which converts a force into an electrical signal; the force being sensed deforms a strain gauge through a mechanical

arrangement. The strain gauge in turn converts the strain to an electrical signal using a Wheatstone bridge and an instrumentation amplifier.

5.3.1   Strain Gauge Load Cell

The load cell may be used for the measurement of static as well as dynamic forces over a wide range of forces with very high reliability. In this transducer the output is obtained as the deformation of an elastic member with high tensile strength. The elastic column is typically made of homogeneous materials, generally steel alloys manufactured to very close tolerances. The main parameters to be considered in designing a load cell are maximum force, relative size and shape, material density, modulus of elasticity, strain sensitivity, deflection and dynamic response. With proper design, a linear relationship between a dimensional change and measured force can be achieved. The materials used should result in low strain hysteresis over repeated loadings and very low creep over long periods of loading. The variation in the modulus of elasticity of the column with temperature, its ultimate strength and ease of fabrication are other parameters to be considered.

   The column type strain gauge load cell is most popular for measuring unidirectional tensile or compressive forces. It consists of a hollow rod made of a material like high carbon steel with chromium and molybdenum or hardened stainless steel which has a large elastic limit and low hysteresis and creep. To facilitate fixing of strain gauges, the middle portion of the rod is made square in cross-section. When a load F acts on a cross-sectional area A, the stress developed is F/A. It is preferable to have a reduced cross-sectional area to obtain large measurable strains. This may be achieved by using a hollow cylinder (tubular column). However slender columns are susceptible to buckling above a particular load. The cylinder length is hence chosen such that there is no buckling in the given range of loading. Typically four strain gauges are bonded to the column. All four may be bonded such as to sense axial strains, or they may be bonded with two gauges sensing in the axial direction and two in the circumferential direction or in Poisson configuration. The second arrangement is more popular. Figure 5.5(a) shows the column load cell with gauges 1 and 3 measuring axial strains and gauges 2 and 4 measuring the circumferential or Poisson strains. Thus, the strain gauges are fixed to the column in Poisson configuration and the measured strains are:

$$\varepsilon_1 = \varepsilon_3 = \frac{\sigma}{E} = \frac{F}{AE}$$

$$\varepsilon_2 = \varepsilon_4 = -v\frac{F}{AE} \tag{5.8a}$$

With a Poisson half bridge arrangement, the resulting bridge output is $(1 + v)$ times that of a quarter bridge with axial gauge alone where v is the Poisson's ratio of the column material. With a full Poisson bridge as shown in Fig. 5.5(b) the input–output relationship for the bridge (excluding resistance R_c is given by)

$$\frac{V_{\text{out}}}{V_{\text{in}}} = \frac{G\varepsilon_1(1+v)}{2 + G\varepsilon_1(1-v)} \tag{5.8b}$$

where G is the gauge factor as before and V_{out} and V_{in} are the bridge output and input, respectively. Such a bridge offers better linearity than a bridge with axial gauges alone. Since all the four gauges experience the same temperature changes, there is temperature compensation inherently built into the bridge circuit. If the applied force is not concentric or if it is at an angle to the axis of the rod, there is a bending moment on the rod. To avoid any error due to this bending strain, the strain gauges are fixed at the centre of the faces. Under this condition, if bending takes place in the rod such that strain gauge 1 is subjected to tensile strain and strain gauge 3 to compressive strain, they produce zero voltage in the bridge circuit since they are connected in opposite arms of the bridge. Thus, the measured load in this transducer is insensitive to any bending load.

As already pointed out, the effect of temperature on the strain gauges is automatically compensated for in the bridge circuit. But to compensate for the effect of temperature on the Young's modulus of the material of the rod, a separate resistance R_c is connected in series with the bridge as shown in Fig. 5.5(b). When the atmospheric temperature increases, Young's modulus decreases; therefore for a given load, the strain increases, increasing the bridge output. This increase is cancelled by the increase in the value of the resistance of resistor R_c, which is kept in the vicinity of the active gauges. The voltage drop across R_c increases due to the increase in temperature; hence the bridge excitation voltage is reduced, which in turn reduces the output voltage of the bridge circuit to the extent it would have increased due to change in Young's modulus. Thus, inaccuracy of such load cells is reduced up to 0.1% of the full-scale reading. Strain gauge force transducers are best suited for static measurements. Figure 5.5(c) shows the photograph of a commercially available load cell.
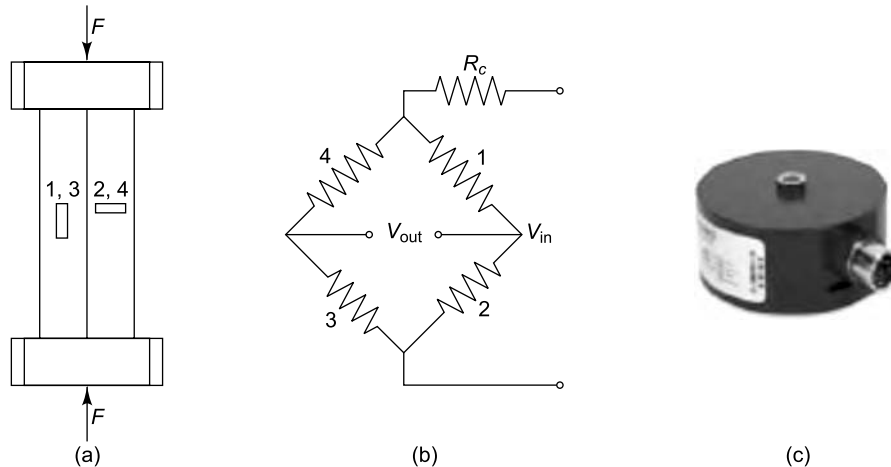
FIGURE 5.5 Column load cell: (a) schematic drawing (b) bridge circuit (c) photograph of load cell
(Courtesy of http://www.pcb.com)

Typical specifications

Force	:	1500 N full-scale
Hysteresis	:	0.1–0.05%
Output	:	2 m V/V
Non-repeatability	:	0.05–0.02%

5.3.2 Piezoelectric Force Transducer

A piezoelectric crystal can conveniently be used in a force transducer since it has all the required characteristics for the same. The piezoelectric force transducer (Fig. 5.6) like the piezoelectric accelerometer works on the principle that the deformation of a piezoelectric crystal produces a charge or voltage output proportional to the force acting on the crystal. However unlike the accelerometer, the force sensor does not have an inertial mass attached to the transducing crystal. Piezoelectric force transducers are suited for the measurement of tensile and compressive forces, for which the crystal has to be physically elongated or compressed for the generation of an output. The transducer responds to forces applied at one or more points on the face of the crystal by producing a proportional charge across perpendicular faces. More than one piezoelectric crystal may be stacked (mechanically in series, but electrically in parallel) to form what are called bimorphs (two elements) or multimorphs (a piezopile consisting of multiple elements), enhancing the charge sensitivity of the transducer in proportion to the number of elements stacked. For modal analysis using shaker excitation, the shaker applies a force through a stinger to the force transducer, which is, in turn, attached to the structure. The force applied to the structure in the axial direction is that transmitted through the piezoelectric crystal (measured) minus the force that is required to accelerate the base of the force gauge. For excitation using an impact hammer, the transducer is fixed to the tip of the hammer and gets compressed when an impact is applied to it. When selecting a force transducer, the way in which the transducer interacts with the excitation device to which it is connected, the exciter or hammer tip, is to be carefully considered, since these can alter the force transducer calibration and cause distortion of the force signal measured, especially at resonances. The mass of the transducer may also result in sensitivity to bending moments which is not desirable.

The force transducer can also be of the ICP type, ICP being PCB's registered trademark standing for 'Integrated Circuit-Piezoelectric' (discussed in Section 3.7.1) and finds wide applications in dynamic structural testing systems for modal analysis and for the prediction of structural response for steady-state continuous and impact-force measurements on small lightweight and delicate structures. It is widely used for the measurement of frequency response functions using a dual-channel analyser, in which case the force transducer is used to measure the input force and an accelerometer (or velocity transducer) is used to measure the response of the structure.

The main advantages of a piezoelectric force transducer are that it has good linearity, low weight, robust construction, as well as excellent and long-term stability. It is characterized by a compact design and is especially suited for applications where space is a premium. It can easily be attached to a stinger or hammer and can also be easily calibrated and mounted. It causes minimal mass loading as compared to other transducers and hence minimal change to the dynamic characteristics of a test structure. Compared to strain gauge sensors it provides excellent resolution. The lowest frequency to be measured depends on the signal conditioner used and is in the range 0.1–1 Hz. Piezoelectric force transducers are to be used in conjunction with regular charge amplifiers.

Typical specifications

Sensitivity	:	300 pC/N
Maximum compression force	:	2000 N
Maximum tensile force	:	300 N
Weight	:	20 gm
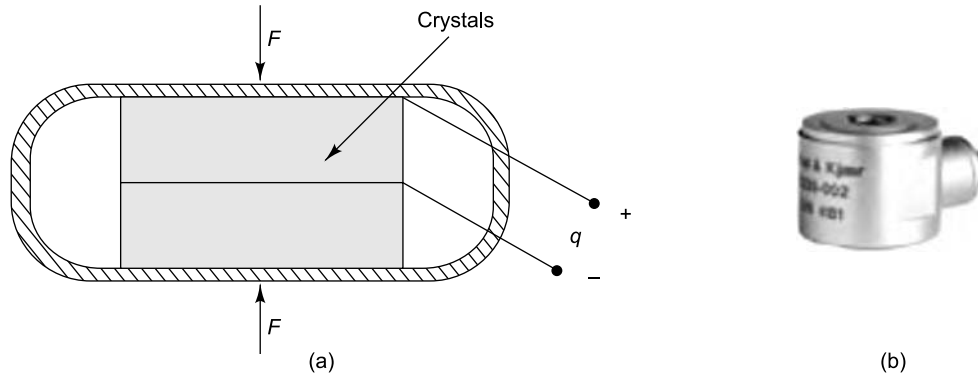
FIGURE 5.6   Piezoelectric force transducer: (a) schematic, (b) photograph
(Courtesy of Brüel & Kjær, Denmark)

5.4   TRANSDUCERS FOR TORQUE MEASUREMENT

Torque measurements are used in process control and monitoring torque is critical to the performance of axles, drive trains, gear drives, electric and hydraulic motors, gas and steam turbines. Torque and speed determine horsepower, which is an indication of system efficiency. There are several methods used for acquiring reliable torque data from rotating objects. The most widely used transducing element for converting torque into an electrical signal is the strain gauge. The gauge is bonded to a beam or structural member that deforms when a torque is applied. The resulting deflection produces a stress, causing the gauge to change its resistance. A Wheatstone bridge is used to convert the resistance change into a calibrated electrical output. Other techniques for the measurement of torque use transducers mounted in the machine train or on the rotating shaft; most of them also utilize strain gages. Each method has its pros and cons and the best solution is obtained from a thorough understanding of the application. The most popular measurement techniques are: (i) strain gauging the shaft and (ii) using in-line torque transducers. The difficulty in both methods is getting power from the stationary supply to the gauges over the stationary/rotating gap and getting the output signal back to the stationary signal conditioning amplifier. Both contact and non-contact methods are in vogue to bridge the gap.

Reaction torque is the moment acting on an object which is not free to rotate. A reaction torque sensor or static torque transducer is used when limited rotation will not damage the power supply and signal output cables. In the design of such a torque transducer, care must be taken to eliminate side loading (bending) and axial loading and to ensure that the transducer is sensitive only to torque loading. Such a transducer finds applications in calibration of torque wrenches, determining the breaking torque of tubes and aircraft fasteners and the opening torque of child-proof containers.

Some torque sensors work on the principle of the Villari effect, which is the phenomenon by which the magnetization of a magnetic material is changed by stretching, compressing, twisting or bending it (as discussed in Section 3.5.4). One component of the torque transducer utilizing this effect is the shaft connecting the sensor to the object being subjected to torque. Twisting the shaft also twists a magneto-elastic (pseudo-magnetic) ring which is rigidly fixed to the shaft. This distorts the crystalline structure of the ring and changes the direction and strength of the magnetic field. This in turn induces current in the

induction coil surrounding the magneto-elastic ring and electrically insulated from it. There is an outside housing that encases the torque sensor and blocks unwanted external magnetic fields.

5.4.1 Strain Gauge Based Transducer

In this method, the existing shaft of the machine can be used by bonding strain gauges on to it. In other cases, a rotary torque transducer may be used. The strain gauge based sensor is nothing but a coupled strain gauged shaft that is incorporated into the existing driveline using a keyway, spline or flange shaft. With strain gages bonded to the shaft, the shaft itself becomes the transducer. This method is suitable provided the applied torque induces at least 150 μm/m. The transducer should be calibrated before it is used. This is done by loading the shaft statically and noting the strains. This is relatively easy in small systems, but becomes extremely cumbersome as loads and shaft sizes increase. One should also be careful in selecting the appropriate location for the strain gauges, in mounting them carefully and protecting them.

The strain gauge based transducer measures the torsional shear strain in a shaft. This makes use of the fact that shear in a plane normal to the axis of the shaft is accompanied by tensional and compressional strains at ±45° to the axis. For this reason, torque gauges similar to that shown in Fig. 5.7(a) are used. The two principal strains occur at ±45° to the direction of pure shear strain produced by the torque. These strains are opposite in polarity, though their absolute values are identical. The magnitude of shear strain at 45° to the axis of the shaft is given by the equation

$$\varepsilon_{45°} = \frac{16T}{\pi d^3 E}(1+v) \tag{5.9a}$$

where T is the torque developed, E is the Young's modulus of the material of the shaft, v is the Poisson's ratio and d is the diameter of the shaft. In order to eliminate the thrust and bending stresses possibly present in the shaft, a second pair of strain gauges may be pasted diametrically opposite to the first pair, on the other side of the shaft. These gauges additionally bring about higher sensitivity and provide temperature compensation. The four gauges are generally wired to form a full Wheatstone bridge. With a full bridge configuration consisting of two 45° gauges ε_1 and ε_2 as shown in Fig. 5.7(a) connected in adjacent arms of a full bridge and two gauges ε_3 and ε_4 bonded 180° away as mirror images of gauges ε_1 and ε_2, respectively, the input–output relationship is given by

$$\frac{V_{out}}{V_{in}} = \frac{G\gamma_{max}}{2} \tag{5.9b}$$

where

$$\gamma_{max} = |\varepsilon_1 - \varepsilon_2| = |\varepsilon_3 - \varepsilon_4| = \frac{32T}{\pi d^3 E}(1+v) \tag{5.9c}$$

Figure 5.7(b) shows a strain gauge based rotary torque transducer. Many of these rotary transducers use standard RS232 digital data to transmit torque, giving fast and effective and easy to use data. Figure 5.7(c) shows a strain gauge based static torque transducer.

Range of specifications available:

Torque rating   :   1–500,000 Nm
Shaft diameter   :   5–1200 mm

Rotational speed	:	0–20,000 rpm
Data sampling rate	:	0–20,000 samples per second
Torque output	:	RS232/RS485/0–10 V/0–20 mA/ ± 10 V /±20 mA
Shaft options	:	Flange/flange, keyway shaft, spline shaft
Accuracy	:	0.1–0.05%
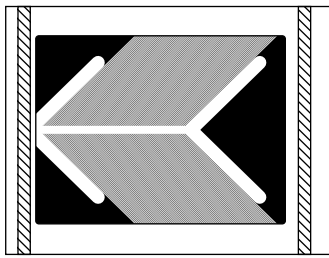
| (a) | (b) | (c) |

FIGURE 5.7   Transducers for torque measurement: (a) torque gauge (Courtesy of http://www.omega.com), (b) rotary torque transducer (Courtesy of http://www.datum-electronics.co.uk), (c) static torque transducer (Courtesy of http://www.datum-electronics.co.uk)

5.4.1.1   *Aids for making torque measurements*   The signal from the strain gauges bonded to a rotating shaft can be retrieved from the shaft using slip rings, rotary transformers or wireless telemetry. Modern torque sensors and torque meters often have additional signal conditioning electronics and a data acquisition system. In the measurement of torque in machinery, three systems are typically employed to feed the excitation voltage to the strain gauge bridge network and to take the output signal: (a) direct contact measurement through slip rings, (b) rotary transformer (non-contacting) or (c) short range telemetry (non-contacting). In all the three cases, accuracy of the order of ± 0.5% full-scale is achievable. These techniques are discussed in detail in Sections 5.6.2–5.6.4.

5.5   PRESSURE TRANSDUCERS

A pressure transducer is a device that converts pressure into an analogue electrical signal. Although there are various types of pressure transducers, one of the most common is the strain-gauge based transducer. The conversion of pressure into an electrical signal is achieved by the physical deformation of strain gauges which are bonded to the diaphragm of the pressure transducer and wired into a Wheatstone bridge configuration. Pressure applied to the pressure transducer produces a deflection of the diaphragm which introduces strain to the gauges. This strain, in turn, produces a change in electrical resistance proportional to the pressure. Pressure transducers are used in many automation applications.

   Many electrical output pressure transducers detect pressure using a mechanical sensing element. These elements may consist of a thin-walled elastic member such as a plate or tube which provides a surface area for the pressure to act upon. When the pressure is not balanced by equal pressure acting on the opposite side of this surface, the element deflects as a result of the pressure. This deflection is then used to produce an electrical output. When another different pressure is allowed on the other side of the

pressure-sensing surface, the transducer will measure differential pressure. If the other side of the surface is evacuated well and sealed, absolute pressure is obtained. The transducer will measure gauge pressure if ambient pressure is allowed on the reference side.

Applications

These transducers have been used successfully, not only for shock/vibration testing, but also in various critical applications including aerospace and nuclear testing. They have typically been used for model testing, wind tunnel and shock tube instrumentation, landing gear hydraulics, rocketry studies, ejection systems and cutting force studies. They have also found applications in combustion, explosion and detonation studies, gait analysis, engine testing, control systems, reactors, building structures, ship structures, auto chassis structural testing, shock and vibration isolation, dynamic response testing and machine health monitoring.

5.5.1   Strain Gauge Pressure Transducers

Although there are various types of pressure transducers, one of the most common is the strain-gauge based transducer. In this sensor the force causing a pressure change, gives rise to a change in the length of a member to which strain gauges are attached and hence a resistance change due to mechanical strain. Since the required deflection is small, the pressure sensing element may be either a diaphragm or a straight tube. If a tube is used, the strain gauges are bonded on it and it is sealed at one end. The pressure difference causes a slight expansion or contraction of the tube diameter. If a diaphragm is used as the sensing element, strain gauges are bonded onto the diaphragm. Pressure applied to the transducer produces a deflection of the diaphragm which introduces strain in the gauges. The strain in turn will produce an electrical resistance change proportional to the pressure. The gauges are wired into a Wheatstone bridge configuration with one to four active gauges. The types of strain gauge elements used in pressure transducers include unbonded metal wire gauges, bonded metal wire gauges, bonded metal foil gauges, thin-film deposited gauges and bonded semiconductor gauges. The advent of integrated circuit technology has led to the development of composite pressure sensors that are very easy to use. These devices commonly employ a semiconductor diaphragm on to which a semiconductor strain gauge and sensor for temperature compensation have been grown. Appropriate signal conditioning is also included in the integrated circuit, providing an output DC voltage or current that is linearly proportional to pressure over a specified range. The strain gauge pressure transducer is advantageous in that it offers solutions that meet varying accuracy, size, ruggedness and cost constraints. These sensors can be used for high and low pressure applications, and can measure absolute or differential pressure. Figure 5.8 shows a typical strain gauge pressure transducer.

FIGURE 5.8   Strain gauge pressure transducer (Courtesy of http:// www.senso-metrics.com)

Typical specifications

Pressure range	:	0.08–1400 MPa
Maximum temperature	:	300 °C
Error	:	±0.1–1% of total dynamic range
Resolution	:	Infinite

5.5.2 Inductive Pressure Transducers

Inductive pressure sensors or variable reluctance pressure (VRP) transducers utilize the operating principle of the LVDT. Though these sensors have been around for some time now, the demand for these devices is increasing due to their suitability for low pressure range applications. In these pressure sensors, a change of fluid pressure is transformed into a proportional change of magnetic reluctance or inductance of a coil. This is achieved by causing the pressure-sensing mechanical element, a bourdon tube, bellows, or diaphragm, to move a part of a magnetic circuit, the core or coil as shown in Fig. 5.9(a). The inductance is a function of the relative motion between the core and the inductive coil. The reluctance is directly proportional to the length of the magnetic flux path and inversely proportional to its permeability. The electromagnetic circuit associated with the coil contains one reluctance element, iron core and air gap path.

In the diaphragm type of pressure transducer which is the most common type of inductive pressure transducer, the change in reluctance of two coils (with a push-pull arrangement) is made use of (Fig. 5.9b). It has the advantage that it is less sensitive to temperature effects than the single coil type. In this transducer, a magnetic stainless steel diaphragm is clamped or welded between two case halves that form two symmetrical pressure cavities. An iron core/coil assembly (encapsulated in a hard compound to maintain maximum stability under high pressures) is embedded in each of the case halves and is electrically connected as a variable transformer (one half of a variable reluctance bridge) constituting the transduction device. When there is a difference in pressures between the two input ports, the diaphragm is deflected. The electrical system is excited by an AC carrier, causing a magnetic flux to be produced in each core and the air gaps formed by the diaphragm. The deflection of the diaphragm with pressure causes the air gap in the magnetic flux path of one core to increase and in the other core to decrease, changing the reluctance and hence the inductance of the two coils. These transducers can be used to measure differential, gauge, or absolute pressure, depending on the pressure maintained in the reference volume side. The inductances may be connected in a Wheatstone bridge arrangement. The change in inductance of the two active arms modulates the carrier signal in direct proportion to the pressure. This amplitude modulated signal must be demodulated to produce a DC output voltage directly proportional to the pressure. In this arrangement it is mainly the air gap reluctance which changes in proportion to the pressure. Hence the reluctance is proportional to the air gap. Figure 5.9(c) shows the photograph of an inductive pressure transducer.

Advantages

The device being rugged withstands environmental qualification requirements such as shock and vibration. Excessive deflection of the diaphragm is mechanically prevented, making bursting nearly impossible and the transducer inherently safe to use. Operational cycles exceeding one million can be withstood, whereas some other types of pressure transducers experience lead wire bending and failure after far fewer cycles. Both sides of the diaphragm can handle most liquids and gases without the need to resort to diaphragm isolators, which are impractical in low-pressure applications. These transducers are, by design, less susceptible to radiation damage than are most other transducers. Since their design uses integral motion as the sensing mechanism, no amplification is required, providing a high output and a high SNR. They can respond to both static and dynamic measurements.

(a)

(b)

(c)
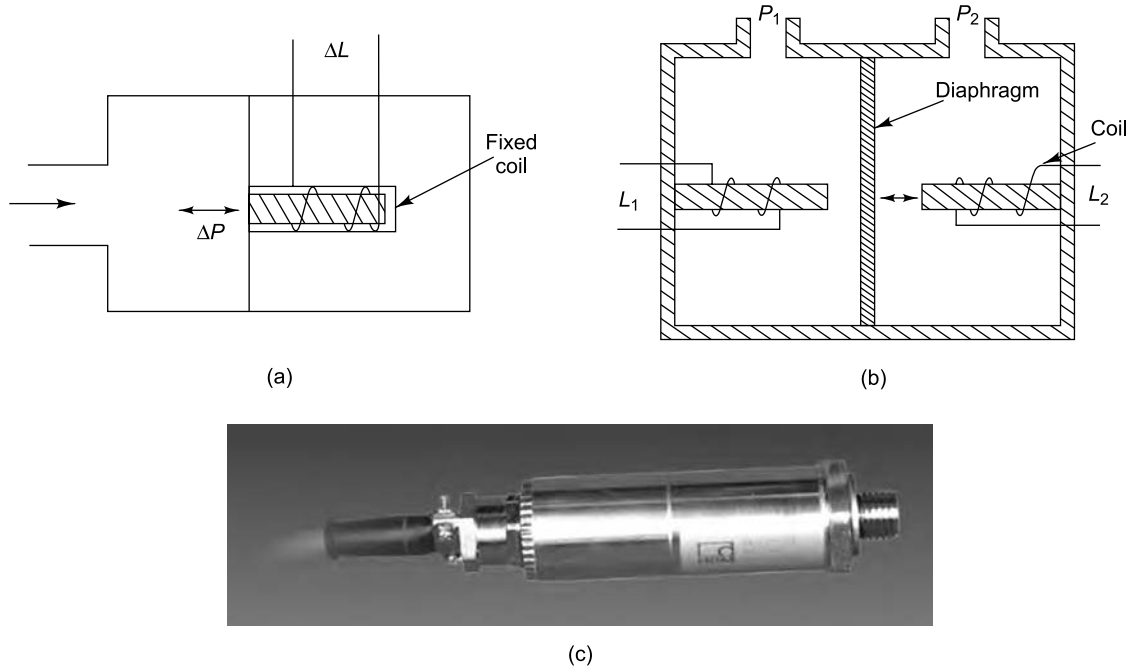
FIGURE 5.9   Inductive pressure transducer: (a) single coil inductive pressure transducer (b) differential reluctive pressure transducer, (c) photograph (Courtesy of HBM, Germany)

Disadvantages

  (i) The frequency response is normally limited by the mechanical construction.
 (ii) The excitation must be AC.
(iii) The transducer must be reactively and resistively balanced at null.
(iv) Magnetic objects and nearby fields can cause transient errors.
 (v) The volumetric displacement tends to be large.
(vi) Mechanical friction can cause wear and errors over a period of time.

Typical specifications

Transducer for absolute pressure

Nominal measuring range	:	0–500 bar
Output signal	:	2 m V/V
Accuracy class	:	0.2%
Nominal temperature	:	−10 to +80 °C

5.5.3   Piezoelectric Pressure Transducers

Piezoelectric sensors use a piezoelectric material to generate a charge or voltage when it is mechanically stressed and thus to measure pressure, acceleration, strain or force. Though the Curies had discovered this effect in 1880, it was not until the 1950s that the piezoelectric effect was used. Two main groups of materials are used for piezoelectric sensors: piezoelectric ceramics and single crystal materials. The ceramic materials (such as Lead Zirconate Titanate, i.e., PZT) have a piezoelectric constant or sensitivity which

is roughly two orders of magnitude higher than those of single crystal materials; the less sensitive crystal materials (gallium phosphate, quartz, tourmaline) however have a much higher long term stability. Depending on how a piezoelectric material is cut, there are three main modes of operation: transverse, longitudinal, and shear. In contrast to the longitudinal and shear effects, the transverse effect allows the fine tuning of transducer sensitivity to the force applied and the element dimensions. Most present day pressure transducers incorporate a quartz element which is sensitive to either compressive or shear loads. The reason why quartz is employed preferentially in transducer designs in spite of the large number of materials available today, is that it has the following excellent properties: high material stress limit, temperature resistance up to 500 °C, very high rigidity, high linearity and negligible hysteresis, almost constant sensitivity over a wide temperature range and ultra high insulation resistance. The finely lapped quartz elements are assembled either singly or in stacks and are usually preloaded with a spring sleeve.

In a piezoelectric pressure transducer, a pressure-sensing diaphragm or membrane reacts normally with the pressure to produce stress on a single disk or a stack of disks made of piezoelectric ceramics or crystalline quartz (Fig. 5.10a). The electrical charges, picked up from the faces of the stack, are proportional to the pressure. In such sensors, a thin membrane and a massive base are used, ensuring that the applied pressure specifically loads the elements in one direction only. Though the voltage generated is supposed to be directly proportional to the applied pressure, it also tends to be sensitive to acceleration. Pressure sensors thus show false signals when they are exposed to vibrations. Sophisticated pressure sensors therefore use acceleration compensation elements in addition to the pressure sensing elements. By carefully matching those elements, the acceleration signal obtained from the compensation element is subtracted from the superimposed pressure and acceleration signal to derive the true pressure information. Some quartz-crystal transducers have been used with amplifiers which permit static measurements, but most piezoelectric pressure transducers are used for dynamic measurement of rapidly varying pressures. Piezoelectric transducers are also used to sense the pressure due to shocks. Frequency response is in the range of 10–50,000 Hz and the pressure ranges are 0–700 bar. Figure 5.10(b) shows the photograph of a piezoelectric pressure sensor.
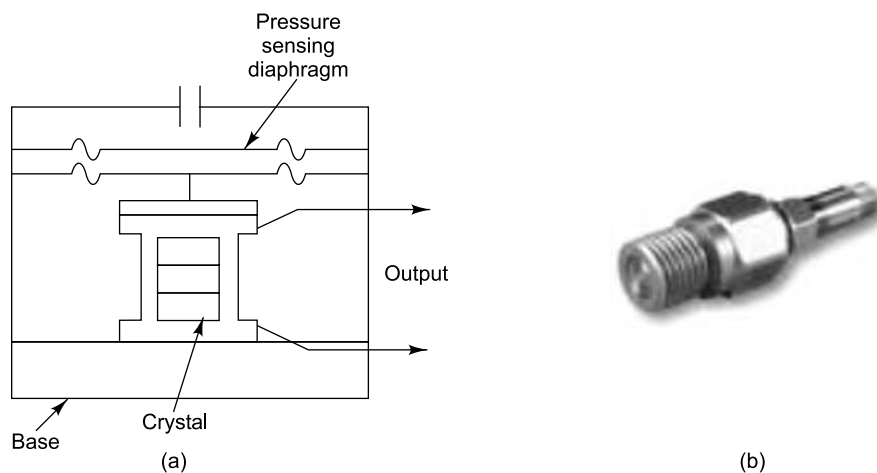
FIGURE 5.10   Piezoelectric pressure transducer: (a) principle of operation, (b) photograph (Courtesy of http://www.pcb.com)

There are piezoelectric transducers of both high and low impedance types. Both these units utilize the same type of piezoelectric sensing element. The main difference between the two is that high impedance units have a charge output which requires a charge amplifier or external impedance converter for charge-to-voltage conversion, while low impedance types incorporate a miniaturized built-in charge-to-voltage converter and require an external power supply coupler to energize the electronics and decouple the subsequent DC bias voltage from the output signal. Generally, low impedance systems have an internally fixed range and time constant and are hence tailored to a particular application. High impedance systems on the other hand, are more versatile; time constant and gain can be controlled by means of the external charge amplifier. Besides, they allow easy short term static calibration and have a wider operating temperature range.

Advantages

Piezoelectric sensors are very rugged and compact. They have an extremely high natural frequency and an excellent linearity over a wide amplitude range. Additionally, piezoelectric technology is insensitive to electromagnetic fields and radiation, enabling measurements under harsh conditions. Some materials used (especially gallium phosphate or tourmaline) have an extreme stability over temperature, enabling sensors to have a working range of up to 1000 °C. Piezoelectrics have the largest strain sensitivity as compared to inductive, capacitive and piezoresistive devices. They also have the widest dynamic range.

Disadvantages

One disadvantage of piezoelectric sensors is that they cannot be used for true static measurements. Elevated temperatures cause an additional drop in internal resistance; therefore, at higher temperatures, only piezoelectric materials that maintain a high internal resistance can be used. However, there are numerous applications that show quasi-static measurements, and many other applications that go to temperatures far beyond 500 °C.

Typical specifications

Measurement range	:	30 MPa
Output voltage	:	±5 V
Sensitivity	:	0.15 mV/kPa
Low frequency response (–5%)	:	0.001 Hz
Resonant frequency	:	>300 kHz
Weight	:	15 gm

5.5.4   Capacitive Pressure Transducers

Capacitive pressure transducers work on the principle that the capacitance between two metal plates changes if the distance between these two plates changes. The capacitance pressure transducer, shown in Fig. 5.11(a) measures the change in capacitance between a metal diaphragm (which serves as one plate of the capacitor) and a fixed metal plate positioned alongside the diaphragm. The pressure acting on the diaphragm causes it to move, changing the capacitance between the two plates. The capacitive sensor can be used as part of an *RC* or *LC* network in an oscillator; alternately it may be used as the reactive arm of an AC bridge, this being the more common of the two arrangements. When it is used in the latter arrangement, the sensor capacitance alone is parameter modulated and another capacitance (an unmodulated reference capacitance) is provided to complete the capacitance half bridge circuit and to

provide thermal compensation. When it is used in an oscillator circuit, the output may be AC, DC, digital or in the form of a phase shift. Some designs use flexible metal diaphragms and silicone oil-filled cavities to isolate the medium from direct influence on the modulated capacitance.

In another configuration which is widely used (Fig. 5.11b), the diaphragm serves as the plate common to two capacitors which are arranged in a push-pull, half-bridge configuration and therefore both capacitors are parameter modulated when the diaphragm is subjected to pressure. The push-pull symmetry arrangement results in the capacitance of one capacitor of the sensor module increasing and the other decreasing when unbalanced pressures act upon the diaphragm. This sensor configuration results in a higher performance in terms of increased sensitivity due to the presence of push-pull symmetry.

Capacitive sensors require dynamic excitation and all capacitive designs contain an internal oscillator (generating a relatively high frequency carrier signal) and signal demodulator to provide AC outputs going down to almost 0 Hz. The range is 0.001–700 bar with a typical error of 0.25%. Units are also available with accuracies to 0.05%. Figure 5.11(c) shows a commercially available capacitive pressure transducer.
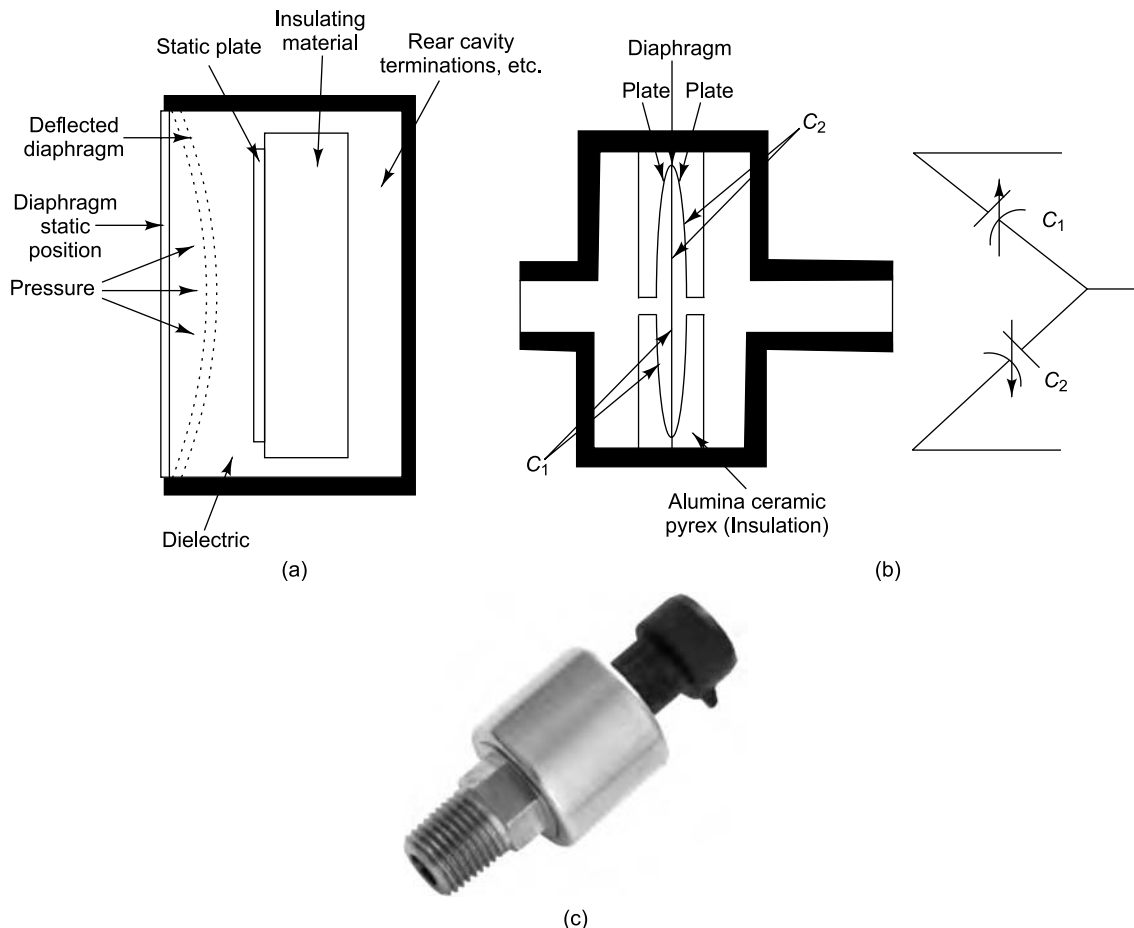
FIGURE 5.11   Capacitive pressure transducer: (a) single active capacitance configuration, (b) two active capacitances configuration, (c) photograph (Courtesy of http://www.kavilco.com)

Advantages

  (i) These transducers are small in size and have a high frequency range.
  (ii) They are inexpensive to produce and have small volumes.
  (iii) A low shock response is possible due to minimum mass required for the diaphragm.
  (iv) They allow the measurement of both static and dynamic quantities.
  (v) These pressure transducers are generally very stable and linear.

Disadvantages

  (i) The high impedance output must be reactively and resistively balanced, bringing about a need for additional signal conditioning amplifiers.
  (ii) Movement of long connecting cables will cause distortion of measured signals.
  (iii) They are sensitive to high temperatures and are more complicated to setup than most pressure sensors.

Typical specifications

Pressure range	:	0–70 bar
Output voltage	:	4.5 V
Response time	:	15 ms maximum
Error	:	<2% of maximum value
Operating temperature	:	–40 to +125 °C
Weight	:	90 gm

5.5.5   Pressure Transducer Errors

A common cause of sensor failure in fluid pressure measurement applications is dynamic impact, which generates a pressure spike and which results in sensor overload, damaging it. A classic example of overloading a pressure sensor is the water hammer phenomenon. To reduce the effects of impacts, sensors are often mounted with a 'snubber' between the sensor and the pressure line. A snubber is usually a mesh filter or sintered material that allows pressurized fluid, but does not allow large volumes of fluid, thus preventing pressure spikes. A snubber is a good choice to protect the sensor, but in applications in which the peak impact pressure is of interest, it is appropriate to select a pressure sensor that does not include overprotection.

    Another major source of errors in many pressure transducers is ambient temperature variation. This could be due to two reasons:

  (i) The zero setting may shift with temperature due to unequal mechanical expansion of the different members of the instrument. In order to minimize the zero error with temperature, the differential expansion of the mechanical components must nearly balance.
  (ii) The calibration factor or sensitivity may change with the ambient temperature owing to a change in elasticity or spring constant of the gauge members. Many metals have a temperature coefficient for Young's modulus of elasticity of about –0.007 °C.

5.6   Equipment which Facilitate Measurements on Rotating Structures

The need for determining the vibration and stress levels under actual operating condition is felt for all equipment even at the design stage and they are obtained analytically. But an experimental validation, done after the prototype is ready, not only verifies the analytical method, but also enables identification of failure-prone situations. Measurements on stationary components do not pose any difficulty since standard methods are available. However, such measurements on rotating components are cumbersome and pose difficulty in taking out cables from the rotating parts. The tachometer is the most commonly used device for directly measuring the rotating speed of a shaft. The stroboscope is a slightly more complicated piece of equipment used to measure rotating speeds of objects or to make them appear stationary for visual inspection. The slip ring unit comes in handy for vibration and strain measurement on rotating structures and uses contacting slip rings and brushes to take leads out from rotating members to external stationary equipment. The telemetry system is another option in which signals are transmitted from transducers on rotating machines to stationary measuring/analyzing devices.

5.6.1   Stroboscope

The stroboscope utilizes the property of persistence of vision. The human eye responds so slowly to light stimuli that it cannot differentiate between two light impulses reaching the eye within 0.1 seconds of each other. Impulses reaching the eye in quick succession are observed by the eye as a continuous unbroken sequence caused by this temporal aliasing or the 'stroboscopic effect'. Joseph Plateau and Simon von Stampfer simultaneously and independently invented the stroboscope, which is a device exploiting this property of the eye to make objects with periodically repetitive motions appear to be slow moving, or stationary, thus enabling them to be observed more conveniently. Commercially available stroboscopes are capable of operating at speeds between 600 and 20,000 rpm.

The stroboscope is used for measuring speeds of rotation of shafts and other parts of machinery without stopping the machine. It is also used for the study of reciprocating and oscillating motion of structures and machine parts and in the study of stresses in machinery in motion and is thus very useful in machinery diagnostics. The stroboscope is also used to measure car engine speeds; such a stroboscope does not have an independent oscillator to control the flash frequency; instead it uses the ignition timing of the engine and is actually meant for measuring the phase angle (i.e. the timing of the spark relative to the position of the camshaft) rather than the frequency. This device is also used as a diagnostic aid in medicine to view the vocal cords. Flashing strobe lamps are also used in discotheques where they give the impression of dancing in slow motion. In high speed photography very short flashes of light are used as a means of producing still photographs of fast-moving objects, such as bullets in flight. The stroboscope, however, cannot be used where the ambient light intensity is high, since it requires a subdued surrounding light for its efficient operation.

5.6.1.1   *Mechanical disk-type stroboscope*   A mechanical disk type stroboscope consists essentially of a whirling disk attached to a motor, the speed of which can be varied and measured. In its simplest form, a rotating disc with a single opening or a number of evenly-spaced holes is placed in the line of sight between the observer and the rotating shaft. A reference mark on the rotating shaft is observed through the openings in the rotating disk. The rotational speed of the disk is adjusted so that it becomes

synchronized with the movement of the observed shaft, making the mark on the shaft appear to be stationary. When this happens, the shaft speed is equal to that of the rotating disk, or some integral multiple of this speed and can be expressed as

$$\text{Shaft speed} = \frac{(\text{disk speed}) \times (\text{number of openings in the disk})}{\text{number of images}}$$

The number of images refers to the number of times a single mark on the shaft appears when viewed through all the openings in the disk when the images appear be stationary. Figure 5.12 shows a mechanical type disk stroboscope.
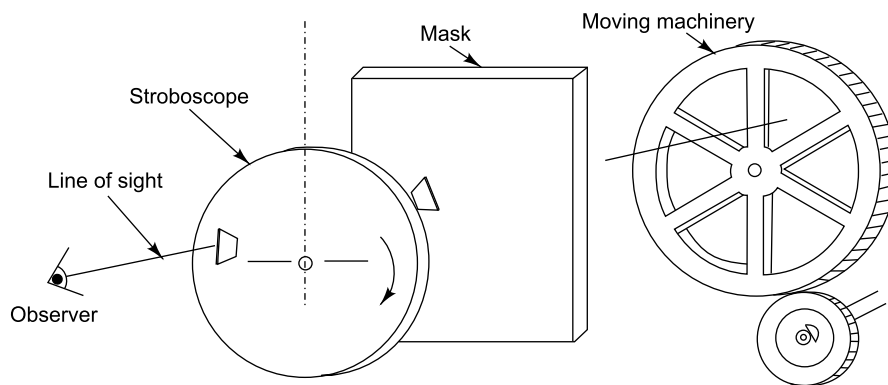
FIGURE 5.12 Mechanical disk-type stroboscope

5.6.1.2 Electrical stroboscope In an electrical stroboscope, the perforated disc is replaced by a neon gas discharge lamp, capable of emitting brief and rapid flashes of light. The rate of flashing is controlled by a variable frequency electronic oscillator, the frequency of which can be read off a dial/display attached to its tuning control. In operation, these flashes of light are directed on to the rotating, reciprocating, or vibrating member and the rate of flashing is adjusted to be equal to the frequency of motion of the target, causing it to appear stationary. Thus a repetitive motion occurring 2400 times a minute will appear to stand still if viewed against repetitive light flashes occurring 2400 times per minute or 40 Hz. If the frequency of the flashes is less than that of the rotating member or shaft, then the light will flash after the shaft makes a complete revolution, and it will appear as though the mark on the shaft is slowly rotating forwards. Similarly, if the frequency of the light flashes is greater than the shaft's rotation rate, the mark will appear to turn backwards. Figure 5.13 demonstrates the use of an electrical stroboscope.
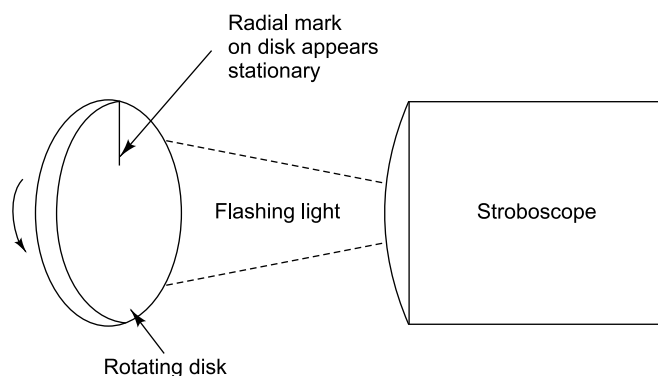
FIGURE 5.13 Flash light stroboscope

Let us consider the following cases:

Case (i) Single mark on the shaft: A single mark on the rotating shaft gets illuminated in the same position for each flash and hence appears to be stationary when the time between flashes is equal to the time for one rotation or in other words when the flashing frequency f_f equals the rotational frequency f_r, i.e., $f_r = f_f$. However it is to be noted that the mark will also appear to be stationary if the shaft speed is an integral multiple of flashing frequency, i.e. $f_r = nf_f$ where $n = 1, 2, 3, 4,...$, etc. Thus a single mark on a shaft rotating at 2400 rpm would complete two revolutions between flashes if these are at the rate of 1200 rpm. The shaft would appear to be stationary at 2400 flashes per minute, and also at 1200, 800, 600, etc. flashes per minute.

However, when the rotational frequency is less than the flashing frequency ($f_r < f_f$) and is a submultiple of f_f, multiple stationary images would be obtained as shown in Fig. 5.14. Let us assume that a flash occurs when the mark is at A. If $f_r = \dfrac{f_f}{2}$, the next flash will occur after half a revolution (i.e. when the mark is at B), and again at A when one revolution has been completed. Subsequently the cycle would be repeated. Apparently the mark will be illuminated at two positions 180° apart causing a double image to appear. With $f_r = \dfrac{f_f}{3}$, the mark will be illuminated at three positions 120° apart causing a treble stationary image to appear. This argument can be extended for cases $f_r = \dfrac{f_f}{4}, \dfrac{f_f}{5}, ...$, etc.
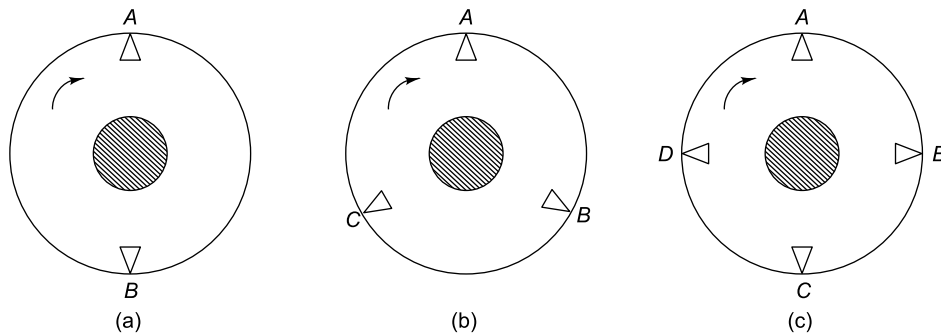
FIGURE 5.14 Multiple images for a single mark

Case (ii) Multiple marks on the shaft: When a number of identical and equally spaced marks are made on the shaft, we get a stationary pattern when $f_r = f_f, 2f_f, 3f_f, ...$, etc. and also for certain values of f_r which are smaller than f_f, i.e. $f_r = \dfrac{f_f}{N}, \dfrac{2f_f}{N}, \dfrac{3f_f}{N}, ...$, etc. where N is the number of marks made on the shaft.

5.6.2 Slip Ring Unit

When it is required that transducers be mounted on rotating parts of machines, some provision must be made to bring excitation power to the transducer from the stationary power source and to take the output signal away to the stationary signal conditioning amplifier and/or recording device. If the measurements can be completed over a few rotations, the connecting wires can simply be allowed to

wind or unwind on the rotating shaft. This is convenient for slow speed machines. When the relative motion involved is small, continuous flexible conductors (often in the form of light coil springs) can be used. However, for measurements during continuous high-speed operation, slip rings, radio telemetry, or some form of magnetic coupling between rotating and stationary parts are required.

The slip ring unit has a wide range of applications, allowing use of inductive and resistive transducers, strain gauges, thermocouples, as well as other transducers. It is an electromechanical device that allows the transmission of power and electrical signals from a stationary to a rotating structure and is similar to the brushes and commutators found in many types of DC motors. Also called a rotary electrical joint, collector or electric swivel, a slip ring can be used in any electromechanical system that requires un-restrained, intermittent or continuous rotation while transmitting power and/or data. It can improve mechanical performance, simplify system operation and eliminate damage-prone wires dangling from movable joints.

The main parts of the slip ring head are a stator with brushes and a rotor with rings; the arrangement is such that it helps easy mounting. Figure 5.15(a) shows two pairs of circular slip rings, one pair for providing excitation voltages (either AC or DC) to the bridge on the rotating shaft and the other for retrieving the signal from strain gauges to the stationary signal conditioning and data acquisition systems. The rotor consists typically of 6 or 12 hard slip rings which are conductive circles or rings or bands mounted on the rotor shaft and insulated from it, while the stator has two brush arrays 180° diametrically apart. The rotating slip rings are generally made of or coated with gold, silver, rhodium or other noble metals and alloys. The stator brushes run in contact with the ring, transferring electrical power or signals to the exterior, non-rotating part of the system. Block-type brushes often are of sintered graphite, while wire-type brushes are alloys of platinum or gold. These are the best suited materials for the application, ensuring a very small amount of wear, as well as high transmission quality. In order to maintain a high signal to noise ratio and good accuracy, the dimensional tolerances, surface hardening of the slip rings and the contact resistance of the brushes are to be carefully considered. The stator is equipped with 6 or 12 solder tags, corresponding to the number of slip rings, for the connection of cables. Each ring is adjacent to the next along the centreline, somewhat like the threads on a bolt; electrical connections from transducers on the rotor or shaft are made to the ring. The rotor part of the slip ring can be fixed either directly or via a mounting plate to the front end of the shaft. The brushes can be engaged even with the shaft running or can be lifted afterwards by just turning a ring, enabling easy replacement of worn brushes. Slip ring brushes, as well as the support bearings internal to these transducers undergo a lot of wear and tear. The slip ring assemblies have to be fitted such that rough dirt, dust, humidity, oil, solvents and gases which could influence the slip ring transmission or reduce the resistance to ground of a connected measuring point cannot reach the slip rings or get into the case. In a typical industrial application, maintaining an oil-free slip ring is difficult. Figure 5.15(b) shows a typical slip ring assembly.

An important point to be considered in using slip rings for transmission of low level instrumentation signals is the electrical noise produced at the sliding contact. A part of the sliding noise is due to thermocouple action if the brushes and rings are of different materials. The other part of the noise is due to random variation of contact resistance caused by surface roughness, vibration, etc. It is possible to reduce this noise by filtering, but this is inconvenient, and not always very effective. The better remedy is to ensure that there is a very small thermo-voltage generated between slip rings and brushes, that the contact resistance between slip rings and brushes is very small (<1 mΩ), and that the variation of contact resistance with speed of rotation, temperature, current, etc. is small. A high quality miniature sliding slip ring may exhibit a contact resistance variation of the order of 0.05 Ω peak to peak and 0.005 Ω RMS.

When the rings are used with strain-gauge circuits, particular care must be taken, since the resistance variation of the sliding contact may be of the same order as the small change in resistance of the strain gauge to be measured. In such cases it is advisable to use the full bridge circuit, as against the half or quarter bridge configurations, so that the sliding contacts can be taken out of the bridge circuit, leading to minimal influence from the slip ring resistance variations. For very precise measurements, the 700 Ω strain gauge full bridge circuit is preferred; however, even with 350 or 120 Ω full, half, or quarter bridges, the errors will be comparatively small. For the most demanding applications, more complex schemes are available to reduce noise to even lower levels.
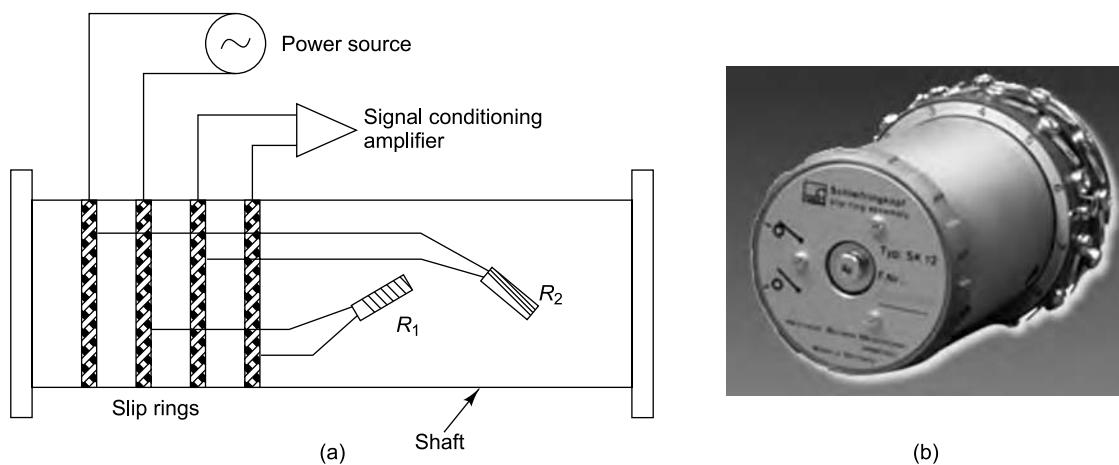
FIGURE **5.15**   Slip ring unit: (a) slip rings, (b) photograph (Courtesy of HBM, Germany)

Disadvantages

  (i) Slip rings have a limitation in terms of the number of simultaneous channels of data that can be collected. The larger the number of channels, the bigger the slip ring becomes, until at around 30 channels, the entire unit becomes very unwieldy. If engineers want to collect dynamic data from a large number of strain gauges, they have to perform two or more test runs. For high speed data transfer or for data transfer in electromagnetic interference (EMI) sensitive environment, fibre optic rotary joints (FORJs) are available.

  (ii) Static charges are another source of problem and disturbances from the same may be got rid of by connecting rotor and stator housing earth with machine earth.

  (iii) Another problem is wear and tear of brushes and rings; this can be minimized by engaging the two only during the time of actual measurements.

  (iv) Besides, difficulty is faced in applications requiring speeds above 100,000 rpm; extreme care is required because of heating and vibration problems.

Typical specifications

Number of slip rings	:	6 or 12
Resistance between slip ring and brush	:	<40 mΩ
Fluctuation of resistance	:	<2 mΩ

Thermo-electric voltage measured between
two brushes when the slip rings are short-

circuited (after running until warm)	:	<10 μV
Current limit	:	2 A
Permitted speed for continuous duty	:	1–6000 rpm

5.6.3 Measurements Using Telemetry System

Slip rings have been relatively satisfactory for measurements on rotating machines; they however have the drawback that the measurement necessitates extension of the rotating shaft and a variety of other modifications and hence their use is not always feasible for in-situ measurements. The use of the short range radio telemetry (meaning distant measurement) technique, where the data is acquired through wireless transmission of radio waves, is more reliable and convenient for measuring parameters such as vibration, stress, torque, etc. in rotating machines. The sensor (which could be a strain gauge, a thermocouple or an accelerometer) is fixed on the rotating shaft (Fig. 5.16a) and is powered by a battery mounted on the shaft itself. The sensor output modulates a radio frequency (RF carrier) emitted by a miniature shaft-mounted transmitter to a rotating loop antenna. A concentric stationary antenna placed very close (less than 50 mm away) to the rotating antenna induces power in the loop antenna on the rotating shaft and picks up the signal and sends it to a stationary receiver, as shown in Fig. 5.16(b). The receiver receives the modulated signal from the stationary antenna and demodulates it to retrieve the information and display it. The received signals may be in the form of amplified analogue DC/AC voltages and may be analyzed using oscilloscopes or precision voltmeters or Fast Fourier Transform (FFT) analyzers/signal analyzers or can be recorded on a multichannel instrumentation tape recorder for later analysis. Besides offering an easy solution, this method provides completely noise-free signals with good accuracy and resolution. No additional signal conditioning of the received signals is generally required. The schematic block diagram of a typical telemetry system is shown in Fig. 5.16(b).

The telemetry system generally comes with the following subsystems:
- A static strain transmitter for static strain and torque measurements.
- A dynamic strain transmitter for dynamic strain measurement using strain gauges and piezoelectric transducers as well as for vibration measurement using accelerometers.
- A piezoelectric module which is used along with the dynamic transmitter as charge converter while using vibration accelerometers and other piezoelectric sensors.
- A signal receiver used for amplification of the received input and for demodulation of the transmitted signal to obtain calibrated strain/vibration data.
- Batteries for power supply to the transmitter. A power on/off switch is also generally provided and is useful in tests where time interval between assembly and test is quite long.

5.6.3.1 *Factors to be considered in designing a telemetry test*   Considerable effort is required in designing a telemetry instrumentation set-up. Some of the important points to be considered before carrying out a telemetry test are as spelt out in the following section. With appropriate selection of transmitters, telemetry instrumentation can be utilized to collect signals from various sensors viz. strain gauges, vibration accelerometers, piezoelectric crystals and thermocouples. These sensors are to be wired to the transmitter with appropriate instrumentation cables. The cables are to be routed along the surface of the rotating component and bonded properly to it so as to prevent them from flying out during rotation. Miniature FM transmitters along with battery have to be mounted on the rotating component. The following points are to be considered:
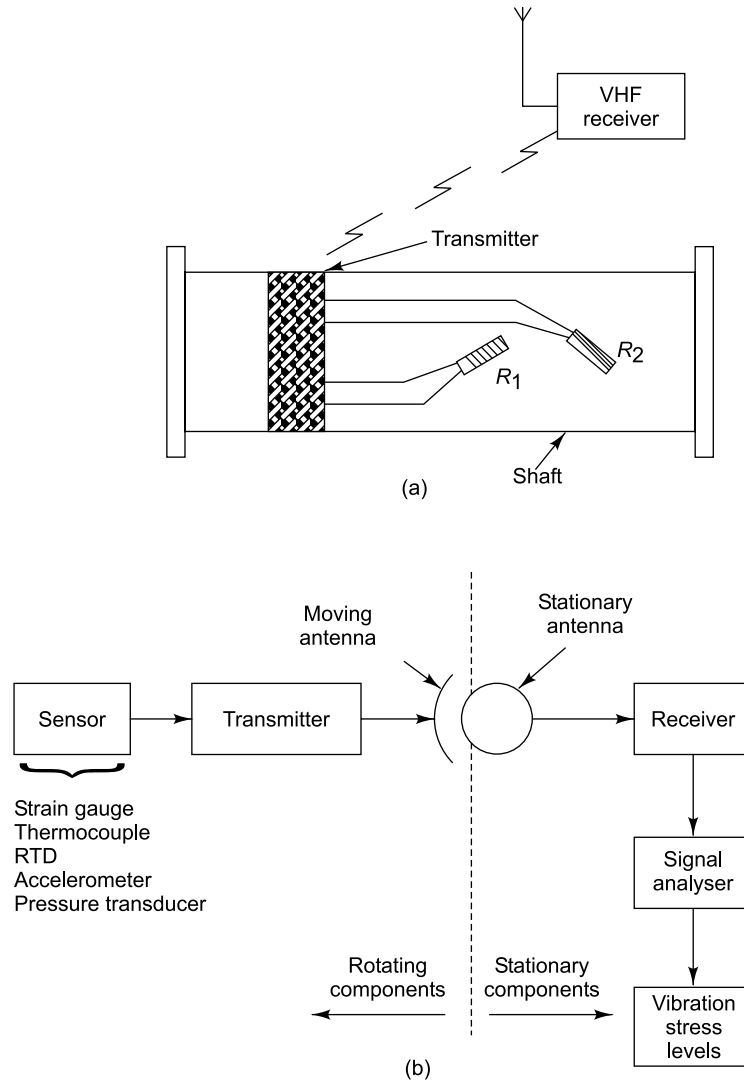
FIGURE 5.16   Telemetry setup: (a) sensor and transmitter on rotating shaft, (b) block diagram

(i)   The sensors mounted on the structure should have sensitivity large enough to give a measurable signal output. They should not create any significant stress or imbalance in the rotating structure. Hence low weight accelerometers are preferred for vibration measurement and foil type resistance gauges or semiconductor strain gauges are preferred for analyzing strain signals.

(ii)   Neither the sensors nor the transmitters should cause any static or dynamic imbalance in the rotating machine; hence they are generally mounted as diametrically opposite pairs.

(iii)   The battery and the transmitter used for these applications typically have limiting static acceleration levels of the order 10,000 or 15,000 g. If large centrifugal forces are expected, as when these devices are located at large radii or when the speed of the machine is high, a very strong transmitter

housing is required. The cables/wires also have to withstand large centrifugal forces. Hence it is preferable to locate the transmitter at locations with the minimum radii from the shaft centre. The fixture and the housing should be light in weight, but at the same time strong enough to withstand the centrifugal forces generated during the operation of the machine.

(iv) There should be provision to mount bridge completion strain gauges and balancing circuits for strain gauge measurements.

(v) The transmitter and the batteries can typically operate up to a maximum temperature of 125 °C. However, when it is required that measurements are carried out at higher temperatures, these units are to be mounted at locations where temperatures are lower, and the sensor wiring has to be routed suitably for connections to the transmitter.

(vi) Both antennas have to be a little flexible for ease of mechanical installation and should be adjusted for maximum coupling so as to obtain maximum induced power and received signal strength.

The telemetry system employs capacitive coupling to transmit the signal from the rotating antenna to the stationary antenna placed at a distance not exceeding 50 mm. In cases where the shafts are of small diameter (less than 200 mm), the antenna design is quite simple with a single loop of wire, properly anchored on the rotating shaft. For shafts of large diameters, more complicated antenna schemes are usually required, such as properly designed dipole antennae, matched to the transmitter output impedance.

5.6.4 Non-contact Rotary Transformer

A rotary transformer can be employed instead of the slip rings or telemetry system described above. This is especially beneficial when it is required to eliminate the noise due to contact vibrations. The system comprises two circular rotary transformers, one to induce the AC bridge supply to the strain gauge bridge and the other to pick up the output signal, as illustrated in Fig. 5.17. The excitation signal is normally a carrier of 20 kHz. The transformers consist of a pair of concentrically wound coils which are suitably aligned for maximum mutual induction, with a high permeability core to improve efficiency.
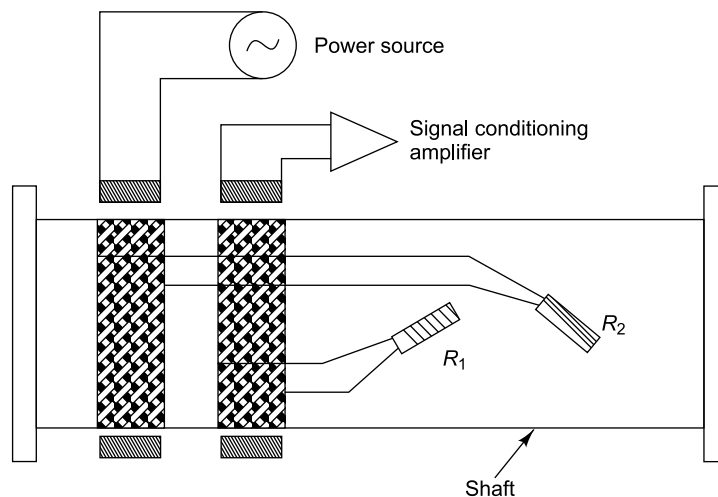

FIGURE 5.17 Rotary transformer

5.7 SIGNAL CONDITIONING AMPLIFIERS

We have seen from discussions in Chapter 3 that the principle of operation of each transducer is unique. Signal conditioning devices are typically used to convert the output of a transducer to a form (convenient voltage level) that is easily measured or recorded or analyzed by the system. In the absence of these conditioners, some signals cannot be measured at all. The operations that a signal conditioning device performs are amplification of small signals, attenuation of large signals, signal shaping and isolation of signals from transducers before they are sent to the recording, measurement or analysis hardware. Other functions include thermocouple compensation for temperature measurements, current sourcing for 2-wire and 4-wire resistance measurements, filtering to remove system noise and fixing shunt resistors for current measurements or for calibration. Some data acquisition systems have built-in signal conditioning components. The signal conditioners for commonly used pickups are shown in Table 5.1. Other transducers may have their own specific signal conditioning units.

TABLE 5.1   Signal conditioners

S.No.	Transducer	Output level	Signal conditioner
1.	Electrodynamic pickup	Low level voltage (50 mV)	Broadband amplifier
2.	Electromagnetic pickup	Low level voltage (50 mV)	Broadband amplifier
3.	Piezoelectric pickup for acceleration/force/pressure	Low level, high impedance charge	Charge amplifier
4.	Inductive transducer for displacement (LVDT)/ angular motion (torsional vibration)/acceleration/pressure	Change in inductance	Carrier frequency amplifier with constant voltage supply
5.	Capacitive pickup for displacement, pressure or force	Change in capacitance	Closed loop AC servo bridge
6.	Eddy current probe	Change in inductance	High frequency carrier amplifier
7.	Resistance strain gauge transducer for stress/force/torque	Change in resistance	Carrier frequency amplifier with constant voltage supply
8.	Fibre optic transducer	Low level voltage (50 mV)	Broadband amplifier
9.	Semi-conductor strain gauge	Large change in resistance	Constant current bridge
10.	Electrodynamic/electromagnetic exciters and PZT exciters	Low level current	Linear power amplifier

5.7.1   Signal Conditioners for Resistive, Inductive and Capacitive Transducers

The signal conditioning device used for a variety of sensors with passive elements such as resistors, capacitors and inductors is the Wheatstone bridge. The operation of this bridge with DC excitation for resistive arms is discussed in the following section. For reactive arms such as those constituted by inductances or capacitances, a bridge with AC excitation may be used.

5.7.1.1   The Wheatstone bridge   The Wheatstone bridge is the most commonly used sensing circuit to measure resistance changes produced when active resistance strain gauges are exposed to strain. This is a very simple circuit allowing measurements of impedances with very high accuracies. Typically the

bridge circuit consists of four 'arms', each consisting of an impedance element which could be resistive, capacitive or inductive. The four impedance elements are arranged in this circuit such that there are two parallel voltage divider circuits, each divider consisting of two elements in series. Bridge excitation voltage is applied to the circuit at the junctions of the two series circuits and the output voltage is measured between the junctions of the series elements in each parallel arm. A resistive bridge is provided with a DC excitation a few Volts in magnitude. Care should be taken to minimize the non-strain resistance changes in the gauges and all other conductors within the bridge circuit.

Consider the simple case when just two resistors R_1 and R_2 are placed in series across a constant voltage supply V_{in}. A portion of the supply voltage is dropped across R_1 and the remaining across R_2.

$$V_{in} = V_{R_1} + V_{R_2} \qquad (5.10)$$

When two more resistors R_3 and R_4 are also placed across the same voltage supply (Fig. 5.18), they too function as a voltage divider. Such a circuit is called a full bridge circuit. Depending upon the values of the four resistors, a differential voltage can also be present between points B and D. This voltage V_{out} can be measured with a good high impedance voltmeter and can be calculated as shown in Eq. (5.11).

$$V_{out} = V_{in} \left[\frac{R_4}{R_4 + R_3} - \frac{R_1}{R_1 + R_2} \right] \qquad (5.11)$$

Corresponding to a non-zero input voltage, the output signal V_{out} is 0 when there is a condition of resistive balance or when

FIGURE 5.18 Full bridge circuits

$$\frac{R_1}{R_2} = \frac{R_4}{R_3} \qquad (5.12)$$

However in actual practice, a true resistive balance is seldom achieved. Rather a zero balance circuit like that shown in Fig. 5.19 is often externally applied to the bridge in order to eliminate the small initial imbalances by changing the ratio of the resistances in the two adjacent arms.

For null balance to occur with this configuration:

$$\frac{R_{R_1}}{R_{R_2}} = \frac{R_4}{R_3} \qquad (5.13a)$$

where

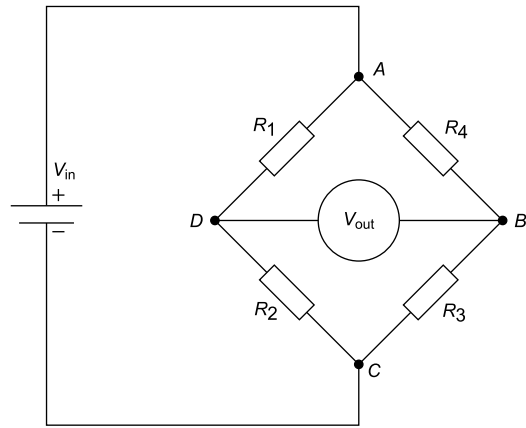$$R_{R_1} = \frac{R_A \times R_1}{R_A + R_1} \qquad (5.13b)$$
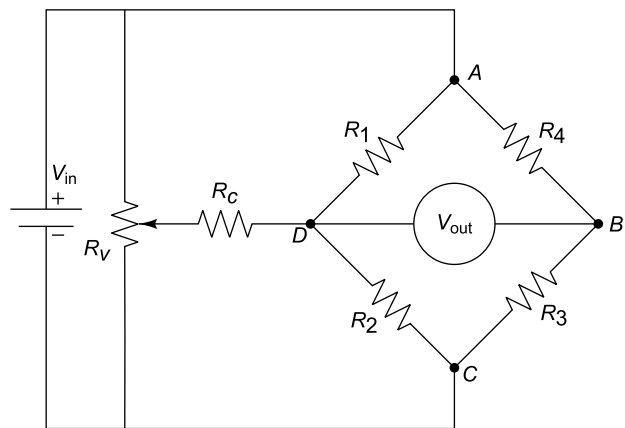

FIGURE 5.19 Wheatstone bridge with external zero-balance circuit

$$R_{R_2} = \frac{R_B \times R_2}{R_B + R_2} \tag{5.13c}$$

and

$$R_A = R_V f_{R_1} + \frac{R_c}{1 - f_{R_1}} \tag{5.13d}$$

$$R_B = R_V \left(1 - f_{R_1}\right) + \frac{R_c}{f_{R_1}} \tag{5.13e}$$

Here R_c is a large fixed (balance limit) resistor and R_V is a variable resistor (across the power supply). F_{R_1} is the fraction of R_V in the upper part.

In actual application, strain gauges may constitute the four arms of a bridge. Typically the four gauges are chosen such that they are 'matched', i.e. they are of identical resistance values. This helps to easily calculate the output voltage of the Wheatstone bridge, although theoretically it is not necessary that the two series circuits be identical. Besides, this choice allows the bridge circuit to self-compensate for temperature changes and drift in sensor output. The measurement of output voltage is used to calculate the exact impedance of the variable elements, which can be then be converted to strain. The bridge is initially balanced, or in other words its output is adjusted to a null or zero voltage state (by adjusting an impedance element). The change in resistance of one element will then induce an output voltage. For instance, if a strain gauge is subjected to a tensile load, its resistance will increase, resulting in a positive output voltage. Therefore, for stress analysis applications, it is the difference in the bridge output before and after the gauges are strained, that is of interest. When the initial resistances and the changes in resistances are known, Eq. (5.11) can be rewritten to give the difference in bridge output before and after strain is applied.

$$V_{\text{out}} = V_{\text{in}} \left[\frac{R_4 + \Delta R_4}{R_4 + \Delta R_4 + R_3 + \Delta R_3} - \frac{R_4}{R_4 + R_3} - \frac{R_1 + \Delta R_1}{R_1 + \Delta R_1 + R_2 + \Delta R_2} + \frac{R_1}{R_1 + R_2} \right] \tag{5.14}$$

where R_1, R_2, R_3 and R_4 are the initial resistances, prior to the application of strain and $\Delta R_1, \Delta R_2, \Delta R_3$ and ΔR_4 are the resistance changes experienced by the bridge arms when the strains are applied to them. It is to be noted that an increase in either R_4 or R_2 will produce a positive change (increase) in bridge output voltage; conversely an increase in R_1 or R_3 will produce a decrease in bridge output voltage. This leads to an important maxim of the Wheatstone bridge: changes of resistance in adjacent arms (R_4 and R_3, for example) have a numerically additive effect on the bridge output when the changes are of opposite signs. When the changes in adjacent arms are of the same sign, they have numerically subtractive effects.

The bridge output can also be expressed in terms of a constant current supplied to the bridge. With such an input, the output of the bridge can be found for any combination of input current and bridge arm resistances. For a bridge having a constant-current power supply, the bridge output can be calculated from the initial resistances and the strain induced resistance changes as shown in Eq. (5.15).

$$V_{\text{out}} = I_{\text{in}} \left(\frac{R_4 \times R_2 - R_1 \times R_3}{R_1 + R_2 + R_3 + R_4} \right) \tag{5.15}$$

When the Wheatstone bridge is used for reactive arms such as capacitors and inductances, the DC excitation voltage is replaced by an AC excitation source. A typical inductive bridge used for measurements

involving inductive accelerometers and LVDTs is shown in Fig. 5.20.

Different bridge configurations are used, depending on the quantities to be measured. They are:

- Quarter bridge—One arm of the bridge is an active element
- Half bridge—Two arms of the bridge are active elements
- Full bridge—All four arms of the bridge are active elements

In the quarter and half bridge cases, the arms of the bridge that are not made up of active elements are called dummies and consist of fixed elements that match the original unaltered values of the active elements. They are essentially used for completion of the bridge. In case the active gauge is placed in a high temperature environment, it is required that the dummies also be placed in a similar environment to take care of temperature compensation.

FIGURE 5.20   Inductive bridge circuit

The following are the errors that will have to be accounted for:

- *Random drift*   This is generally traceable to a poor strain gauge installation and is virtually impossible to correct after installation. This can be avoided by ensuring proper bonding between the structure and the gauge and avoiding leakage paths to the ground or between leads.
- *Strain averaging*   This is always present when the gauge is installed over a strain gradient. The steeper and more non-linear the gradient, the worse the error. This can be minimized by using a gauge with smaller grid, but it is never completely eliminated.
- *Thermal output*   This is a very dangerous and troublesome error because its magnitude is completely independent of strain. It can be eliminated using self temperature compensation techniques.
- *Gauge factor variation with temperature*   This is a very minor error in stress analysis and can usually be ignored.
- *Wheatstone bridge non-linearity*   This is present in all quarter-bridge circuits and some half- and full-bridge configurations. It amounts to a relatively minor error at strain levels normally encountered in experimental stress analysis, but must be carefully considered in high elongation measurements.
- *Transverse sensitivity*   This is present any time the gauge is installed in a strain field for which the ratio of the transverse and longitudinal strains is other than –0.285. It is relatively minor for measurements made with longitudinal gauges, but data from rosettes usually require numerical correction because one grid is exposed to a high transverse strain.
- *Lead wire effects*   This error is due to strain gauges being mounted at a distance from the measuring equipment and the problem of changing lead wire temperature. Any change in the lead wire resistance will be indistinguishable from changes in the resistance of the strain gauge itself. Special methods of wiring strain gauges to a bridge are employed to correct for lead wire effects to cancel at least a part of the effect. Typically if the lead-wire resistance exceeds 0.1% of the nominal gauge resistance, this source of error becomes significant. Therefore, in industrial applications, these errors are minimized by locating the transmitter directly at the sensor.
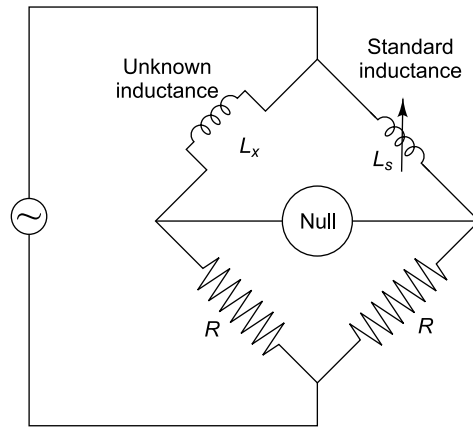
- *Non-linearity errors*   Non-linearity is absent only when the resistance changes are such that the currents in the bridge arms remain constant, that is, when $(\Delta R_1/R_1) + (\Delta R_2/R_2) = 0$ and $(\Delta R_4/R_4) + (\Delta R_3/R_3) = 0$. The only configurations producing complete linearity are bridges employing equal and opposite arms, i.e. two fully active arms and four fully active arms. The error due to non-linearity, when present, is generally small and becomes significant only for large values of measured strains.

5.7.1.2   *The Instrumentation Carrier Frequency Amplifier*   The instrumentation carrier frequency amplifier is widely used in instrumentation wherever mechanical signals from passive pickups of the resistive or inductive or capacitive type are used. Such transducers incorporate strain gauge bridges, differential transformers, capacitive bridges and thus the carrier frequency amplifier serves as the signal conditioning amplifier for the measurement of mechanical quantities such as stresses, torques, displacements, accelerations, forces and pressures. It can be used either for static strain or dynamic strain measurements or both or for dynamic measurements of accelerations, pressures, etc. A typical carrier frequency amplifier is shown in Fig. 5.21. It does the following functions: supplies AC excitation to the bridge, provides bridge completion for quarter- and half-bridge installations, zero balances the bridge, shunt calibrates the system of gauges, conditions and amplifies the signal from the bridge, sets gauge factor for direct readout of strain or sets gain for scaling output signals from the instrument. The oscillator produces a high frequency signal and this carrier signal serves as the AC excitation voltage for the transducer. Its frequency should be at least 10 times the highest frequency of the measured input (physical) signal. Carrier frequency amplifiers used in vibration measurement usually have a carrier frequency of 5 kHz. They can easily measure vibration frequencies in the range of 0–500 Hz.
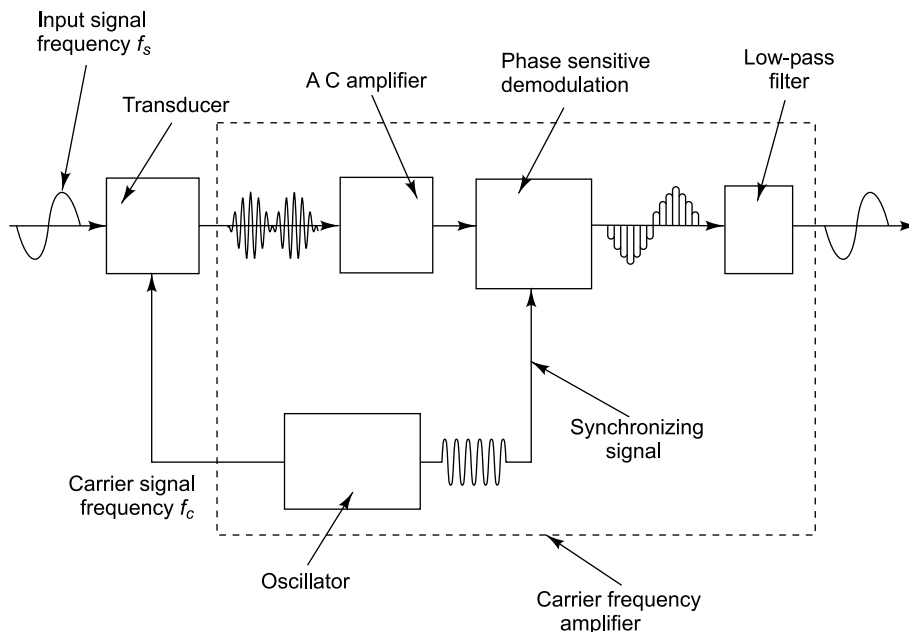
FIGURE 5.21   Instrumentation carrier frequency amplifier

The amplitude of the carrier signal is modified by the amplitude of the input signal or in other words, the carrier signal is amplitude modulated by the input signal. If the input signal is represented by the function $A_s \sin \omega_s t$ and the carrier signal by the function $A_c \sin \omega_c t$, then the modulated signal x_m is given by

$$x_m = (A_s \sin \omega_s t)(A_c \sin \omega_c t)$$

$$= \frac{A_s A_c}{2} \sin\left[(\omega_c - \omega_s)t + 90°\right] + \frac{A_s A_c}{2} \sin\left[(\omega_c + \omega_s)t - 90°\right] \quad (5.16)$$

The above equation shows that the modulated signal has two side bands corresponding to two frequencies $(\omega_c + \omega_s)$ and $(\omega_c - \omega_s)$. This amplitude modulated waveform is shown in Fig. 5.22. After amplification, the desired signal is extracted from the modulated signal by using a demodulator and a low pass filter. All dynamic signals undergo some attenuation. Errors are generally small at low frequencies, but can become very large if the frequency bandwidth of the carrier frequency amplifier is insufficient to accommodate the high frequencies present in the signal.

(a)

(b)

(c)
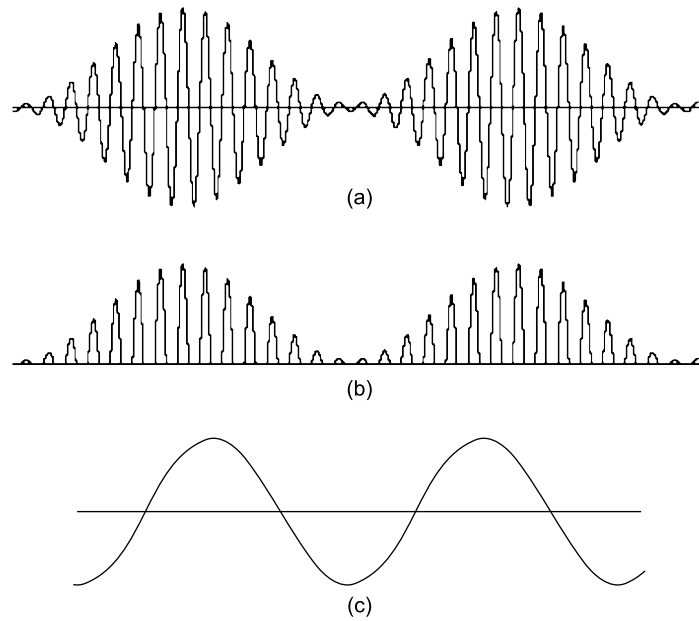
FIGURE 5.22   Amplitude modulation: (a) amplitude modulated signal, (b) rectified signal, (c) demodulated signal

Typical specifications of a carrier frequency amplifier

Carrier frequency	:	5 kHz ± 1%
Strain gauges	:	40–1200 Ω
Inductive transducer	:	2–20 mH
Usable frequency range	:	0–500 Hz

5.7.2 Charge Amplifiers

Piezoelectric crystals are extensively used in accelerometers, pressure sensors, load cells and other transducers. The output of a piezoelectric transducer is a charge signal in the range of some picoCoulombs with a very high impedance which is to be converted to a low impedance voltage so that it can be fed to a standard AC measuring, recording or analyzing equipment. The same holds good for a photodiode as well. This has led to the development of a special kind of a signal conditioning amplifier called a charge amplifier which is charge-to-voltage converter. It is essentially a device with an input capacitance, presenting a very high equivalent input impedance at low frequencies. All sensors with charge output are to be used with charge amplifiers. Besides they should be used with special low noise cables with length not exceeding 10 m. It is also preferable not to move the cable during measurement and to ensure that all connector nuts are tightened to obtain accurate results.

The purpose of a charge amplifier is to convert the high impedance charge (q) which is the output of the piezoelectric transducer (and the input to the charge amplifier) into a usable output voltage (with low output impedance). This amplifier offers some advantages over the usual voltage amplifiers for such applications. Its principle was first patented by W.P. Kistler in 1950 and it gained practical significance in the 1960s. A charge amplifier is usually constructed using operational amplifiers (op amps) with a capacitive feedback. It consists of a high-gain inverting voltage amplifier with a Metal Oxide Semiconductor Field Effect Transistor (MOSFET) or Junction gate Field Effect Transistor (JFET) at its input to achieve high input impedance. It thus acts in a manner similar to an integrator. Since the transducer acts as a differentiator, the transfer functions of the two cancel each other out and the output voltage across the feedback capacitor is proportional to the charge produced by the transducer. Stray capacitance at the input to the amplifier is not detrimental to its operation because this capacitance is always at a virtual ground. A block diagram of the input stage of a charge amplifier is shown in Fig. 5.23.
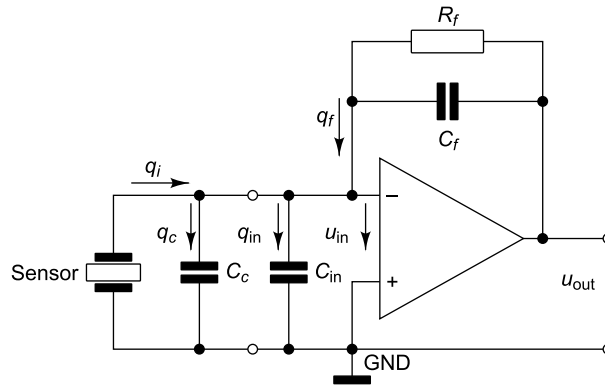

FIGURE 5.23   Input stage of charge amplifier

The input charge signal q_{in} is applied to the summing point (inverting input) of the amplifier stage and is balanced by the capacitive feedback circuit. The feedback voltage signal is thus proportional to the input charge. The charge equation at the input node is therefore:

$$q_i = q_c + q_{in} + q_f \tag{5.17}$$

Using the equation

$$u = qC \qquad (5.18)$$

and substituting for q_c, q_{in} and q_f in Eq. (5.17),

$$q_i = u_{in}.(C_c+C_{in}) + u_f.C_f \qquad (5.19)$$

where C_c is the cable capacitance, C_{in} is the amplifier input capacitance, C_f is the feedback capacitor and u_{in} is the voltage at the inverting input. Since the voltage difference between the inverting and the non-inverting inputs of a differential amplifier is zero under normal operating conditions, we can assume that the input voltage $u_{in} = 0$. Therefore

$$q_i = u_f.C_f \qquad (5.20)$$

Hence the output voltage u_{out} may be written as

$$u_{out} = u_f = \frac{q_i}{C_f} \qquad (5.21)$$

Thus it is seen that the output voltage of a charge amplifier depends only on the input charge and the feedback capacitance. The input and cable capacitances do not influence the output signal. This is a very big advantage when making measurements with cables of different lengths and types.

Two of the important considerations in the use of charge amplifiers are the time constant (TC) and drift. The time constant of an AC coupled circuit is defined as its discharge time. Such a circuit will cause a step input to decay to 37% of its original value in a time duration equivalent to one time constant. Time constant of a charge amplifier is $R_f C_f$ where R_f and C_f are the time constant resistor and capacitor respectively as shown in Fig. 5.23. Drift is defined as an undesirable change in output signal over a period of time; this change is not a function of the measured variable (input to the charge amplifier) and the user has no control over it. Drift in a charge amplifier may be due to low insulation resistance at its input or due to leakage current of the front end MOSFET or JFET. Both drift and time constant simultaneously affect a charge amplifier's output. Many commercially available charge amplifiers have selectable time constants which can be altered by changing the time constant resistor R_f depending on the desired frequency response. The time constant determines the lowest usable frequency; the longer the time constant, the better the low end frequency response and the longer the usable measuring time. When measuring events with wide pulse widths, the time constant should be at least 100 times the total event duration, failing which the DC component of the output signal will decay towards zero before the event is completed.

The advantages of a charge amplifier are:
 (i) It enables quasi-static measurements in certain situations, such as constant pressures on a piezoelectric transducer lasting several minutes.
 (ii) The output voltage is proportional to the charge produced by the piezoelectric transducer.
 (iii) The amount of charge present is not affected by the cable capacitance.
 (iv) The sensitivity and the time constant are independent of the capacitance of the crystal and also that of connecting cables. Gain is dependent only on the feedback capacitor, unlike ordinary voltage amplifiers.

There are certain disadvantages of charge amplifiers which may be present in certain applications. They are:
 (i) The SNR tends to be small.

(ii) The natural frequency of the transducer is reduced due to loss of stiffness caused by what amounts to a short circuit across the crystal.

5.7.3 Displacement Measuring Unit/Meter

This is a circuit used with capacitive transducers. These sensors are typically high impedance sensors, especially at low frequencies. The capacitance C of a parallel plate capacitor is $C = (\varepsilon A/d)$ where A is the common area between the plates, ε is dielectric constant and d is gap between the plates. A change in capacitance may be brought about by a change in any of the above three quantities and may be measured by using a reactive bridge with AC excitation. The reactance of a capacitance

C is $X_c = \dfrac{1}{i\omega C}$. All impedance quantities in the bridge must be complex, accounting for both magnitude and phase, both of which must be balanced to get null output voltage. Each series arm of such a bridge contains a capacitor and resistor. While both resistances are known, one of the two capacitances is the unknown capacitance being measured, while the other is a standard variable capacitor. Figure 5.24 shows a capacitive bridge.
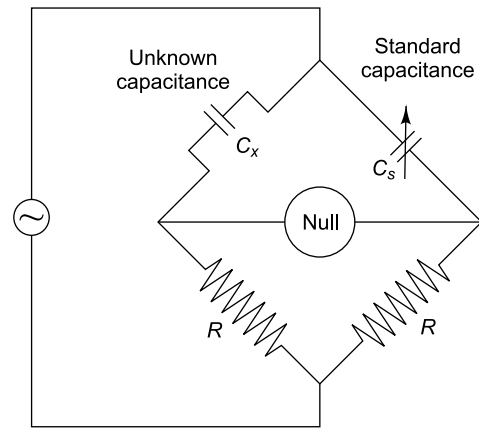
FIGURE 5.24 Capacitive bridge circuit

Another arrangement using a simple operational amplifier may be used as shown in Fig. 5.25.
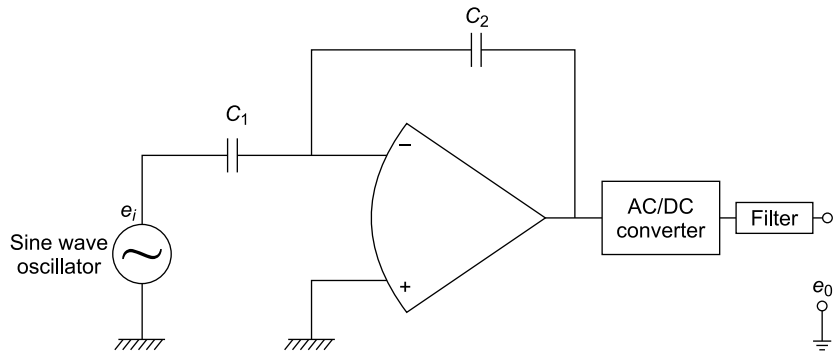

FIGURE 5.25 Displacement measuring circuit

The output voltage e_0 of the circuit may be written as

$$e_0 = \frac{-Z_2}{Z_1} e_i \tag{5.22}$$

where Z_1 and Z_2 are the impedances of the capacitors. Writing Z_1 and Z_2 in terms of capacitances C_1 and C_2 we have

$$e_0 = \frac{-C_1}{C_2} e_i \tag{5.23}$$

In many commonly used capacitive devices, it is the change in gap which leads to a change in capacitance. With capacitor $C_2 = \dfrac{\varepsilon_2 A_2}{d_2}$ as the transduction element operating on the principle of variation in gap and C_1 a fixed capacitor, we have

$$e_0 = \frac{-C_1 d_2}{\varepsilon_2 A_2} e_i \tag{5.24}$$

We get e_0 proportional to gap d_2. If d is the initial distance between the plates of the capacitor, and a displacement Δd is given to one of the plates, then a relation between the output voltage e_0 and the displacement Δd can be obtained. Thus, by measuring e_0, we can obtain the displacement Δd. In this case, one plate of the capacitor C_2 can be mounted directly on the machine and the other plate kept floating, so that the change in capacitance due to the vibration is picked up as output e_0.

Alternately, if the change in capacitance is brought about by a change in dielectric constant or the common area between plates, then capacitor C_2 may be made to represent the standard fixed capacitance and C_1 the capacitance due to the sensor in Fig. 5.25. It can easily be proved that in this case, e_0 is proportional to ε_1 or A_1.

5.7.4 Linear Power Amplifiers

A power amplifier is a circuit that is designed to amplify low-power signals in the range of a few μW or mW, so that they can drive loads such as small vibration exciters with the associated vibration heads, or loudspeakers. The power amplifier takes a signal from a source device, such as a function generator in the case of a vibration exciter, or from a low power audio-frequency source (such as a record player, CD player or cassette player), in the case of a loudspeaker, boosts it and makes it suitable for driving them. Ideally, the only difference between the input and the output signals is the power. All power amplifiers have a power supply, an input stage, and an output stage. The purpose of the input stage or front end of a power amplifier is to receive and prepare the input signals for amplification by the output stage. The output stage of the amplifier is the portion which actually converts the weak input signal into a much more powerful replica of the input which is capable of driving high power to the output device. This portion of the amplifier generally uses a number of 'power transistors' or Metal Oxide Semiconductor Field Effect Transistors (MOSFETs) and is responsible for generating most of the heat in the unit. It is the output stage of the amplifier which interfaces to the exciters/speakers.

There are two categories of power amplifiers that may be used in vibration testing: constant voltage (voltage mode negative feedback) amplifier and constant current (current mode negative feedback) amplifier, offering a highly linear output with selectable output impedance. The constant voltage amplifier produces an output voltage proportional to the input voltage, while a constant current amplifier produces an output current proportional to the input voltage. Constant voltage amplifiers are the most common and are preferred for general purpose applications, having a low output impedance (typically 0.01 Ω). For the kind of applications described in this book (vibration excitation), however, a constant current amplifier is advantageous, primarily when the exciter is used as a force generator or where non-feedback control is required using the mid-frequency range of the exciter. This demands a high impedance output (typically 3 kΩ). Constant current amplifiers in conjunction with exciters lead to distinct advantages such as the following:

(i) The force generated will always be directly proportional to, and in phase with the drive signal.

(ii) There will be no force drop-outs at structural resonances if the amplifier has adequate power rating.

(iii) The spectral content of the force generated will match that of the drive signal exactly.

Power amplifier modules for shaker applications are capable of controlling all connected electrodynamic shakers. Most of the commercially available amplifiers also provide continuous monitoring and metering, having displays showing RMS values of current and voltage. The waveforms can also be monitored by taking leads from the output sockets to oscilloscopes. There is also a display indicating any clipping of output voltage caused by too high a driving voltage, or output current peaks and any power device failures. A safety management system is also provided to monitor functions such as temperature, excessive current and excessive travel of exciter head and avoids the destruction of the amplifier in case of short circuit. Since these amplifiers come with selectable ranges of operating voltages and currents, they can be readily adapted to shakers from a wide range of manufacturers.

The frequency response of an exciter driven by a constant current will show three different regions. The first two regions are characterized by the spring-mass system consisting of the moving element and its spring with a resonance of around 20 Hz. In the third region, axial resonances of the moving element will occur; this is generally above 3 kHz for big exciters and sets the upper operational frequency limit of the exciter. A response curve for an exciter with a constant voltage input will show the same three regions of control, but since the lower resonance is considerably damped, an easier control of the level is achieved.

Analogue audio amplifiers with a rated sinusoidal power output up to 2000 VA are available in the market. The Class of an amplifier refers to the design of the circuitry within the amplifier. For audio amplifiers, the common Classes are Class A, B, AB, C and D. There are also a number of other classes of amplifiers, such as G, H, S, etc. Most of these amplifiers have a large frequency range, from DC to 10s or 100s of kHz. The output signals of all amplifiers contain additional (unwanted) components that are not present in the input signal, known as harmonic distortion. In addition to distortion, all amplifiers generate a certain amount of noise. A high SNR and a low distortion factor are desirable features.

Typical specifications

Output voltage	:	10 V RMS from DC to 10 kHz
Output current	:	10 A DC
	:	2 A for 40 Hz–10 kHz
Frequency range	:	40 Hz–10 kHz
Frequency response	:	DC to 10 kHz < ± 0.5 dB
		AC input 15 Hz to 100 kHz < ±3 dB
Input impedance	:	>10 kΩ
Output impedance	:	>50 Ω
Gain at 1 kHz in low impedance (voltage output) mode	:	5 V/V ± 1 dB
In high impedance (current output) mode	:	14 A/V ± 2 dB
Harmonic distortion	:	<0.5%

FURTHER READINGS

1. Beeby, S.P., Beeby, S., Ensel, G., Kraft M. and White, N.M., *MEMS Mechanical Sensors,* Artech House Inc., eprint 2004.
2. Buzdugan, Gh., Mihailescu, E. and Rades, M., *Vibration Measurement (Mechanics: Dynamical Systems)*, Springer, Netherland, 1986.
3. Collacott, R.A., *Vibration Monitoring and Diagnosis*, George Godwin Ltd., London, 1979.
4. De Silva, C.W., *Mechatronics: An Integrated Approach*, CRC Press, Boca Raton, FL, 2005.
5. De Silva, C.W., *Vibration Monitoring, Testing, and Instrumentation*, CRC Press, Boca Raton, FL, 2007.
6. Doebelin, E.O., *Measurement Systems: Application and Design*, McGraw-Hill Professional, Boston, 2004.
7. Dyer, S.A., *Survey of Instrumentation and Measurement*, John Wiley & Sons, Canada, 2001.
8. Fraden, J., *Handbook of Modern Sensors*, Springer-Verlag, New York, 1996.
9. Gatti, P. and Ferrari, V., *Applied Structural and Mechanical Vibrations—Theory, Methods and Measuring Instrumentation* (eBook), Taylor & Francis, 1999.
10. He, J. and Fu, Z.-F., *Modal Analysis*, Butterworth-Heinemann, Oxford, 2001.
11. Khazan, A.D., *Transducers and Their Elements*, PTR Prentice Hall, Englewood Cliffs, NJ, 1994.
12. McConnell, K.G. and Varoto, P.S., *Vibration Testing: Theory and Practice*, Wiley-IEEE, New York, 1995.
13. Measurements Group Inc., *Experimental Stress Analysis, Notebook*, Raleigh, NC 27611.
14. Murty, D.V.S., *Transducers and Instrumentation*, Prentice Hall of India Pvt. Ltd., New Delhi, 2004.
15. Nakra, B.C. and Chaudhry, K.K., *Instrumentation Measurement and Analysis*, Tata McGraw-Hill, New Delhi, 2004.
16. Parr, E.A., *Industrial Control Handbook*, Industrial Press Inc., New York, 1998.
17. Piersol, A.G., *Harris' Shock and Vibration Handbook*, McGraw-Hill Professional, eBook, 2002.
18. Prudenziati, M., *Handbook of Sensors*, Vol. 1, Sec. IV, pp. 189–206, Elsevier Science Inc., New York, 1994.
19. Rangan, C.S, Sarma, G.R. and Mani, V.S.V., *Instrumentation Devices and Systems*, Tata McGraw Hill Publishing Company Ltd., New Delhi, 1983.
20. Tse, F.S. and Morse, I.E., *Measurement and Instrumentation in Engineering: Principles and Basic Laboratory Experiments*, Marcel Dekker, New York, 1989.
21. Window, A.L., *Strain Gauge Technology*, Springer, Berlin, 1992.
22. Webster, J.G., *Mechanical Variables Measurement: Solid, Fluid, and Thermal*, CRC Press, Boca Raton, FL, 2000.
23. http://www.bksv.com
24. http://www.hbm.com
25. http://www.kavilco.com
26. http://www.omega.com
27. http://www.pcb.com

28. http://www.sensorland.com
29. http://www.senso-metrics.com
30. http://www.sensorsmag.com/articles/0103/19/main.shtml
31. http://www.sensorsportal.com
32. http://sensors-transducers.globalspec.com

Chapter

6

Fundamentals of Signal Analysis

6.1 Introduction

The overall magnitude of the vibration signal which is obtained using any one of the transducers we have discussed in Chapter 3 is not sufficient, but often spectral amplitudes or statistical descriptors are required. Historically, people have studied experimental data in the time domain using oscilloscopes, strip chart recorders, etc. The observer then had to rely on his or her powers of observation to make interpretations from the time domain data and this was difficult. For diagnostics, it is important to have information regarding the salient frequency components of the vibration signal. Simple $1 \times$ rpm information can be got using tachometers or stroboscopes or eddy current probes responding to key phasor information. However, for comprehensive information on the spectral content of the signal being measured, one needs more sophisticated instruments. The advances in electronics and signal processing have led to the development of very efficient real time analysers/spectrum analysers/signal analysers/fast Fourier transform (FFT) analysers, which give frequency information. Using these analysers, complex vibration waveforms, often referred to as vibration signatures, can be broken down into their spectral components, and defined as a sum of harmonic functions of discrete amplitudes, frequencies and phases.

This chapter throws light on several aspects of signal analysis which involves representation, transformation and manipulation of signals: for example, separation of two signals, i.e. signal from a combination of signal plus noise, changing from one domain to another, say from time to frequency domain or vice versa, or obtaining statistical parameters from a signal. Signal processing is extensively used these days in machinery diagnostics and in applications in acoustics including the study of jet noise, traffic noise, machinery noise and sound navigation and ranging (SONAR), underwater acoustics and instrumentation. This chapter also discusses the issues to be dealt with while carrying out dynamic measurements using a signal analyser, whether it is dedicated or personal computer (PC) based. The various terms used in signal processing have been dealt with, without too much mathematical encumbrance.

6.2 Various Steps in Data Acquisition and Processing

Most signals in raw form are continuous-time signals, but all subsequent analysis is done on digital signals due to extensive availability of digital signal processing software/hardware. The first step in any measurement is data collection and for this a transducer which converts the physical/mechanical quantity

to a proportional electrical quantity is used. The conversion may happen in one stage as in the case of a thermocouple, where a change in temperature is converted to a proportional change in voltage; or it may be done in many stages as in the case of an inductive vibrometer, where the change in displacement is converted first to a change in inductance, and subsequently to a change in voltage using a Wheatstone's bridge arrangement. The next stage is data recording, which may be done using either a dedicated signal recording device such as an instrumentation tape recorder or a digital recorder or more commonly by PC based recording. The third stage involves editing the data, converting it to the required format and pre-processing it before doing any actual processing to extract required signal descriptors. This is followed by data qualification since the nature of the signal dictates the analysis techniques to be used. Before analysing a signal, it is required to see if the signal is harmonic or periodic or random. If it is random, checks to see if it obeys stationarity can be done by observing if all data records have same mean, RMS, etc. (for more information refer to Chapter 7). The last and final step is data analysis, which involves working on individual records and on multiple records. Depending on the requirement, one might want to obtain the spectrum/cepstrum/probability density histogram/statistical averages, etc. For many of the present day applications, real-time operation is often desirable, i.e. the output has to be computed at the same rate at which the continuous time signal is sampled.

6.2.1 Data Acquisition Systems

Data acquisition systems, as the name implies, are used to collect data, record them and analyse them. They have a wide range of measurement applications, including collection, characterization, monitoring and control of data. The data acquisition process has become considerably simplified these days through advances in electronics and has become more accurate, versatile and reliable. Data acquisition systems typically acquire measured physical parameters such as displacement, velocity, acceleration, temperature, pressure, flow or sound pressure level or take a specific action based on the data received. A wide range of data acquisition components and solutions are available in the market, ranging from simple recorders to PC plug-in cards to sophisticated mainframe computer systems. The specific requirements of any application will dictate the resolution, accuracy, number of channels and speed requirements for a data acquisition system. Such a system can also include a multiplexer, which is used for signal switching in multichannel applications. Generally, in the multiplexed mode, only one channel is connected at a time to the measuring device, with break-before-make switching (i.e. the input is disconnected before a new input is connected). Figure 6.1 shows a typical data acquisition and processing system for a system identification application.
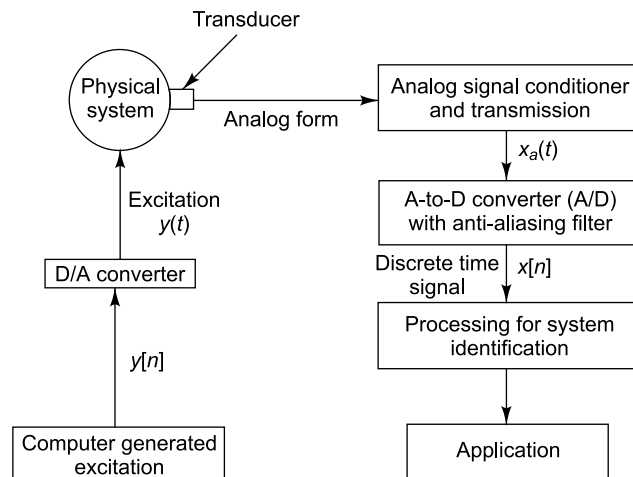

FIGURE 6.1 Data acquisition and processing

6.2.2   Analogue and Digital Signals

While most of the data obtained using transducers is generally analogue, the processing on these signals is done after digitizing them as mentioned earlier. Hence, it is required to know what these signal forms imply. A signal $x(t)$ is said to be an analogue signal if it is continuous along both x- and y-axis. Here the x-axis is typically time and the y-axis could represent any mechanical or electrical quantity. A signal is said to be a discrete-time signal, if it is discrete along the x-axis only and continuous along the y-axis. In order to convert a signal from continuous time to discrete-time, a process called sampling is used; this involves measuring the value of the signal at regular intervals of time and each value is referred to as a sample. The discrete-time signal is thus related to the analogue signal from which it was obtained by sampling. The original analogue signal can be reconstructed from its discrete-time counterpart if the sampling was done in the first place without aliasing (as discussed in Section 6.6) and if the sampling frequency is known. A digital signal is one that is discrete along both the x- and y-axes and is generally called a sequence, since it depicts on the x-axis sample number, which is a dimensionless quantity. Figure 6.2 shows all three kinds of signals. In a digital signal, the variable on the x-axis is made discrete by sampling, and that on the y-axis by a process called quantization (discretization of amplitude). The notation $\{x(n)\}$ is used to represent a complete sequence and $x(n)$ the nth sample of the sequence; however the latter is often used to denote the sequence $\{x(n)\}$ itself. Typical sequence operations are addition, multiplication, shifting (delaying/advancing), etc.
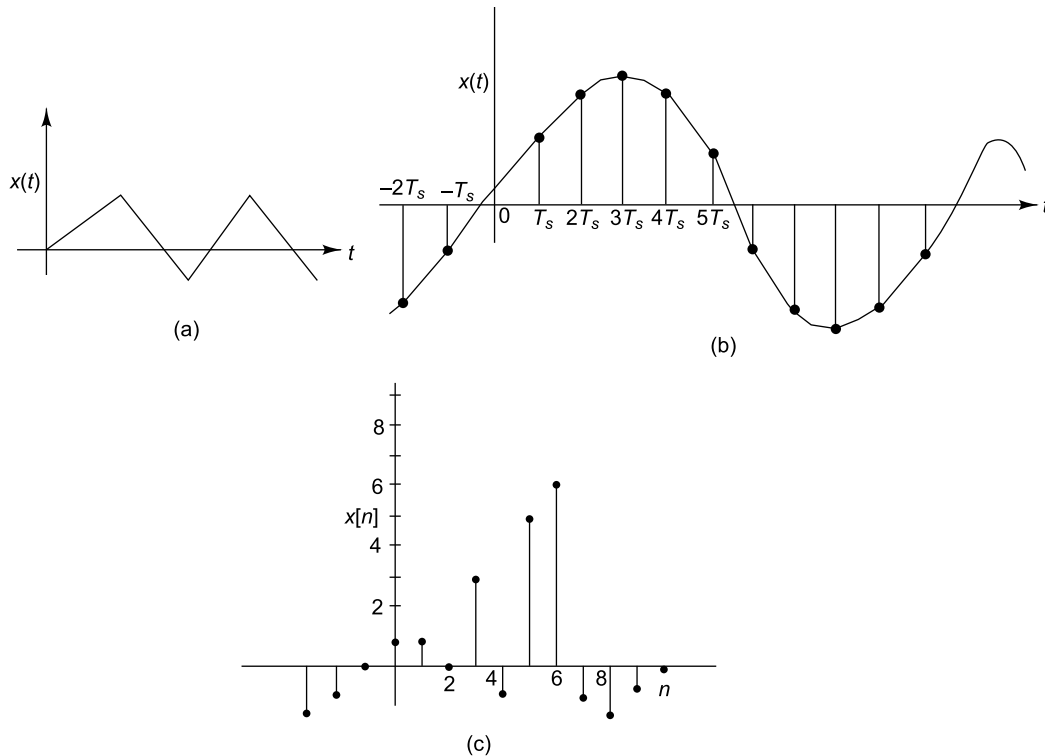

FIGURE 6.2   Various types of signals: (a) analogue signal, (b) discrete-time signal, (c) digital signal

6.2.3   Analogue-to-Digital Converter (A/D Converter or ADC)

The ADC is the heart of most data acquisition systems and is an electronic device that converts analogue voltages acquired from transducers into their equivalent digital form (in binary or hexadecimal code) representing the quantized amplitude values closest to the actual values of the inputs. There are several types of A/D conversion techniques; they can generally be divided into two types: integrating and non-integrating; the former measures the average input value over a defined time interval, thereby rejecting many types of noise, while the latter samples the instantaneous value of the input (including noise) during a very short time interval and converts it to digital form. One of the simplest forms of ADC is one with uniform sampling, i.e. sampling the analogue signal $x_c(t)$ at regular intervals of time with a clock rate T_s or with sampling frequency $f_s = 1/T_s$ as shown in Fig. 6.3(a). The output of the sample-and-hold circuit $x_o(t)$ is a staircase waveform and $x[n]$ is a digital signal as shown in Fig. 6.3(b). The sample-and-hold circuit is designed to sample $x_c(t)$ as instantaneously as possible and to hold the sample value as nearly constant as possible till the next sample is taken. Typical issues involved in analogue-to-digital
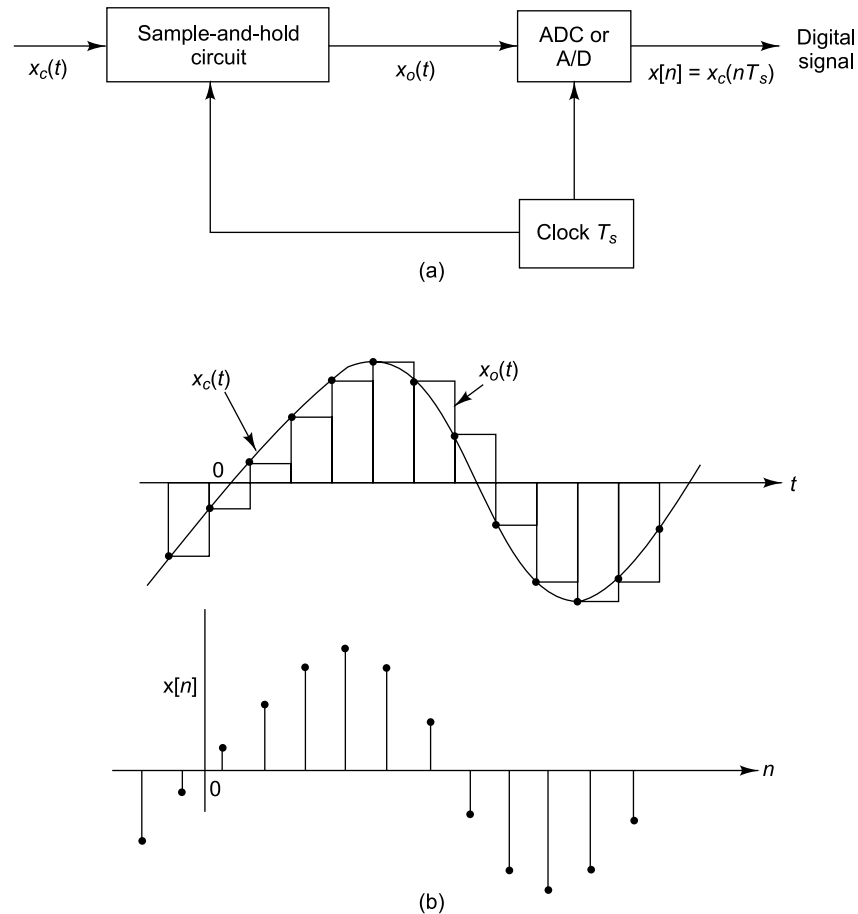

FIGURE 6.3   Analogue-to-digital conversion: (a) schematic, (b) typical staircase waveform and digital signal

conversion are the number of bits to be used, which has a direct bearing on the quantization error, the need for sample-and-hold circuits, the required sampling frequency, linearity, etc. and these are discussed in the subsequent sections. The number of bits is decided by the least significant bit (LSB), which is nothing but the percentage of full-scale represented by 1 bit; for example in a 10 bit ADC with a full-scale voltage of 10 V, 1 LSB corresponds to $\dfrac{1}{2^{10}} \times 10\,\text{V} = \dfrac{1}{1024} \times 10\,\text{V} = 0.00976$ V or 9.76 mV. The sampling frequency is decided by the highest frequency in the signal. Too low a sampling frequency results in what is called aliasing, while too high a sampling frequency involves complicated circuitry. Sampling is generally not invertible, i.e. given the output $x[n]$, it is not possible, in general, to reconstruct $x_c(t)$, the input to the sampler, since many continuous-time signals can produce the same output sequence of samples.

6.2.3.1 *Common terms in ADC or DAC parlance*   Some common terms pertaining to the performance of an ADC or DAC are shown further.

Resolution: This is defined as the fraction of the full-scale range represented by the smallest signal increment that can be detected by a data acquisition system, e.g. 0.0244% of full-scale for a 12 bit ADC. It can also be expressed in bits, e.g. as a system having 12-bit resolution or in proportions, i.e. one part in 4,096.

Sampling rate: This represents the speed at which a data acquisition system collects data and is expressed in samples per second. For multichannel data acquisition systems, the sample rate is typically the speed of the ADC for all channels together. To obtain individual channel sample rates, the speed of the ADC has to be divided by the number of channels being sampled.

Accuracy: This is the difference between the expected and measured output voltages in terms of the change caused in the LSB during conversion. Most converters have an accuracy of at least ±1 LSB.

Linearity: This represents how closely the transfer function from input voltage to binary value or its inverse represents a linear function and is usually specified to ±1 LSB.

Monotonicity: A DAC is said to be non-monotonic if there is a momentary reversal in the expected direction of change.

Zero offset: The zero offset represents the output voltage of a DAC for an input code of zero. It is generally less than 1/2 LSB for unipolar DACs, but bipolar DACs can have considerable zero offset.

Stability: This indicates how constant the full-scale output is with respect to age and variations in temperature and power supply.

Settling time: This is an indication of the time taken for the output signal to settle to within 1/2 LSB of its final value after a given change in input scale (usually full-scale).

Glitches: These are transients appearing at the output when new digital data are applied to a DAC.

Single-ended (SE) input: Data acquisition devices have either SE inputs or differential inputs or both, and refer to the way a signal is wired to a data acquisition device. In SE wiring, each analogue input has a unique high connection, but all channels share a common ground connection.

Differential input: In a differential input configuration, each signal channel has a unique high and unique low connection.

Communication options: One might also want to look into the several different communication options available in ADCs or DACs. Serial communication systems are preferred when measurements have to be made far away from the computer and long cables are required. RS232 is the most common serial

communication option, but has the drawback that it supports transmission distances up to 15 m and SE inputs only; besides it has a very low data rate, up to 20 kbits/s only. For communication over longer distances or at higher bit rates, SE methods are unsuitable. Differential data signals nullify the effect of induced noise signals and ground shifts and thus offer better performance for most applications. RS422 comes under this category of communication devices and is meant for higher bit rates and transmission distances than RS232. However, it is RS485 which is supreme for truly multipoint communication networks, catering to a maximum of 32 drivers and receivers on a single (2-wire) bus. RS485 is superior to RS232 and supports transmission distances up to 1.5 km. The universal serial bus (USB) is ideal for data acquisition applications, offering several advantages over conventional series and parallel connections, including higher bandwidth (up to 12 Mbits/s) and the ability to provide power to the peripheral device, making only one cable sufficient to link the data acquisition device to the PC. Computer data acquisition boards plug directly into the computer bus and hence are fast and inexpensive. The standard parallel port on a computer can also be used to connect to a data acquisition device. Parallel port systems often support very high sample rates, but the distance between the computer and the data acquisition device is limited to a few metres.

6.2.3.2   *Tracking ADC* The tracking ADC (Fig. 6.4) is a simple form of A/D converter. Let $x_c(t)$ be the analogue signal to be converted to digital form. This analogue signal and a reference signal x_r, form the inputs to a comparator. This reference signal is obtained as the output of a DAC by triggering a binary counter. The output of the counter is converted to analogue form by the D/A converter to produce the reference voltage given by

$$x_r = \left[\left(\frac{a_1}{2}\right) + \left(\frac{a_2}{2^2}\right) + \ldots + \left(\frac{a_n}{2^n}\right)\right] x_{FS}, \tag{6.1}$$

where x_{FS} is the full-scale output of the DAC, a_1 is the most significant bit (MSB) corresponding to $(x_{FS}/2)$, and a_n is the LSB corresponding to $(x_{FS}/2^n)$.

The comparator operates such that its output $x_d = 1$ for $x_c > x_r$, and $x_d = 0$ for $x_c < x_r$. The counter counts up as long as it receives 1, i.e. as long as $x_c > x_r$ and the output of the DAC then keeps increasing with increase in digital count of the counter. When the DAC output voltage exceeds the analogue input, i.e. when $x_r > x_c$, the output of the comparator becomes 0, and the counter stops counting, causing the DAC to reveal the digital code corresponding to the analogue input.

As an example, consider a 10-bit ADC with a full-scale voltage of 10 V. Let us say we would like to find the digital code corresponding to an analogue input signal of 0.1 V or
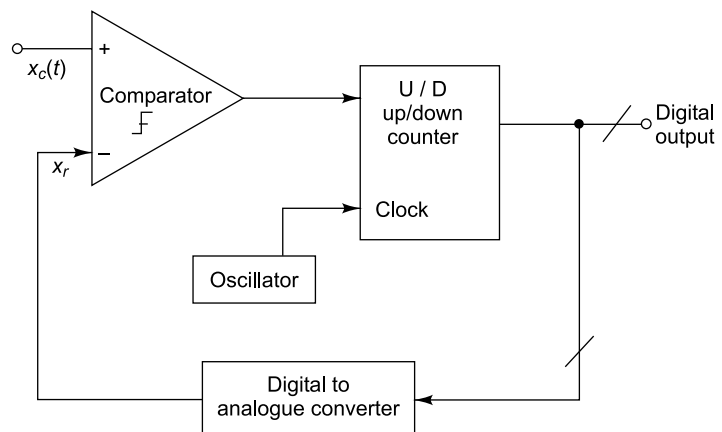

Figure 6.4   Tracking ADC

100 mV. The voltage corresponding to the LSB for this converter would be $(1/2^{10}) \times 10$ V = 9.76 mV. When the counter takes on a count of 1011, the corresponding reference input to the comparator becomes 9.76×11 = 107.426 mV, which is greater than the input voltage x_c of 100 mV. The output of the comparator becomes 0, causing the counter to stop counting. The output of the converter for 0.1 V input would thus be 1011. The quantization error is $9.76 \times 11 - 100$ mV = 7.426 mV. It may be expected that for a 12 bit ADC, we will get a smaller value of quantization noise. Table 6.1 gives an indication of the resolution obtained with different number of bits in an ADC.

TABLE 6.1   Number of bits in ADC and resolution

No. of bits (n)	2^n	1 bit with 10 V FS	Resolution (%)
8	256	39 mV	0.390625
10	1024	9.76 mV	0.0976
12	4096	2.4 mV	0.024414
16	65,536	150 μV	0.001526
32	131,072	76.2 μV	0.000763

Statistical evaluation of quantization error: The quantization error x may be assumed to have a uniform distribution (for more details on probability distribution, see Section 7.3), the actual value of the quantized signal lying within ±0.5 LSB of the analogue input. Let the probability density function of the error be denoted as

$$p(x) = 1/\text{LSB}, \quad -0.5 \text{ LSB} < x < 0.5 \text{ LSB} \qquad (6.2)$$
$$= 0, \text{ otherwise}$$

$$\text{Variance of the error } \sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \, p(x) \, dx \qquad (6.3a)$$

$$= \frac{1}{\text{LSB}} \int_{-0.5\text{LSB}}^{0.5\text{LSB}} x^2 \, dx = \frac{\text{LSB}^2}{12} \qquad (6.3b)$$

$$\text{Standard deviation of error} = \sigma_x = \frac{\text{LSB}}{\sqrt{12}} = 0.29 \, \text{LSB} \qquad (6.3c)$$

Let us assume that the random vibration of a structure is to be measured with an accelerometer and that the vibration record is to be converted to a digital format for analysis over the frequency range 0–2 kHz with a signal-to-noise ratio (SNR) of at least 80 dB. Let us find the sampling rate and the number of bits/data point required in the A/D conversion. A sampling frequency greater than twice the highest frequency of interest would be required (this is discussed in Section 6.6). Hence, a sampling frequency = 2½ × 2 = 5 kHz may be taken, giving rise to a sampling rate of (1/5000) = 0.2 ms. It is to be noted here that SNR is specified in terms of dB which is defined as $20 \log_{10} (A / A_{\text{ref}}) = 10 \log_{10} (P / P_{\text{ref}})$ where A and A_{ref} are amplitude-like quantities and P and P_{ref} are power-like quantities.

For a minimum SNR of 80 dB, $\dfrac{S}{N} = \dfrac{2^n \times \text{LSB}}{0.29 \times \text{LSB}} = \dfrac{2^n}{0.29}$. Hence, $20 \log_{10} \left(\dfrac{2^n}{0.29} \right) \geq 80$ dB implies

that $\dfrac{2^n}{0.29} \geq 10{,}000$ or $2^n \geq 2900$. This implies that the number of bits n should be at least 12.

6.2.3.3 *Successive approximation ADC*   The successive approximation ADC is much faster than the tracking ADC because it uses digital logic to converge on the value closest to the input voltage, starting with the most significant bit (MSB) and proceeding to the LSB. As in the tracking ADC, a comparator and a DAC are used for the process. Figure 6.5(a) shows a flow chart of the conversion process. Figure 6.5(b) illustrates schematically a 4-bit successive approximation ADC.
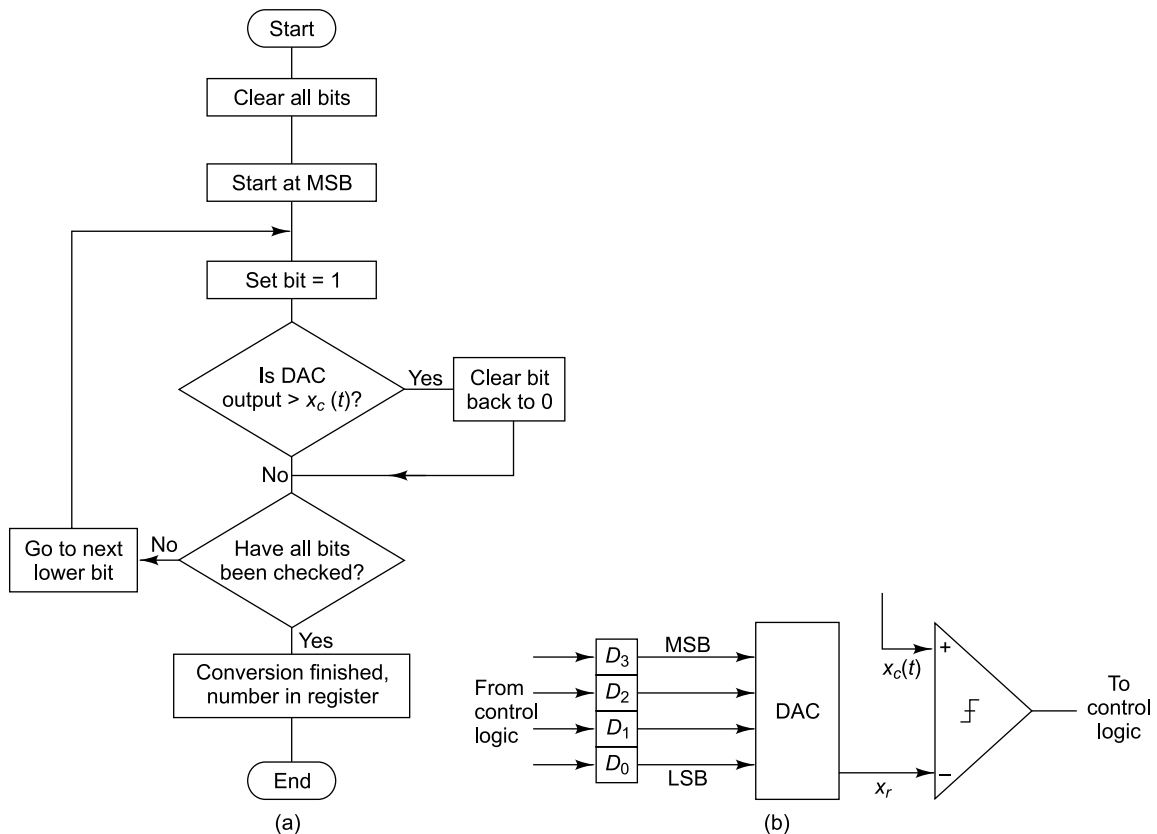

FIGURE 6.5   Illustration of successive approximation ADC: (a) flow chart, (b) 4-bit ADC

6.2.4 Digital-to-Analogue Conversion

Some data acquisition systems contain a digital to analogue converter or D/A converter (DAC) which does the opposite job of an ADC. This device converts binary numbers into analogue DC voltages or currents on receipt of command signals from the control hardware. D/A converters are used where precise control of external devices is required through an analogue voltage. Thus the voltage or current from a D/A converter can be used as excitation input to a shaker. In its simplest form, the DAC consists of a network of resistors and switches which provide currents proportional to the weight of each bit, multiplied by its binary value in the parallel digital word. The digital data source is sampled and held in an input register in parallel digital form to serve as a steady input to the DAC. The 0's and 1's are converted to clean digital form using gates which are either on or off depending on the incoming signal.

This binary signal can then be converted to an analogue signal using a summing amplifier. The main drawback of this method is that very high precision of the summing resistors is required, and this is especially unsatisfactory for a DAC with a large number of bits. A network with the so called R-2R ladder of resistances is used to overcome this problem. The DAC with a summing current to voltage converter uses an operational amplifier to sum the currents from all the bits and convert them to voltages as shown in Fig. 6.6 for a 4-bit DAC. This summing amplifier produces the output

$$V_{\text{out}} = V_{\text{ref}} \left[\frac{D_0}{2^4} + \frac{D_1}{2^3} + \frac{D_2}{2^2} + \frac{D_3}{2^1} \right] \frac{R_f}{R} \qquad (6.4)$$

In the above equation D_0 represents the LSB and D_3 the MSB as before and the D's take on the values 0 or 1. The digital inputs close the switches for a logical 1 and keep it grounded for a logical 0. This can be extended to a DAC with any number of bits using an appropriate R-2R ladder. Offset and gain controls are also generally incorporated in the circuit.
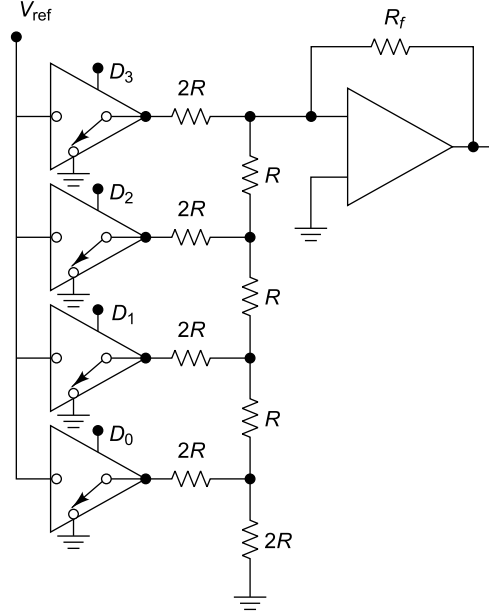
FIGURE 6.6 R-2R Summing amplifier

6.3 SOME IMPORTANT SIGNAL OPERATIONS

The signal operations described here are amplitude scaling, translation or time shifting, time scaling and time inversion/folding. We may need these operations in the course of signal analysis. Though the independent variable considered here and in subsequent sections is time, these operations hold good for signals which are functions of other independent variables such as space or frequency.

6.3.1 Amplitude Scaling

This is typically the kind of operation that takes place in a voltage amplifier. A signal $\phi(t)$ is said to be an amplitude scaled version of a signal $f(t)$ if

$$\phi(t) = k f(t) \qquad (6.5)$$

Figure 6.7(a) shows a signal $f(t)$ and Figs. 6.7(b) and (c) show two different cases of amplitude scaling.

FIGURE 6.7 Amplitude scaling: (a) $f(t)$, (b) $\phi(t) = 2f(t)$, (c) $\phi(t) = -f(t)$

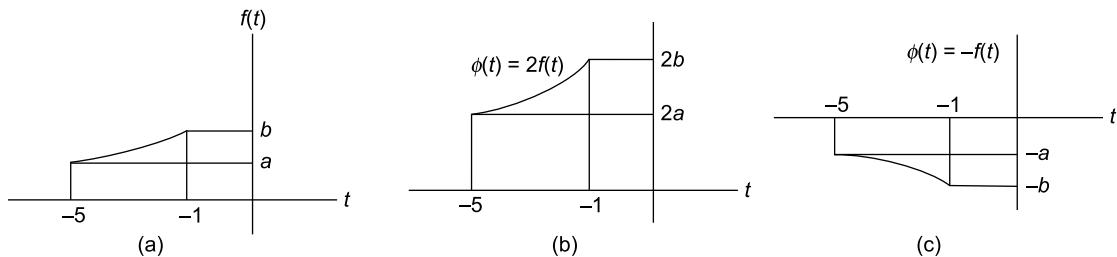## 6.3.2   Translation/Time Shifting

This happens with delaying or advancing operations. Such operations are required while computing auto- or cross-correlations of signals. A signal $\phi(t)$ is said to be a translated version of $f(t)$ if it takes on one of the following forms. We assume that t_0 is positive in these equations.

(i) $\phi(t)$ is said to be a delayed version of $f(t)$ if

$$\phi(t) = f(t - t_0) \tag{6.6a}$$

The signal $f(t)$ gets shifted to the right by t_0 in this case as seen from Figs 6.8(a) and (b).

(ii) $\phi(t)$ is said to be an advanced version of $f(t)$ if

$$\phi(t) = f(t + t_0) \tag{6.6b}$$

In this case, the signal $f(t)$ gets shifted to the left-hand side by t_0 as seen from Fig. 6.8(c).
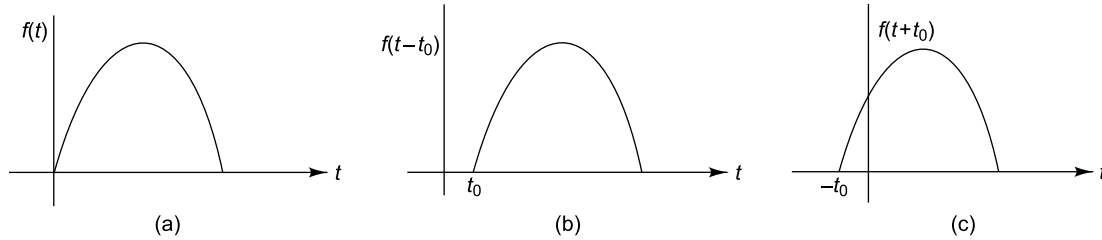
FIGURE 6.8   Time shifting: (a) $f(t)$, (b) $\phi(t) = f(t - t_0)$, (c) $\phi(t) = f(t + t_0)$

6.3.3   Time Scaling

A signal is said to be time scaled if it undergoes compression or expansion on the time axis. This is the operation that takes place when we record a signal at one speed and replay at another. Frequency scaling is done in filters like 1/1 and 1/3 octave filters and in wavelet transforms.

6.3.3.1   Expansion   When we record a signal at some speed and replay at a much lower speed, as when we view a recorded picture in slow motion, we are, in effect, expanding the signal in the time scale. $\phi(t)$ is said to be an expanded version of $f(t)$ if the following equation is satisfied.

$$\phi(t) = f\left(\frac{t}{a}\right) \tag{6.7a}$$

Here a is a positive constant.

6.3.3.2   Compression $\phi(t)$ is said to be a compressed version of $f(t)$ if

$$\phi(t) = f(at) \tag{6.7b}$$

where $a > 1$ as before.

Figures 6.9(a)–(c) show $f(t)$, $f(2t)$ and $f\left(\dfrac{t}{2}\right)$, respectively.

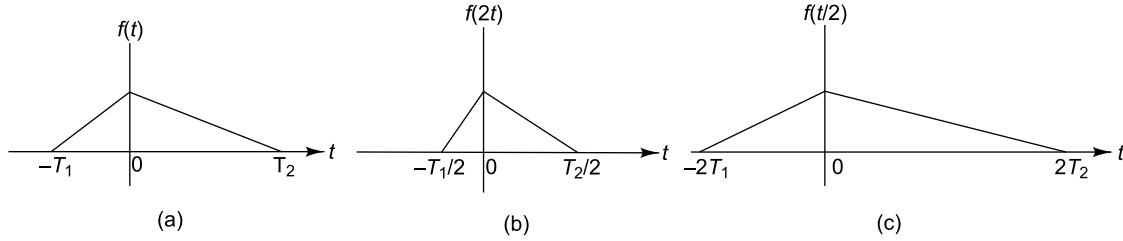**FIGURE 6.9**   Time scaling: (a) $f(t)$, (b) $f(2t)$, (c) $f\left(\dfrac{t}{2}\right)$

6.3.4   Time Inversion/Reversal/Folding/Flipping

This happens when we replay a tape in the direction opposite to that in which it was recorded. A signal $\phi(t)$ is said to be a time-reversed version of $f(t)$ if

$$\phi(t) = f(-t) \tag{6.8}$$

Figures 6.10(a) and (b) show the original and time-reversed versions of $f(t)$. The latter is seen to be a mirror image of the former about the axis $t = 0$ or is said to be a folded or flipped version.
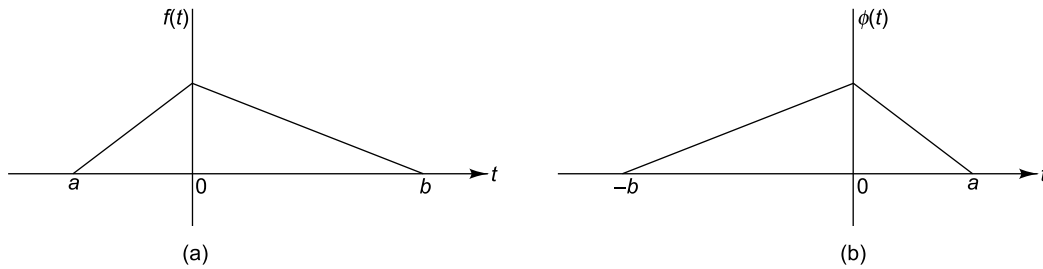
FIGURE 6.10   Time inversion: (a) $f(t)$, (b) $\phi(t) = f(-t)$

6.3.5   Even and Odd Parts of a Signal

Examples of even and odd signals are the cosine and sine waves, respectively, plotted as a function of time. The former is so called because it is symmetric about the y-axis, while the latter is antisymmetric. The even part $x_e(t)$ of a real signal $x(t)$ is such that

$$x_e(t) = x_e(-t) \tag{6.9a}$$

$$x_e(t) = \frac{x(t) + x(-t)}{2} \tag{6.9b}$$

The odd part $x_0(t)$ of a signal $x(t)$ is such that

$$x_0(t) = -x_0(-t) \tag{6.10a}$$

$$x_0(t) = \frac{x(t) - x(-t)}{2} \tag{6.10b}$$

Figure 6.11(a)–(c) shows $x(t)$, $x_e(t)$ and $x_0(t)$, respectively.

FIGURE 6.11   Even and odd parts of a signal: (a) $x(t)$, (b) $x_e(t)$, (c) $x_0(t)$

For a complex signal $x(t)$, we have the conjugate symmetric and conjugate antisymmetric parts which are the even and odd counterparts of a real signal. The conjugate symmetric part of a complex signal is defined as

$$x_e(t) = \frac{x(t) + x^*(-t)}{2} \tag{6.11a}$$

and
$$x_e(t) = x_e^*(-t) \tag{6.11b}$$

The conjugate antisymmetric part is defined as

$$x_0(t) = \frac{x(t) - x^*(-t)}{2} \tag{6.12a}$$

and
$$x_0(t) = -x_0^*(-t) \tag{6.12b}$$

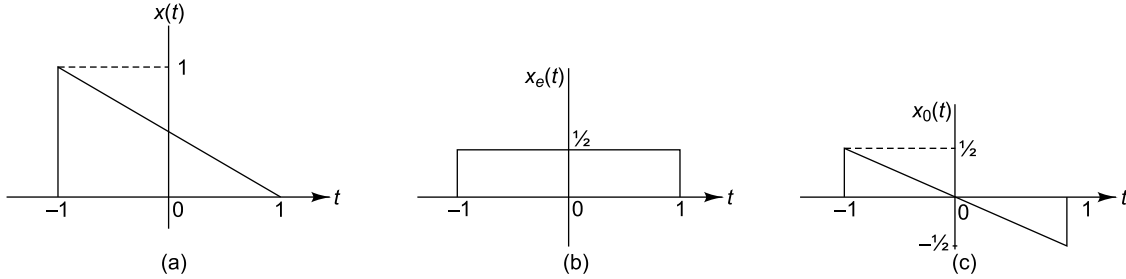## 6.4   SOME IMPORTANT CONCEPTS RELATED TO SIGNALS AND SYSTEMS

6.4.1   Unit Impulse Function

The concept of a unit impulse function is very important in the study of signals and systems. Such a signal can be described as shown in Fig. 6.12(a) and in the equation below:

$$\delta(t) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}, -\frac{\varepsilon}{2} \leq t \leq \frac{\varepsilon}{2}$$

$$= 0, |t| > \frac{\varepsilon}{2} \tag{6.13a}$$

Here ε is a very small, arbitrary quantity. The area under the function is computed by integrating $\delta(t)$ as shown in Eq. (6.13b) and is found to be unity. It is for this reason that this function is called a unit impulse.

$$A = \int_{-\infty}^{\infty} \delta(t)dt = \lim_{\varepsilon \to 0} \int_{-\varepsilon/2}^{\varepsilon/2} \frac{dt}{\varepsilon} = 1 \tag{6.13b}$$

Delta functions can be translated to any point t_0 in time and can be scaled in magnitude by a constant A to obtain $A\delta(t - t_0)$.

$$A\delta(t - t_0) = \infty, \; t = t_0 \tag{6.14a}$$

$$= 0, \; t \neq t_0$$

It follows that

$$\int_{t_0 - \varepsilon}^{t_0 + \varepsilon} A\delta(t - t_0)\,dt = A \tag{6.14b}$$

Figures 6.12(a)–(c) show the delta function, its delayed version and advanced version, respectively
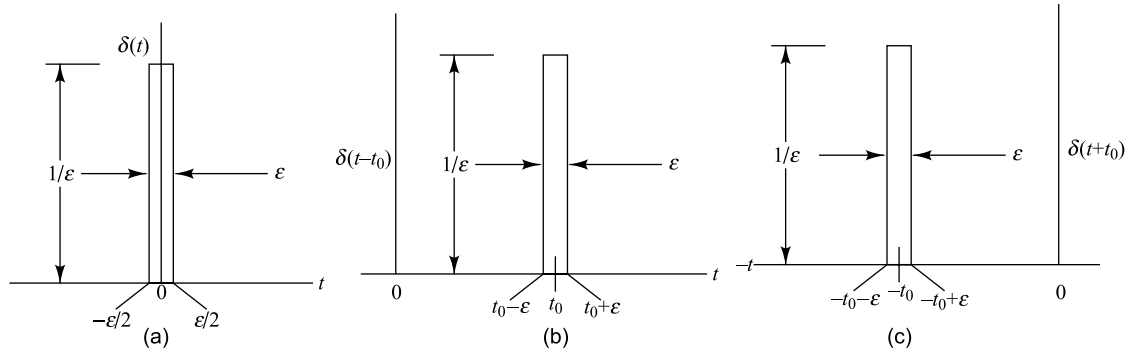

Figure 6.12 Delta function: (a) $\delta(t)$, (b) delayed version $\delta(t - t_0)$, (c) advanced version $\delta(t + t_0)$

Besides, when any arbitrary function $x(t)$ is multiplied by the translated delta function $\delta(t - t_0)$ and integrated over time, the value of $x(t)$ at $t = t_0$ is obtained, this being called the sifting property of the delta function. That is

$$\int_{-\infty}^{\infty} x(t)\delta(t - t_0)\,dt = x(t_0) \tag{6.15a}$$

It is also to be noted that convolving any signal $x(t)$ with the time-delayed Dirac delta function $\delta(t - T)$ produces a time delay in $x(t)$ of the same amount as shown below:

$$x(t) * \delta(t - T) = \int_{-\infty}^{\infty} x(\tau)\delta(t - T - \tau)\,d\tau = x(t - T) \tag{6.15b}$$

A point of great importance is that any arbitrary signal $x(t)$ can be represented as

$$x(t) = \int_{-\infty}^{\infty} x(\tau)\delta(t - \tau)\,d\tau \tag{6.16}$$

Equation (6.16) implies that $x(t)$ may be thought of as the superposition of an infinite number of magnitude scaled and translated (in time) impulse functions. Thus it is clear that delta functions can be used as basic building blocks for synthesizing arbitrary functions in the time domain, in much the same way that unit sinusoidal functions can be used for synthesizing periodic functions (please refer to Fourier series representation in Section 1.2.3.1).

6.4.1.1   *Unit sample sequence*   The discrete-time counterpart of the unit impulse signal is the unit sample sequence $\delta[n]$.

$$\delta[n] = 1, \, n = 0$$

$$= 0, \, n \neq 0 \tag{6.17}$$

As in the case of the translated delta function for the continuous time case, we can have the shifted unit sample functions as shown in Fig. 6.13.

$$\delta[n - n_0] = 1, \, n = n_0$$
$$= 0, \, n \neq n_0 \tag{6.18a}$$

$$\delta[n + n_0] = 1, \, n = -n_0$$
$$= 0, \, n \neq -n_0 \tag{6.18b}$$
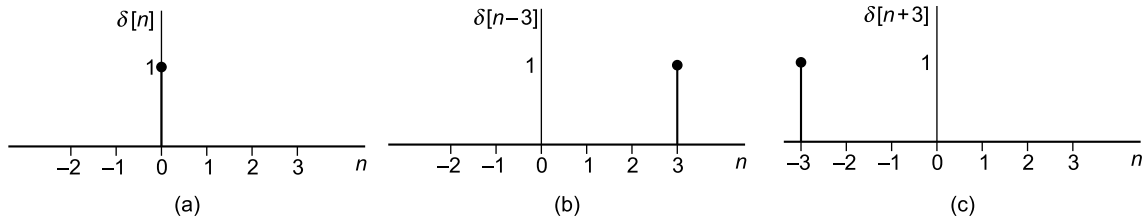
FIGURE 6.13   Unit sample sequence: (a) $\delta[n]$, (b) delayed version $\delta[n - 3]$, (c) advanced version $\delta[n + 3]$

Representation of any arbitrary sequence $x[n]$ may be made in terms of $\delta[n]$ as shown in Fig. 6.14. The signal shown may be expressed mathematically as

$$x[n] = a_1\delta[n - 1] + a_2\delta[n - 2] + a_5\delta[n - 5] + a_{-3}\,\delta[n + 3] \tag{6.19}$$
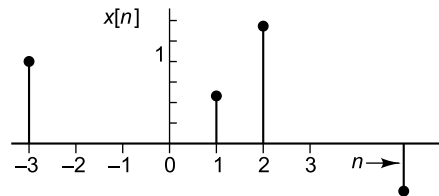
FIGURE 6.14   Representation of a sequence in terms of unit impulses

In general, any sequence may be represented as

$$x[n] = \sum_{k=-\infty}^{\infty} a_k\delta[n - k] \tag{6.20a}$$

or more generally as

$$x[n] = \sum_{k=-\infty}^{\infty} x[k]\delta[n - k] \tag{6.20b}$$

This is nothing but the convolution sum representing the superposition of magnitude-scaled and shifted unit impulses. $x[k]$ is the scale factor for the kth sample in the sequence.

6.4.2 Impulse Response Function

This is a very important function characterizing the time domain behaviour of a system which (i) is physically realizable, (ii) is stable, (iii) has constant or time-invariant parameters and (iv) is linear. The impulse response function (IRF) $h(t)$ is the output of the system (with zero initial conditions) to a unit impulse function. This implies that $h(t) = y(t)$ with $x(t) = \delta(t)$, the unit impulse function. The IRF is purely a function of the system properties and because it gives a complete description of the system's characteristics, it can be used to predict the response to any input. For a single input single output (SISO) system, the input $x(t)$ and the output $y(t)$ are related through the convolution integral

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t-\tau)\,d\tau \tag{6.21}$$

It is seen that Eq. (6.21) is a representation of the output in the form of the superposition of an infinite number of magnitude scaled and translated IRFs. These equations make the study of the input-output relationship easy to handle. If a physical system can be modelled as a stable, linear time-invariant system within allowable approximations, the opportunity should not be lost.

6.4.3 Linear Time-invariant System

For such a system, the properties of both linearity and time-invariance hold good. Convolution integral can be used to represent the system as we have just seen. An operator A is said to be linear if for any set of admissible values $x_1, x_2, x_3, \ldots x_N$ and constants $a_1, a_2, a_3, \ldots a_N$:

$$A\left[\sum_{i=1}^{N} a_i x_i\right] = \sum_{i=1}^{N} a_i A[x_i] \tag{6.22}$$

An operator A is said to be shift-invariant if for any shift t_0 of the input $x(t)$ to $x(t + t_0)$, there results a similar shift of output $y(t)$ to $y(t + t_0)$ for any t_0:

i.e. $$y(t + t_0) = A\left[x(t + t_0)\right] \tag{6.23}$$

Thus for a linear, time-invariant system, the response $y(t)$ to an arbitrary input $x(t)$ is equal to the convolution of the input with the impulse response of the system $h(t)$ as shown in Fig. 6.21 and in the equation below:

$$y(t) = x(t) * h(t)$$

$$= \int_{-\infty}^{\infty} x(\tau)h(t-\tau)\,d\tau \tag{6.24a}$$

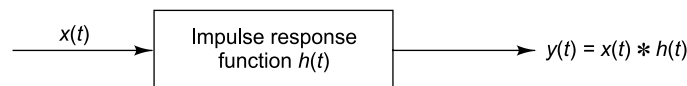where $*$ represents the convolution operator.

FIGURE 6.15 Mathematical representation of an LTI system

Substituting $t - \tau = t'$ in Eq. (6.24a), we get

$$y(t) = \int\limits_{-\infty}^{\infty} x(t - t')h(t')\,dt' \qquad (6.24\text{b})$$

Instead of variable t', this equation can be written in terms of variable τ to give

$$y(t) = \int\limits_{-\infty}^{\infty} h(\tau)x(t - \tau)\,d\tau \qquad (6.24\text{c})$$

$$= h(t) * x(t)$$

From Eqs (6.24a)–(6.24c) it is seen that

$$y(t) = h(t) * x(t) = x(t) * h(t) \qquad (6.25)$$

This implies that we can change the order of the IRF and system input in the computation of the system output in a series combination of the system. This is a very important relationship, serving as the foundation for engineering analysis of linear systems. An equivalent description of the IRF in the frequency domain is the Fourier transform of the impulse response, and this is called the frequency response function (FRF) of the system. According to the convolution theorem which will be seen later, the convolution of the input with the IRF is equivalent to multiplication of the input spectrum with the FRF of the system. Since multiplication is an operation which is easier to perform than convolution, frequency domain analysis is often preferred to time domain analysis with the final result being transformed back into the time domain if required.

6.4.3.1 *Discrete-time convolution or convolution sum*   For the discrete-time case, we have the concept of a linear shift-invariant (LSI) system, akin to the linear time-invariant (LTI) system and the convolution integral is replaced by the convolution sum. Although the convolution sum is analogous to the convolution integral, it is to be remembered that it is not an approximation to the integral. While the integral plays mainly a theoretical role for continuous systems, the sum often serves as an explicit realization of a discrete-time linear system.

For any input $x[n]$, the output $y[n]$ is

$$y[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] = \sum_{k=-\infty}^{\infty} h[k]x[n-k] \qquad (6.26)$$

where $h[n]$ is the IRF.

6.4.4 Causal Systems and Signals

This relates to the cause and the effect produced in a system. A causal system is one, the output of which is a function of the present and past inputs only, not future inputs. Causal signals may be defined in a similar fashion. Causal signals are those which are zero for $t < 0$. Since the IRF $h(t)$ is the output of a system to the unit impulse function $\delta(t)$ which is zero for $t < 0$, $h(t) = 0$ for $t < 0$ for a causal system, or in other words the IRF is a causal signal. Non-causal signals are those that exist for $t < 0$ as well as for $t \geq 0$. Anticausal signals are those which are zero for all $t \geq 0$. Examples of such signals are shown in Fig. 6.16.
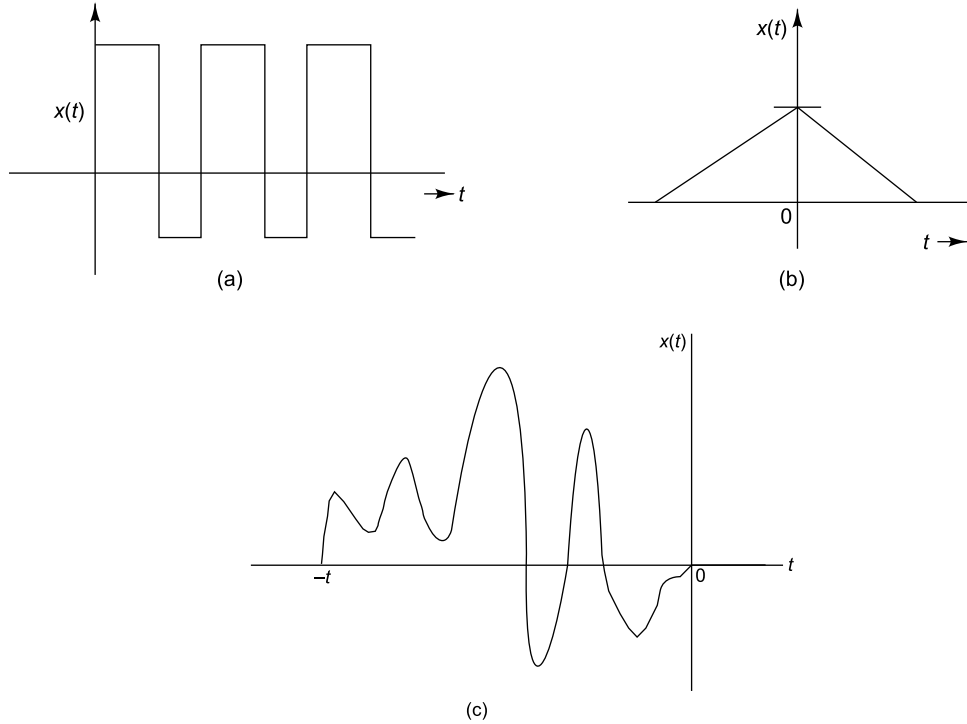
FIGURE 6.16 Causality of signals: (a) causal signal, (b) non-causal signal, (c) anticausal signal

6.4.5 Stability

This is a property of a system, not a signal. There are several definitions for stability available in literature. We shall discuss one that is appropriate for causal, linear, time-invariant systems. A system is stable in the *bounded-input, bounded-output* (BIBO) sense if and only if output $y(t)$ is bounded for all possible bounded inputs $x(t)$. If even one bounded input gives rise to an unbounded output, then the system is said to be BIBO unstable.

An input $x(t)$ is said to be bounded if

$$|x(t)| < B_x < \infty \tag{6.27}$$

For such a bounded input to an LTI system, we can expect that

$$|y(t)| \leq \int_{-\infty}^{\infty} |x(\tau)| \, |h(t-\tau)| \, d\tau \tag{6.28}$$

$$\leq B_x \int_{-\infty}^{\infty} |h(t-\tau)| \, d\tau$$

Since any output is obtained as the convolution of the input with the IRF for an LTI system, this implies that if the IRF is absolutely integrable, then the output to every bounded input is bounded. From this it follows that for an asymptotically stable system [for which $h(t)$ decays to zero as time goes to infinity],

$$\int_{-\infty}^{\infty} |h(t)| \, dt < B_h < \infty \tag{6.29}$$

and $$|y(t)| \le B_x B_h < \infty \tag{6.30}$$

It is to be noted that the IRF tells us all we want to know about a system. For instance, if we want to know about the stability and causality of a system, we just have to inspect $h(t)$ of the system.

For stability of a discrete-time system, the following necessary and sufficient condition has to be satisfied.

$$S = \sum_{k=-\infty}^{\infty} |h(k)| < \infty \tag{6.31}$$

6.4.6   Frequency Response Function

The dynamic properties of a system may be described in the frequency domain by what is called the FRF. For a physical understanding of the FRF, it is worth mentioning that a stable LTI system, in response to a harmonic excitation at a particular frequency, produces a harmonic output at the same frequency. The FRF at a particular frequency is the response of such a system to unit harmonic input at that frequency and is generally plotted at all possible frequencies in the frequency range of interest. The FRF and hence magnitude and phase lag of the output are a function of the system parameters alone, being dependent neither on the input, nor on the output. The system gain is defined as

$$|H(\omega)| = \left| \frac{Y(\omega)}{X(\omega)} \right| \tag{6.32}$$

and the associated phase lag as

$$\phi(\omega) = \tan^{-1} \frac{\text{Im}\{H(\omega)\}}{\text{Re}\{H(\omega)\}} \tag{6.33}$$

Here $Y(\omega)$ and $X(\omega)$ constitute the Fourier transforms of the output and input, respectively. The FRF is also the Fourier transform of the IRF. Thus

$$|H(\omega)| = \int_{-\infty}^{\infty} h(t) e^{-i\omega t} \, dt \tag{6.34a}$$

It follows that the IRF is the inverse Fourier transform of the FRF.

$$h(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} H(\omega) e^{i\omega t} \, d\omega \tag{6.34b}$$

It may be noted that the lower limit of integration may be taken as zero in the transform Eqs (6.34a) and (6.34b) since $h(t) = 0$ for $t < 0$ for a physically realizable or causal system and since negative frequencies ω do not have any practical meaning. The FRF is a complex, frequency dependent quantity. It is generally described graphically in three different forms:

(i) $|H(\omega)|$ versus frequency and $\angle H(\omega)$ versus frequency, constituting a set of two plots. Conventionally phase is defined such that phase lags are considered positive. If log–log scales are used for the axes in the magnitude plots, these are called Bodé plots.

(ii) Re$\{H(\omega)\}$ versus frequency and Im$\{H(\omega)\}$ versus frequency, constituting a set of two plots.

(iii) Polar plot where Re$\{H(\omega)\}$ is plotted versus Im$\{H(\omega)\}$ as ω varies from zero to infinity.

It may be observed that the frequency information is not explicit in this notation and has to be marked manually. The various plots are discussed in detail in Chapters 8 and 10.

6.4.6.1 *Obtaining FRF of a system*

The FRF of a system may be obtained in the following two ways:

(i) Excite the system with $\delta(t)$ to obtain $h(t)$ as output; then $H(\omega)$ is obtained by the Fourier transformation of $h(t)$ as described in Eq. (6.34a).

(ii) Alternatively, $H(\omega)$ may be obtained by imparting swept sine or pure random excitation to obtain $H(\omega)$ straightaway from a frequency domain operation.

In the first method, the system is excited with a unit impulse to obtain $h(t)$. The Fourier transform of the unit impulse is given by

$$\mathfrak{J}\{\delta(t)\} = \int_{-\infty}^{\infty} \delta(t)e^{-i\omega t}\,dt = 1 \tag{6.35}$$

This implies that the unit impulse has a spectrum with ω or f extending from $-\infty$ to $+\infty$ with spectral height unity. Thus in the context of vibration testing, when we excite a system with a unit impulse in the time domain, we are in effect imparting harmonic excitation at all frequencies simultaneously to the system, causing it to respond simultaneously at all its natural frequencies. This is the basis of a free vibration test, which is a time domain method used for bringing out the system parameters in the form of the IRF. The FRF $H(\omega)$ may then be got by doing a Fourier transformation on the IRF.

In the second method, we obtain the FRF by imparting harmonic excitation. With the function $e^{i\omega t}$ as input, the output of an LTI system is obtained from the convolution operation as

$$y(t) = \int_{-\infty}^{\infty} h(\tau)x(t-\tau)\,d\tau = \int_{-\infty}^{\infty} h(\tau)e^{i\omega(t-r)}\,d\tau$$

$$= e^{i\omega t}\int_{-\infty}^{\infty} h(\tau)e^{-i\omega^{\tau}}\,d\tau = e^{i\omega t}\,H(\omega) \tag{6.36}$$

This equation is a statement of the fact that complex exponentials are eigenfunctions of LTI systems and $H(\omega)$ are the eigenvalues. When we subject a system to swept sine excitation, we impart sequentially harmonic excitations of different frequencies. The output in this case is directly brought out as the FRF since this is a frequency domain method. To avoid the large amount of time taken for testing in the frequency domain using swept sine excitation, we may resort to pure random excitation, where instead of exciting the system at increasing or decreasing frequencies sequentially, we excite the system at all frequencies simultaneously, thus obtaining the FRF in one shot and thereby reducing testing time considerably. The FRF that we get, whether it is from swept sine excitation or random excitation, is nevertheless the same, being defined only by the system. This forms the basis for frequency domain methods.

6.5 FREQUENCY DOMAIN ANALYSIS

Most of the early test methods for system identification were time domain methods. Later these methods were replaced by frequency domain methods. Narrow band frequency analysis was originally done using

constant-bandwidth narrow band analysers. A good description of frequency analysers is given in Section 8.11.1 . These methods were exceedingly slow, especially for large signal bandwidths and were limited to a 1.5 to 2.0 Hz resolution. A statistically averaged analysis of data from 0 to 500 Hz with 2 Hz resolution would typically take 2 hr to complete. With the progresses in the electronics industry in the late 1960s and in signal processing techniques in the early 1970s, the FFT-based spectrum analysers were born. The FFT behaves like a bank of parallel filters. The FFT-based machines inherently provide phase information which is required for the high-speed FRF computation. The popularity of the two channel machines led to the demand for multichannel FFT machines. Frequency analysis using these machines makes use of the Fourier transform pair of equations to convert a signal from the time domain to the frequency domain and vice versa. As has been seen in Chapter 1, Fourier's approach is based on the assumption that any waveform can be represented as the superposition of a series of sine and cosine waves at different frequencies. A Fourier series is used for representation of continuous-time periodic signals. For continuous-time aperiodic transient signals, the Fourier transform or analysis equation is used. This is defined as

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt \tag{6.37a}$$

The inverse Fourier transform or synthesis equation is given by

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega \tag{6.37b}$$

All signals are not Fourier transformable. The question that arises is which class of signals can be represented by Fourier and inverse Fourier transforms. The existence of the transform is assured for any $x(t)$ that satisfies the Dirichlet conditions. (Readers are suggested to refer to any book on signal analysis.) The first of these conditions is that

$$\int_{-\infty}^{\infty} |f(t)| dt < \infty \tag{6.38a}$$

Since

$$|X(\omega)| = \left| \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt \right| < \int_{-\infty}^{\infty} |x(t)| |e^{-i\omega t}| dt \leq \int_{-\infty}^{\infty} |x(t)| dt < \infty \tag{6.38b}$$

it follows that for the Fourier transform to exist $|X(\omega)| < \infty$ for all ω. Thus the sufficient condition for convergence is

$$|X(\omega)| \leq \int_{-\infty}^{\infty} |x(t)| dt < \infty \tag{6.39}$$

Therefore, if $x(t)$ is absolutely integrable, then $X(\omega)$ exists, or in other words any stable function has a Fourier transform. Any finite length function thus has a Fourier transform because it is absolutely integrable. We know that the output $y(t)$ of an LTI system with input $x(t)$, and finite-duration impulse response $h(t)$ can be obtained from the convolution integral. Since Eq. (6.40) has to be satisfied for BIBO stability, it can be said that any system with a finite-duration impulse response (FIR) will be stable and will have a Fourier transform.

$$S = \int_{-\infty}^{\infty} |h(t)| \, dt < \infty \tag{6.40}$$

Some functions are not absolutely integrable, but are square integrable as shown further.

$$\int_{-\infty}^{\infty} |x(t)|^2 \, dt < \infty \tag{6.41}$$

If we relax the condition of uniform convergence and have mean square convergence instead, then those functions $x(t)$ which are not absolutely integrable may also be assumed to have Fourier transforms.

6.5.1 Symmetry Properties of the Fourier Transform

Table 6.2 gives the symmetry properties of the Fourier transform. The conjugate symmetric and conjugate antisymmetric parts of a complex function, as described in Eqs (6.11) and (6.12) are used here. The corresponding even and odd parts of a real function are as described in Eqs (6.9) and (6.10).

TABLE 6.2   Symmetry properties of the Fourier transform

Signal $x(t)$	*Fourier transform $X(\omega)$*
$x(t) = \dfrac{1}{2\pi} \displaystyle\int_{-\infty}^{\infty} X(\omega) e^{i\omega t} \, d\omega$	$X(\omega) = \displaystyle\int_{-\infty}^{\infty} x(t) e^{-i\omega t} \, dt$
$x^*(t)$	$X^*(-\omega)$
$x^*(-t)$	$X^*(\omega)$
$\mathrm{Re}\,\{x(t)\}$	$X_e(\omega)$, conjugate symmetric part of $X(\omega)$
$i\,\mathrm{Im}\,\{x(t)\}$	$X_o(\omega)$, conjugate antisymmetric part of $X(\omega)$
$x_e(t)$, conjugate symmetric part of $x(t)$	$\mathrm{Re}\,\{X(\omega)\}$, i.e. $X_R(\omega)$
$x_o(t)$, conjugate antisymmetric part of $x(t)$	$i\,\mathrm{Im}\{X(\omega)\}$, i.e. $iX_I(\omega)$
For $x(t)$ real, the following properties apply	
$X(\omega) = X^*(-\omega)$	
$X_R(\omega) = X_R(-\omega)$, real part is even	
$X_1(\omega) = -X_1(-\omega)$, imaginary part is odd	
$|X(\omega)| = |X(-\omega)|$, magnitude is even	
$\angle X(\omega) = -\angle X(-\omega)$, phase is odd	
$x_e(t)$, even part of $x(t)$	$X_R(\omega)$
$x_o(t)$, odd part of $x(t)$	$iX_1(\omega)$

6.5.2 Fourier Transform Theorems

The Fourier transform is very powerful mainly due to the many theorems describing the properties of the transformation; these theorems provide insight into the nature of physical systems. The forward and inverse Fourier transform equations offer complementary views of the same signal. Hence, if some operation is performed on one equation of the pair, then it follows intuitively and can be proved mathematically that an equivalent operation can be performed on the other equation. In all the equations below we will use the notation $x(t) \leftrightarrow X(\omega)$ which means that $x(t)$ has a Fourier transform $X(\omega)$.

6.5.2.1 *Linearity* This theorem states that if a function is scaled in the time domain, its transform also gets scaled by the same amount. Thus adding two functions in the time domain corresponds to adding the two frequency spectra. If $x_1(t) \leftrightarrow X_1(\omega)$ and $x_2(t) \leftrightarrow X_2(\omega)$, then

$$ax_1(t) + bx_2(t) \leftrightarrow aX_1(\omega) + bX_2(\omega) \tag{6.42}$$

6.5.2.2 *Scaling* This theorem states that if a signal is compressed in the time domain, it gets expanded in the frequency domain and vice versa. A change in the scale of the time axis by a positive real constant a thus changes the scale of the frequency axis in the spectrum by $1/a$. Thus

$$x(at) \leftrightarrow \frac{1}{|a|} X\left(\frac{\omega}{a}\right) \tag{6.43a}$$

$$x(t / a) \leftrightarrow |a| X(\omega a) \tag{6.43b}$$

An implication of this theorem is that for any real $x(t)$,

$$x(-t) \leftrightarrow X(-\omega) \tag{6.43c}$$

This is got by letting $a = -1$ in Eq. (6.43a). In other words, flipping the time function about the origin corresponds to flipping its spectrum about the origin. If $x(t)$ is complex, then

$$x^*(-t) \leftrightarrow X^*(\omega) \tag{6.43d}$$

6.5.2.3 *Time shifting and frequency shifting* **Time shift:** Shifting a function in time (delaying or advancing) induces a phase shift proportional to the frequency and to the amount of time shift.

Thus
$$x(t - t_0) \leftrightarrow X(\omega)\, e^{-i\omega t_0} \tag{6.44a}$$

Frequency shift: Conversely, shifting in frequency causes the time function to be multiplied by a unit phasor with angle proportional to time and to the amount of shift.

i.e.
$$x(t)\, e^{i\omega t_0} \leftrightarrow X(\omega - \omega_0) \tag{6.44b}$$

6.5.2.4 *Duality between time and frequency domains (transform of a transform)* In order to convert from the frequency domain back to the time domain, we normally use the inverse Fourier transform. However, if instead, we take a forward Fourier transform of the spectrum, the result is a time function which has been reversed about the y-axis. This implies that

$$X(\omega) \leftrightarrow x(-t) \tag{6.45a}$$

and
$$X(t) \leftrightarrow x(-\omega) \tag{6.45b}$$

For example, $\text{rect}(t) \leftrightarrow \text{sinc}(f)$ implies that $\text{sinc}(t) \leftrightarrow \text{rect}(-f)$ where the rectangular and sinc functions are as shown below.

$$\text{rect}\left(\frac{t}{\tau}\right) = 1, |t| < \frac{\tau}{2}$$
$$= 0, |t| > \frac{\tau}{2} \tag{6.46}$$

$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t} \tag{6.47}$$

6.5.2.5 Convolution   This theorem states that if two functions are convolved in the time domain, then their Fourier transforms are multiplied in the frequency domain. Conversely, if they are multiplied in the time domain, then their Fourier transforms are convolved in the frequency domain. If $x_1(t) \leftrightarrow$ $X_1(\omega)$ and $x_2(t) \leftrightarrow X_2(\omega)$ and if $y(t) = x_1(t) * x_2(t) = \int\limits_{-\infty}^{\infty} x_1(\tau) x_2(t - \tau) d\tau$, then

$$x_1(t) * x_2(t) \leftrightarrow X_1(\omega) \,.\, X_2(\omega) \tag{6.48}$$

Typically for a practical LTI system, the output $y(t)$ is the convolution of the input $x(t)$ and the IRF $h(t)$. The convolution theorem implies that the Fourier transform of the output of such a system is the product of the Fourier transform of the input and the FRF, which is the Fourier transform of the IRF. Convolution obeys the commutative, associative, and distributive laws of algebra as shown below.

Commutative law   : $x_1(t) * x_2(t) = x_2(t) * x_1(t)$ $\qquad$ (6.49a)

Associative law   : $x_1(t) * ((x_2(t) * (x_3(t)) = (x_1(t) * x_2(t)) * x_3(t)$ $\qquad$ (6.49b)

Distributive law   : $x_1(t) * ((x_2(t) + (x_3(t)) = (x_1(t) * x_2(t)) + (x_1(t) * x_3(t))$ $\qquad$ (6.49c)

6.5.2.6 Modulation or windowing   This is similar to convolution; here multiplication in the time domain is equivalent to convolution in the frequency domain. If $x(t) \leftrightarrow X(\omega)$ and $w(t) \leftrightarrow W(\omega)$, then

$$y(t) = x(t).w(t) \leftrightarrow Y(\omega) = \frac{1}{2\pi} \int\limits_{-\pi}^{\pi} X(\theta).w(\omega - \theta) d\theta \tag{6.50}$$

From Euler's relationship $\cos \theta = \dfrac{e^{i\theta} + e^{-i\theta}}{2}$, we infer that multiplication of a time function by a cosine signal splits the frequency spectrum, causing half of the spectrum to shift to a negative frequency to the left-hand side and half to shift to the right-hand side to the corresponding positive frequency on the frequency axis. The frequency shift theorem states that $x(t)e^{i\omega t_0} \leftrightarrow X(\omega - \omega_0)$ and $x(t) e^{-i\omega t_0} \leftrightarrow X(\omega + \omega_0)$ (from Eq. 6.44b). From the linearity theorem it follows that

$$x(t) \cos \omega_0 t \leftrightarrow \frac{1}{2}[X(\omega + \omega_0) + X(\omega - \omega_0)] \tag{6.51a}$$

and

$$x(t) \sin \omega_0 t \leftrightarrow \frac{1}{2i}[X(\omega - \omega_0) - X(\omega + \omega_0)] \tag{6.51b}$$

The modulation theorem is the basis of amplitude modulation.

6.5.2.7 Differentiation in time   Differentiating a function in time causes a 90° phase shift in the spectrum and scales the magnitude of the spectrum in proportion to the frequency. Repeated differentiation leads to the result

$$\frac{d^n x(t)}{dt^n} \leftrightarrow (i\omega)^n\, X(\omega) \tag{6.52}$$

This theorem helps us understand why differentiation of a signal has the reputation of being a noisy operation. If there is high frequency noise in a signal, it will get greatly amplified by differentiation.

6.5.2.8 *Differentiation in frequency* This theorem states that multiplying a function $x(t)$ by t amounts to differentiating the signal in the frequency domain.

$$tx(t) \leftrightarrow i\frac{dX(\omega)}{d\omega} \tag{6.53a}$$

Repeated multiplications lead to the result

$$(-it)^n\, x(t) \leftrightarrow \frac{d^n X(\omega)}{d\omega^n} \tag{6.53b}$$

6.5.2.9 *Integration* Integrating a function in the time domain introduces a $-90°$ phase shift in the spectrum and scales the magnitude of the spectrum inversely with frequency. Thus

$$\int_{-\infty}^{t} x(u)\,du \leftrightarrow \frac{X(\omega)}{i\omega} + C, \tag{6.54}$$

where C is a constant.

From this theorem we see that the integration operation is akin to passing the signal through a low pass filter which blurs the signal. Practically speaking, integrators are more popular than differentiators for the reason that they do not amplify high frequency noise.

6.5.2.10 *Parseval/Rayleigh theorem* This theorem states that the total power in any physical system can be computed, either by integrating over the time domain or over the frequency domain. What is called Parseval's energy conservation theorem in the context of Fourier series representation is often called Rayleigh's theorem in the context of Fourier transforms.

$$\int_{-\infty}^{\infty} |x(t)|^2\, dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)|^2\, d\omega \tag{6.55a}$$

The quantity $|X(\omega)|^2$ is called the energy density spectrum or power spectral density and is a quantity which is extensively used for the description of random signals for which the Fourier transform does not exist. The integral on the left-hand side gives the total amount of energy in the signal as computed in the time domain, whereas the integral on the right hand side gives the total amount of energy computed in the frequency domain. The modulus is found since the integrand is, in general, complex and therefore it is the magnitude of this quantity which is integrated. A more general formulation of Parseval's theorem for two signals is as follows. If $x_1(t) \leftrightarrow X_1(\omega)$ and $x_2(t) \leftrightarrow X_2(\omega)$, then

$$\int_{-\infty}^{\infty} x_1(t)x_2^*(t)\,dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_1(\omega)X_2^*(\omega)\,d\omega \tag{6.55b}$$

Table 6.3 summarizes all the Fourier transform theorems seen and Table 6.4 shows the Fourier transforms of some common signals. Figure 6.17 illustrates the scaling property of the Fourier transform and Fig. 6.18 shows the Fourier transform of some common functions.

TABLE **6.3** Fourier transform theorems

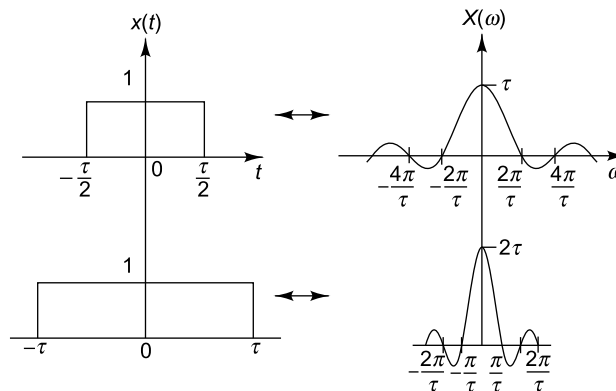| *Time signal*	*Corresponding transform*
$x(t) = \dfrac{1}{2\pi} \int\limits_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega$	$X(\omega) = \int\limits_{-\infty}^{\infty} x(t) e^{-i\omega t} dt$
$ax_1(t) + bx_2(t)$	$aX_1(\omega) + bX_2(\omega)$
$x(at)$	$\dfrac{1}{|a|} X\left(\dfrac{\omega}{a}\right)$
$x^*(-t)$	$X^*(\omega)$
$x(t - t_0)$	$X(\omega)\, e^{-i\omega t_0}$
$x(t) e^{i\omega t_0}$	$X(\omega - \omega_0)$
$x(-t)$	$X(\omega)$
$X(t)$	$x(-\omega)$
$\int\limits_{-\infty}^{\infty} x_1(\tau) x_2(t - \tau)\, d\tau$	$X_1(\omega) X_2(\omega)$
$x(t) \cos \omega_0 t$	$\dfrac{1}{2}[X(\omega + \omega_0) + X(\omega - \omega_0)]$
$x(t) \sin \omega_0 t$	$\dfrac{1}{2}[X(\omega - \omega_0) - X(\omega + \omega_0)]$
$\int\limits_{-\infty}^{t} x(u)\, du$	$\dfrac{X(\omega)}{i\omega} + C,$ where C is a constant
$\dfrac{d^n x(t)}{dt^n}$	$(i\omega)^n X(\omega)$
$(-it)^n x(t)$	$\dfrac{d^n X(\omega)}{d\omega^n}$
$\int\limits_{-\infty}^{\infty} x(t + \tau) x^*(\tau)\, dt$	$|X(\omega)|^2$
$\int\limits_{-\infty}^{\infty} |x(t)|^2\, dt = \dfrac{1}{2\pi} \int\limits_{-\infty}^{\infty} |X(\omega)|^2\, d\omega$	

FIGURE **6.17** Fourier transform–scaling property

TABLE 6.4   Fourier transforms of common signals

Function	Time waveform g(t)	Spectrum G(f)
Rectangular	$\text{rect}\left(\dfrac{T}{\tau}\right)$	$\tau\,\text{sinc}(\tau f)$
Constant	1	$\delta(f)$
Impulse at $t = t_0$	$\delta(t - t_0)$	$\exp(-i2\pi t_0 f)$
Sinc	$\text{sinc}\,(2Wt)$	$\dfrac{1}{2W}\text{rect}\left(\dfrac{f}{2W}\right)$
Phasor	$\exp[i(2\pi f_0 t + \phi)]$	$\exp(i\phi)\delta(f - f_0)$
Harmonic	$\cos(2\pi f_c t + \phi)$	$\dfrac{1}{2}\left[\exp(i\phi)\delta(f - f_c) + \exp(-i\phi)\delta(f + f_c)\right]$
Impulse train	$\displaystyle\sum_{k=-\infty}^{\infty}\delta(t - kT)$	$f_0\displaystyle\sum_{n=-\infty}^{\infty}\delta(f - nf_0),\ f_0 = \dfrac{1}{T}$

In the above equations, rect and sinc functions are as defined in Eqs. (6.46) and (6.47).
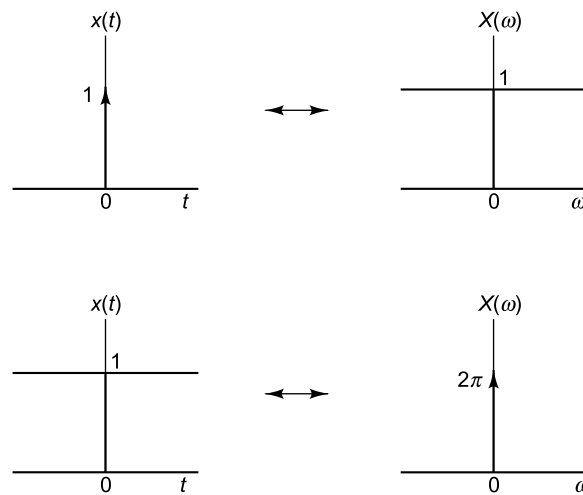

FIGURE 6.18   Common Fourier transforms

6.5.3   Fourier Transform of Sequences

The Fourier transform of a sequence $x[n]$ can be represented as

$$X(e^{i\omega}) = \sum_{n=-\infty}^{\infty} x[n]e^{-i\omega n} \tag{6.56a}$$

and the inverse Fourier transform is defined as

$$x[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(e^{i\omega})e^{i\omega n}\,d\omega \tag{6.56b}$$

$X(e^{j\omega})$ is a continuous function of ω and is periodic in ω with period 2π. $x[n]$ is seen to be a superposition of infinitesimally small complex sinusoids of the form $\dfrac{1}{2\pi}X(e^{i\omega})e^{i\omega n}d\omega$ with frequency ω ranging over a length 2π. Thus there is a clear equivalence between the Fourier series representation of a continuous-time periodic signal and the Fourier transform of a discrete-time signal, and all properties of the Fourier series can be applied to the Fourier transform representation of a sequence with appropriate interpretation of variables. All the Fourier transform theorems explained earlier hold good for sequences also, with some minor variations.

6.6 SAMPLING OF CONTINUOUS TIME SIGNALS

Often any signal to be analysed is analogue in nature as mentioned earlier, typically being a continuous function of time (or space in a digitized image), or any other independent variable. It is generally captured as a series of values recorded on a tape, data logger or computer. It must then be converted into digital form by an analogue to digital converter for further processing. For this conversion, it is necessary to sample the continuous-time signal in some manner. Any analogue signal $x_c(t)$ can be converted to a discrete-time form by sensing values of $x_c(t)$ at regular intervals of time nT_s (where n is an integer), to obtain sequence $x[n]$. T_s is called the sampling period or sampling interval and this type of sampling done at regular intervals is called uniform sampling. It is customary to specify a sampling rate or frequency $f_s = (1/T_s)$ in hertz, rather than the sampling period in seconds. Each sample value resulting from sampling is associated with the instant of time when it was captured. The samples can be used to reconstruct the original signal by simply joining the points to form the envelope of the original analogue signal $x_c(t)$ and this process is called reconstruction; for this the sampling frequency should be known. If there is a rapid variation of the signal with respect to time, it implies that there are high frequency components in the signal. Intuitively one could say that the higher the frequencies of the components present in the signal, the higher the sampling frequency should be. If the sampling frequency is not high enough, or if we perform 'undersampling', some of the information that is present in the signal may be lost due to a phenomenon called aliasing. In cinematography, aliasing manifests as the wagon-wheel effect, in which a spoked wheel appears to rotate too slowly or even backwards due to an insufficient frame rate. The reconstructed signals are better with higher sampling frequencies or with 'oversampling'; however, this calls for faster ADCs and more storage, increasing the cost of signal processing. Therefore, the consideration of sampling frequency is very important and one should weigh the pros and cons of each application, and be aware of the tradeoffs involved. In practice, the sampling interval is typically quite small, of the order of milliseconds or microseconds for vibration and acoustic signals. Figure 6.19 shows the effect of sampling at a sufficiently high frequency.
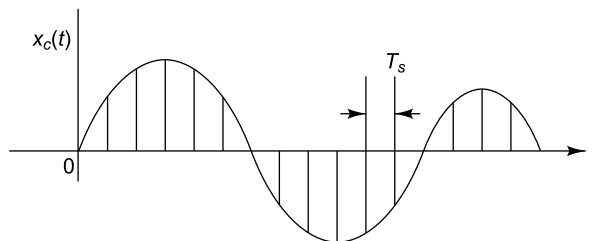

FIGURE 6.19 Sampling at a sufficiently high rate

6.6.1 Undersampling and Aliasing

If the sampling rate (frequency) is not high enough, it is impossible to rebuild the original signal from the samples. In fact the actual high frequency components in the original signal take on the identity or alias of lower frequencies and this phenomenon is called aliasing. Specifically, undersampling causes signal components with a frequency higher than half the sampling frequency to appear as low frequencies which are not actually present in the original signal, or which are indistinguishable from genuine low frequency components and get superimposed over them, causing distortion of the signal. Besides, some of the frequencies in the original signal may be irreparably lost in the reconstructed signal. Aliasing occurs because signal frequencies can overlap if the sampling frequency is too low and frequencies 'fold' about half the sampling frequency, which is also called the folding frequency as will be evident from the theoretical explanation in Section 6.6.3.1. If for instance, we listen to music with high frequency components, which has been reconstructed after sampling at too low a frequency, the low frequency aliases of the undersampled high frequencies will be audible. Figure 6.20 shows two cases of undersampling and aliasing: (i) $f_s = 2f_0$ and (ii) $f_s < 2f_0$, where f_0 is the frequency of the sine wave and f_s is the sampling frequency. The reconstructed signals built from the data samples will not represent the original signals at all in both cases since the sampled values do not have enough information to do so. To prevent or reduce aliasing, two things can be done: (i) Increase the sampling rate, to above twice the highest frequency of interest, (ii) Introduce an anti-aliasing (low pass) filter to limit the frequencies in the analogue signal to less than half the sampling frequency.
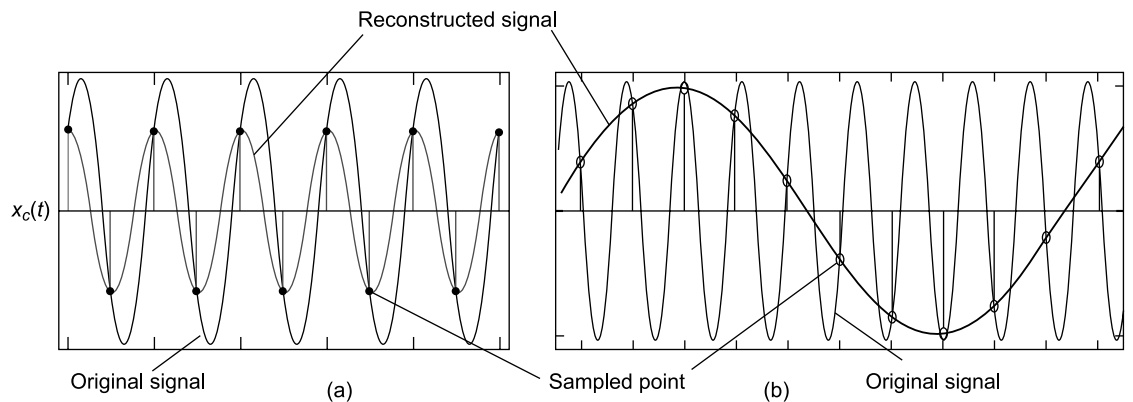

FIGURE 6.20   Depiction of undersampling: (a) $f_s = 2 f_0$, (b) $f_s < 2 f_0$

6.6.2 Anti-aliasing Filter

Sometimes the highest frequency components of a signal may not be of interest for the analysis or may simply be noise. The amount of aliasing due to such signals may be reduced by restricting the bandwidth and filtering out these higher frequency components before sampling the signal. This is done using an anti-aliasing filter, which is nothing but an ideal low pass filter (which filters out high frequency components and lets lower frequency components through) with very sharp cut off characteristics. Theoretically, a perfect boxcar low pass filter with a cut-off at exactly the Nyquist frequency (highest frequency in the signal) should eliminate all the frequencies above the Nyquist frequency. However ideal

filters with infinite attenuation at the cut-off frequencies cannot be realized. All practical filters only cause considerable attenuation of frequencies beyond the cut-off frequency, and do not remove them completely. However, the low frequency aliases thus generated have very low amplitude levels, so as not to cause a problem. Due to the sharp roll off requirements, anti-aliasing filters are expensive. Through combinations of digital and analogue low pass filter technology, industry has been able to provide anti-aliasing filters having an attenuation of 60–120 dB/octave beyond the cut-off frequency. The cut-off frequency is generally set at 60 to 80% of the Nyquist frequency (highest frequency in the signal), so as to drive the aliasing signals into the analyser noise. Figure 6.21(a) shows the characteristics of an ideal anti-aliasing filter and Figure 6.21(b) that of an actual anti-aliasing filter.
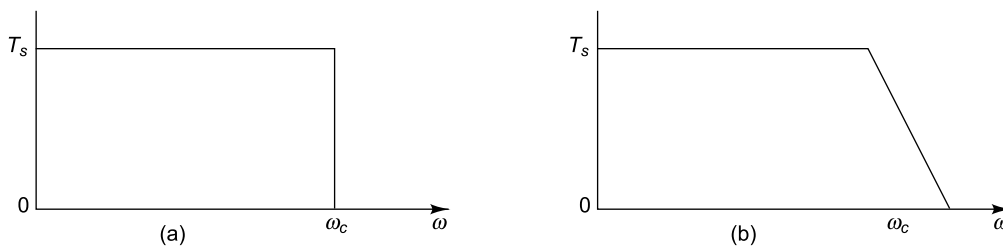

FIGURE 6.21   Anti-aliasing filter: (a) ideal, (b) actual

For example if the highest frequency of interest in a signal is 500 Hz, the filter is typically set at 300–400 Hz depending on the sharpness of the anti-aliasing filter. This implies that one will lose the frequency components between 300 and 500 Hz, since their magnitude and phase information have got modified. This is the price that one pays for anti-aliasing protection.

6.6.3   SAMPLING THEORY

In the case of periodic or regular sampling, a sequence of samples $x[n]$ is obtained from a continuous time signal $x_c(t)$ according to the relation

$$x[n] = x_c(nT_s), \ -\infty < n < \infty \tag{6.57}$$

where T_s is the sampling period and $f_s = (1/T_s)$ is the sampling frequency. It is convenient to mathematically represent the sampling process in the two stages depicted in Figure 6.22(a). It consists of an impulse train modulator followed by conversion of the impulse train to a sequence. Let $x_c(t)$ represent the analogue signal to be sampled. Sampling in the time domain is done mathematically by multiplying $x_c(t)$ with a periodic impulse train of period T_s (train of delta functions T_s apart). This corresponds to a point by point multiplication of the sampling impulse function with $x_c(t)$. By doing so, only the values of $x_c(t)$ corresponding to the instants at which the impulse train occurs are sifted, the product being zero for all other times. This is analogous to sampling $x_c(t)$ with a frequency $f_s = (1/T_s)$, resulting in the new sampled signal denoted as $x_s(t)$. Figure 6.22(b) illustrates a continuous time signal, the result of impulse train sampling and the corresponding output sequence. The essential difference between $x_s(t)$ and $x[n]$ is that $x_s(t)$ is, in a sense a continuous time signal (specifically an impulse train) which is zero except at integral multiples of T_s. The sequence $x[n]$ on the other hand, is indexed on the integer variable n (which is just a constant) and does not contain any explicit information about sampling rate. Furthermore, samples of $x_c(t)$ are represented as finite numbers in $x[n]$, rather than as the areas of impulses as in $x_s(t)$.
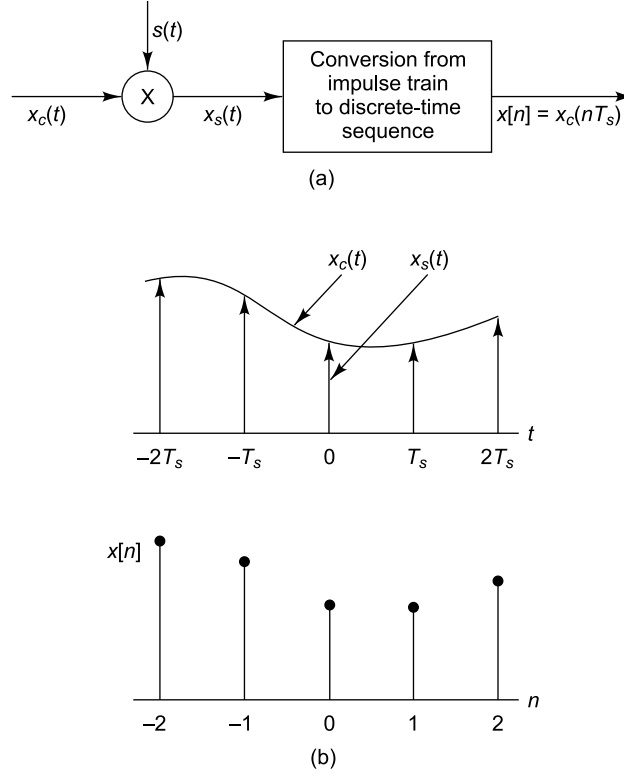
(a)

(b)

FIGURE 6.22 Sampling operation: (a) conversion from analogue signal to sequence, (b) sampling of continuous-time signal and resulting output sequence

6.6.3.1 Frequency domain representation of sampling To derive the relation between input and output of an ideal continuous-to-discrete (C/D) converter in the frequency domain, let us first consider conversion of $x_c(t)$ to $x_s(t)$ through an impulse train modulator. The modulating signal or sampling signal $s(t)$ is a periodic impulse train.

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \tag{6.58}$$

where $\delta(t)$ is the unit impulse function or Dirac delta function.

With time interval T_s between consecutive samples of the continuous time signal $x_c(t)$, the sampled time signal $x_s(t)$ can be represented as

$$x_s(t) = x_c(t).s(t)$$

$$= x_c(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s)$$

$$= \sum_{n=-\infty}^{\infty} x_c(nT_s)\delta(t - nT_s) \tag{6.59}$$

Through sifting property of the impulse function, $x_s(t)$ can be expressed as

$$x_s(t) = x_c(nT_s), \ n = \pm 1, \pm 2, \pm 3, \dots \tag{6.59a}$$

The spectrum of the sampled signal $x_s(t)$, may be obtained by taking its Fourier transform.

$$X_s(f) = \Im\{x_s(t)\} = \int_{-\infty}^{\infty} x_s(t) e^{-i2\pi ft} dt \tag{6.60}$$

It may be noted that $x_s(t)$ is obtained as the product of $x_c(t)$ and $s(t)$, which is a train of impulses. The multiplication of two functions in the time domain corresponds to the convolution of the Fourier transforms of the two functions in the frequency domain. Thus, the Fourier transform of $x_s(t)$ is the convolution of the Fourier transform of $X_c(\Omega)$ and $S(\Omega)$ which can be expressed as

$$X_s(\Omega) = \frac{1}{2\pi} X_c(\Omega) * \Im\left\{ \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \right\} \tag{6.61}$$

Since we know the spectrum of the original signal, $x_c(t)$, we need find only the Fourier transform of the train of impulses. Since $s(t)$ is an impulse train which is a periodic function in the time domain, it can be represented by a Fourier series. Thus, we may write

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) = \sum_{n=-\infty}^{\infty} c_n e^{i2\pi nt/T_s} \tag{6.62a}$$

where the Fourier series coefficients are

$$c_k = \frac{1}{T_s} \int_{-T_s/2}^{T_s/2} \sum_{n=-\infty}^{\infty} \delta(t - nT_s) e^{-i2\pi nt/T_s} dt = \frac{1}{T_s} \tag{6.62b}$$

The Fourier series of an impulse function with spacing T_s in the time domain is seen to result in an impulse train with spacing $1/T_s$ in the frequency domain. In evaluating this, the limits of integration are specified for one period only (avoiding placement of impulses at the integration limits). Thus, $s(t)$ assumes the following simplified expression, which is easily Fourier transformable.

$$s(t) = \sum_{n=-\infty}^{\infty} \delta(t - nT_s) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} e^{i2\pi nt/T_s} \tag{6.62c}$$

Since any signal can be synthesized from its inverse Fourier transform and since

$$\Im^{-1}\{\delta(\omega - \omega_0)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \delta(\omega - \omega_0) e^{i\omega t} d\omega = e^{i\omega_0 t} \tag{6.63}$$

we can express $e^{i2\pi nt/T_s}$ in Eq. (6.62c) as

$$e^{i2\pi nt/T_s} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \delta\left(\Omega - \frac{2\pi}{T_s}\right) e^{i\Omega t} d\Omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} \delta(\Omega - \Omega_s) e^{i\Omega t} d\Omega \tag{6.64}$$

Therefore, Fourier transform of Eq. (6.62c) is

$$S(\Omega) = \Im\left\{ \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \right\} = \frac{2\pi}{T_s} \sum_{k=-\infty}^{\infty} \delta\left(\Omega - k\frac{2\pi}{T_s}\right) = \frac{2\pi}{T_s} \sum_{k=-\infty}^{\infty} \delta(\Omega - k\Omega_s) \tag{6.65}$$

Thus, the Fourier transform of a periodic impulse train in the time domain is a periodic impulse train in the frequency domain.

We can now express the Fourier transform of the sampled signal as follows:

$$X_s(\Omega) = \frac{1}{2\pi} X_c(\Omega) * \frac{2\pi}{T_s} \sum_{k=-\infty}^{\infty} \delta(\Omega - k\Omega_s)$$

$$= \frac{1}{T_s} \sum_{k=-\infty}^{\infty} X_c(\Omega - k\Omega_s) \tag{6.66}$$

Equation (6.66) provides the relationship between Fourier transforms of the input and output of the impulse train modulator. It implies that the Fourier transform of $x_s(t)$ consists of periodically repeated copies of $X_c(\Omega)$, i.e. the Fourier transform of $x_c(t)$. The copies of $X_c(\Omega)$ are shifted by integral multiples of the sampling frequency and then superimposed to produce the periodic Fourier transform.

It is also to be remembered that $x_s(t)$ is the sampled signal and $x[n]$ the resulting sequence. Since $x[n] = x_c(nT_s)$, to get the Fourier transform of $x[n]$, i.e. $\Im\{x[n]\} = X(e^{j\omega})$, we have to do some normalization of the frequency axis so that frequency $\Omega = \Omega_s$ in $X_s(\Omega)$ is normalized to $\omega = 2\pi$ for $X(e^{j\omega})$. This is in view of the fact that $X(e^{j\omega})$ is periodic in ω with period 2π, while $X_s(\Omega)$ is periodic in Ω with period Ω_s. Thus, $X(e^{j\omega})$ is simply a frequency scaled version of $X_s(\Omega)$ with the scaling specified by $\omega = \Omega T_s$. Therefore,

$$X_s(\Omega) = X(e^{j\omega}) = X(e^{j\Omega T_s}) \tag{6.67}$$

Hence, $X(e^{j\omega}) = \dfrac{1}{T_s} \displaystyle\sum_{k=-\infty}^{\infty} X_c\left(\dfrac{\omega}{T_s} - k\dfrac{2\pi}{T_s}\right)$ \hfill (6.68)

This is the Fourier transform of the sequence $x[n]$.

6.6.4   Nyquist–Shannon Sampling Theorem

The sampling theorem was first proposed by Harry Nyquist in 1928, but he did not explicitly consider the problem of sampling and reconstruction of continuous signals. Later in 1949 Claude E. Shannon came up with the proof of the sampling theorem. The theorem describes two important processes: the sampling process, in which an analogue signal is converted to a discrete-time signal and the reconstruction process, in which the original analogue signal is recovered from the discrete-time signal. This theorem can be stated as follows.

Let $x_c(t)$ be a band-limited signal with $X_c(\Omega) = 0$ for $|\Omega| > \Omega_N$. Then $x_c(t)$ is uniquely determined by its samples $x[n] = x_c[nT_s]$, $n = 0, \pm1, \pm2,...$ if $\Omega_s = \dfrac{2\pi}{T_s} > 2\Omega_N$. The highest frequency in the signal Ω_N is commonly referred to as the Nyquist frequency. The Nyquist sampling theorem thus states that the sampling frequency Ω_s should be greater than twice the highest frequency Ω_N present in the signal in order to prevent aliasing from occurring and in order to be able to reproduce the original signal exactly. Given the highest frequency Ω_N in the original signal, the theorem gives the lower bound on the sampling frequency for which perfect reconstruction can be achieved. This lower bound $2\Omega_N$ is called the Nyquist rate. If instead, the sampling frequency is given, the theorem gives an upper bound for frequency components $\Omega_N < \Omega_s/2$. This upper bound is called the Nyquist frequency and is denoted as Ω_N. Another name for the Nyquist frequency is folding frequency. When the condition $\Omega_s > 2\Omega_N$ is met for the highest frequency component in the original signal, it is said that the Nyquist criterion is satisfied. The signal to be sampled must be band-limited in the sense that the spectral amplitudes of frequency components with frequencies above the cut-off frequency should be zero or close to zero, so that their

influence on the reconstructed signal will be minimal. However, in practice, a signal is never perfectly band-limited and one may be forced to use an anti-aliasing filter.

6.6.5 Mathematical Sampling and Aliasing: Harmonic Signal

Understanding what aliasing does to individual sinusoids is a big help in understanding what happens when they are superimposed over each other. Envisaging the sampling process with sinusoidal signals is relatively easy since they have only one frequency component. Figure 6.23 shows the effect of sampling a cosine signal in the frequency domain. Figure 6.23(a) shows the Fourier transform of the signal $x_c(t) = \cos(\Omega_0 t)$. Figure 6.23(b) shows the spectrum of the impulse train which does the sampling. Figure 6.23(c) shows the Fourier transform of the sampled signal $x_s(t)$ for the case of no aliasing, i.e. with $\Omega_0 < (\Omega_s/2)$. It can be seen that we have copies or replicas of the original spectrum shown in Figure 6.23(a) about $\Omega_s, 2\Omega_s, 3\Omega_s$, etc. It is seen that these replicas do not overlap, making it possible to extract the original spectrum from that of the sampled signal by passing the sampled signal through a reconstruction (low pass) filter will gain T_s and cut-off frequency $\Omega_0 < \Omega_c < (\Omega_s/2)$ as shown in Fig. 6.23(d). Thus for cases where sampling has been done with a sufficiently high sampling frequency as to not cause aliasing, the choice of Ω_c is straightforward. Figure 6.23(e) corresponds to the Fourier transform of the low pass filtered output for $\Omega_0 < (\Omega_s/2) = (\pi/T_s)$ with $\Omega_c = (\Omega_s/2)$. Figure 6.23(f) shows the Fourier transform of $x_s(t)$ for $\Omega_0 > (\Omega_s/2)$, i.e. with aliasing arising from undersampling. It can be seen that the replicas overlap in this case and hence the original signal spectrum cannot be retrieved using a low-pass filter as shown in Fig. 6.23(g) which corresponds to the Fourier transform of the low-pass filtered output for $\Omega_0 > (\pi/T_s)$ with $\Omega_c = (\Omega_s/2)$. The reconstructed output is $x_r(t) = \cos(\Omega_0 t)$ without aliasing (Fig. 6.23d) and with aliasing the reconstructed output is $x_r(t) = \cos(\Omega_s - \Omega_0)t$ as seen in Fig. 6.23(g), i.e. the frequency Ω_0 has taken on the identity or alias of the low frequency signal $(\Omega_s - \Omega_0)$ as a consequence of the sampling and reconstruction processes.

6.6.6 Band-Limited Sampling and Aliasing

In practice, signals are quite often band-limited signals with a definite cut-off frequency. To avoid aliasing in the sampled band-limited signal, we must ensure that the spectrum replicas do not overlap. Figure 6.24 shows the sampling and reconstruction processes for a band-limited signal $x_c(t)$. Figure 6.24(a) shows the Fourier transform of $x_c(t)$, where the highest non-zero frequency component in $X_c(\Omega)$ is Ω_N. Figure 6.24(b) represents the periodic impulse train $S(\Omega)$. Figure 6.24(c) shows $X_s(\Omega)$ obtained as a result of convolving $X_c(\Omega)$ with $S(\Omega)$. From this figure, it is evident that when $\Omega_s - \Omega_N > \Omega_N$ or $\Omega_s > 2\Omega_N$, the replicas of $X_c(\Omega)$ do not overlap and therefore when they are added together in Eq. (6.66), there is formed (to within a scale factor of $1/T_s$), a replica of $X_c(\Omega)$ at each integral multiple of Ω_s. Therefore, $X_c(\Omega)$ can be recovered from $X_s(\Omega)$ through the process $X_r(\Omega) = H_r(\Omega) X_s(\Omega)$ where $H_r(\Omega)$ is an ideal low pass filter having frequency response as shown in Fig. 6.24(d) and with cut-off frequency Ω_c such that $\Omega_N < \Omega_c < (\Omega_s - \Omega_N)$. The recovered signal $X_r(\Omega)$ would be as shown in Fig. 6.24(e), where it is assumed that $\Omega_s > 2\Omega_N$. If this inequality does not hold good, i.e. if $\Omega_s < 2\Omega_N$, the copies of $X_c(\Omega)$ overlap, so that when they are added together, $X_c(\Omega)$ is no longer recoverable by low pass filtering. This is illustrated in Fig. 6.24(f). In this case, the reconstructed output $x_r(t)$ has got aliased.
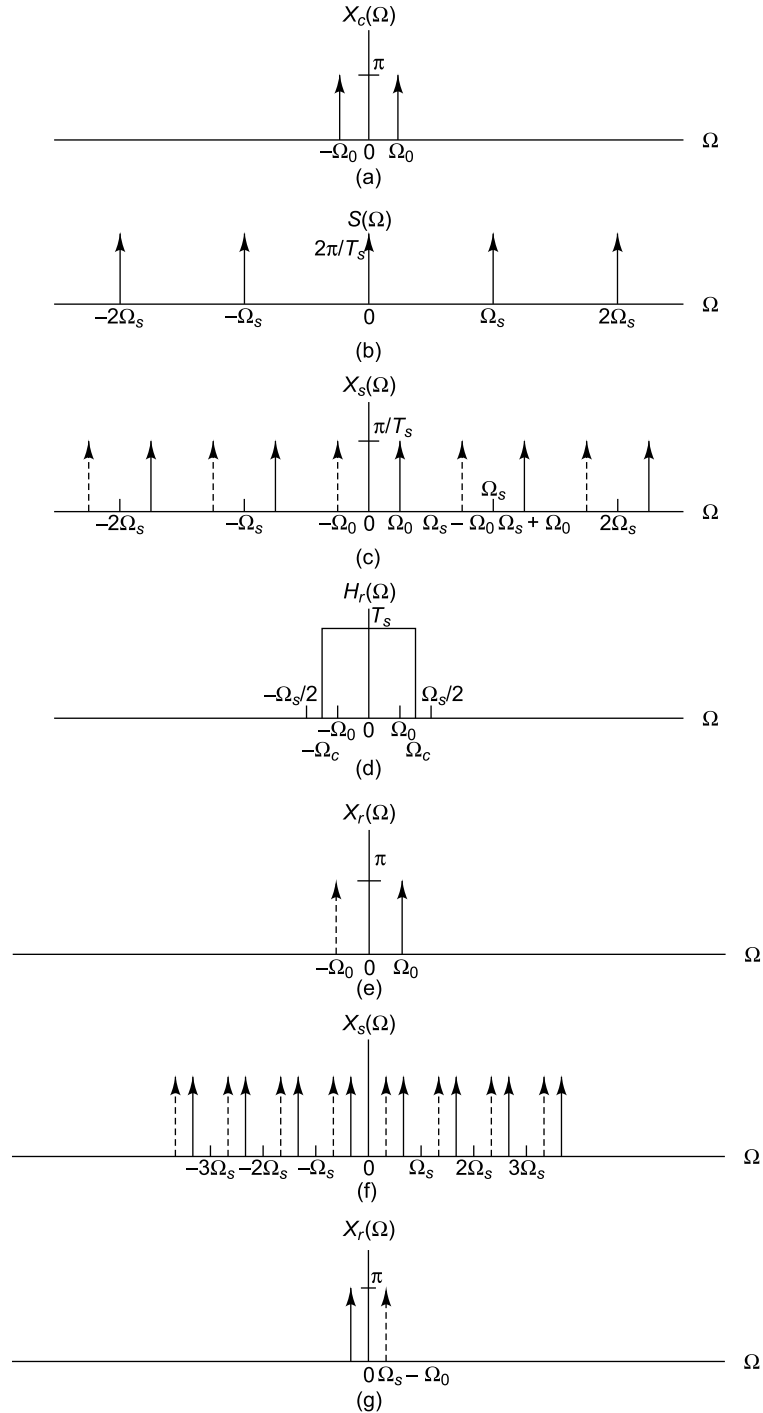
Figure 6.23   Sampling and aliasing for a cosine signal: (a) $X_c(\Omega)$, (b) $S(\Omega)$ with $\Omega_0 < (\Omega_s/2)$, i.e. no aliasing, (c) $X_s(\Omega)$ with $\Omega_0 < (\Omega_s/2)$, i.e. no aliasing, (d) $H_r(\Omega)$ with $\Omega_0 < \Omega_c < (\Omega_s/2)$, (e) $X_r(\Omega)$ for the case without aliasing, (f) $X_s(\Omega)$ with $\Omega_0 > (\Omega_s/2)$, i.e. with aliasing, (g) $X_r(\Omega)$ for the case with aliasing
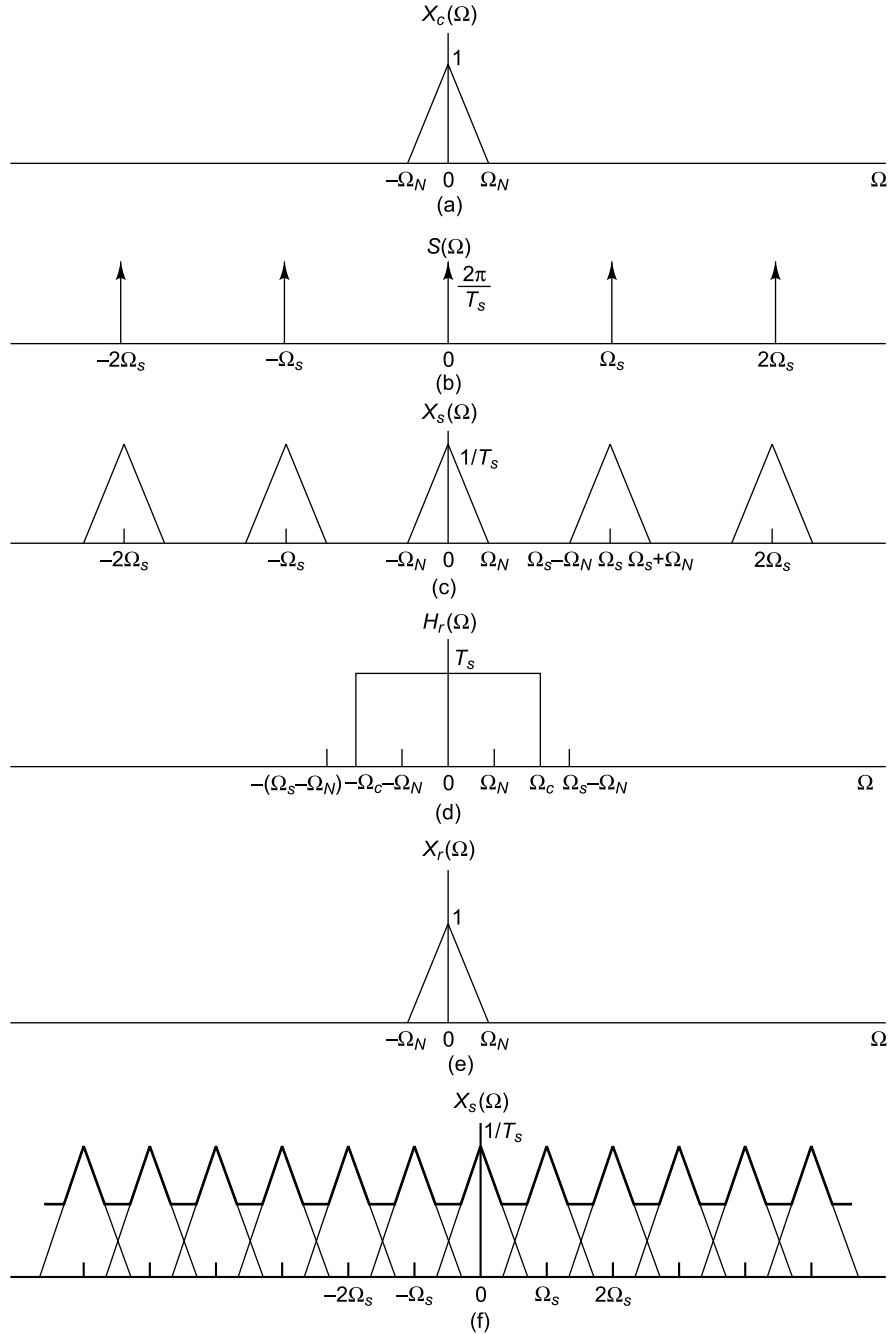
Figure 6.24 Depiction of sampling and reconstruction for a band-limited signal: (a) $X_c(\Omega)$, (b) $S(\Omega)$ with $\Omega_N < \dfrac{\Omega_s}{2}$, i.e. no aliasing, (c) $X_s(\Omega)$ with $\Omega_N < \dfrac{\Omega_s}{2}$, i.e. no aliasing, (d) $H_r(\Omega)$ with $\Omega_N < \Omega_c < (\Omega_s - \Omega_N)$ (e) $X_r(\Omega)$ for the case without aliasing, (f) $X_s(\Omega)$ with $\Omega_s < 2\Omega_N$, i.e. with aliasing

6.6.7 Reconstruction of Original Signal Using Reconstruction Filters

Reconstruction of the original signal is possible if sampling has been done at a sufficiently high frequency in the first place without leading to aliasing and provided that we know the sampling period. Reconstruction is an interpolation process that mathematically defines a continuous time signal $x_s(t)$ from the discrete samples of sequence $x[n]$ and at instants of time between the sampling instants nT_s. Given the sequence $x[n]$, it is possible to form the impulse train $x_s(t)$ in which the impulses have areas equal to the successive sequence values.

$$x_s(t) = \sum_{N=-\infty}^{\infty} x[n]\delta(t - nT_S) \tag{6.69}$$

The nth sample is associated with the impulse at $t = nT_s$, where T_s is the sampling period and $\Omega_s = \dfrac{2\pi}{T_s}$ is the sampling frequency. If the impulse train is fed to an ideal low pass continuous-time filter with impulse response $h_r(t)$ and frequency response $H_r(\Omega)$, the output is

$$x_r(t) = \sum_{n=-\infty}^{\infty} x[n]h_r(t - nT_s) \tag{6.70}$$

Reconstruction of a band-limited signal from its samples $x[n]$ is done as shown in Fig. 6.25.
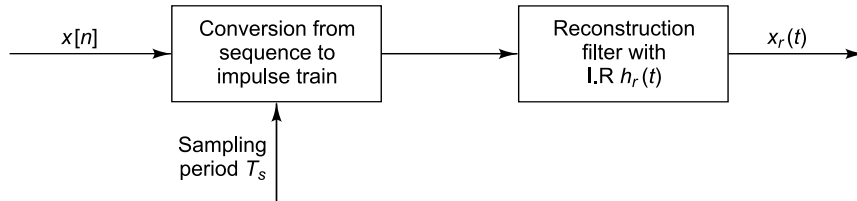

<div align="center">

FIGURE 6.25   Reconstruction process

</div>

This ideal reconstruction (low-pass) filter has a gain T_s to compensate for the factor $1/T_s$ incurred during the sampling process. It is a rectangular function in the frequency domain with a cut-off frequency $\Omega_N < \Omega_c < \Omega_s - \Omega_N$. This corresponds to a sinc function in the time domain as shown below:

$$h_r(t) = \frac{\sin(\pi t / T_s)}{\pi t / T_s} \tag{6.71}$$

Figure 6.26 shows the impulse and FRFs of the reconstruction filter.
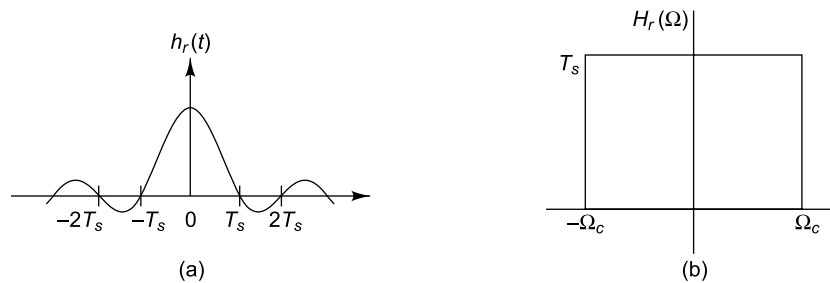

<div align="center">

FIGURE 6.26   Ideal reconstruction filter: (a) $h_r(t)$, (b) $H_r(\Omega)$

</div>

This function $h_r(t)$ is of infinite duration in the time domain, which means that it is impossible to have a perfect reconstruction filter in this domain. Substituting for $h_r(t)$ from Eq. (6.71) into Eq. (6.70), we get

$$x_r(t) = \sum_{n=-\infty}^{\infty} x[n] \frac{\sin\left[\pi(t - nT_s)/T_s\right]}{\pi(t - nT_s)/T_s} \tag{6.72}$$

Thus the ideal reconstruction process is obtained by multiplication in the frequency domain of $X_r(\Omega)$ and $H_r(\Omega)$ or the convolution in the time domain as shown by Eq. (6.72) and involves superposition of scaled and delayed sinc functions as depicted in Fig. 6.27. The scaled and time-shifted sinc functions are continuous, making the superposition of these also continuous. Ideal reconstruction cannot be realized in practice since it implies that each sample contributes to the reconstructed signal at almost all instants of time, requiring summing an infinite number of terms. Instead, some type of approximation of the sinc function, which is finite in length, has to be used and leads to an interpolation error.

6.7 THE FAST FOURIER TRANSFORM

The continuous Fourier transform converts a continuous-time signal of infinite duration into a continuous frequency spectrum with an infinite number of sinusoidal components as we have seen before. A discrete Fourier transform (DFT) on the other hand, deals with sequences which are of finite duration and periodic and which are obtained by sampling a continuous-time signal at regular intervals. This is similar to the description given in Section 6.5.3, except that the sequence in this case is assumed to be periodic and of finite duration. It is assumed that the sequence is obtained from a continuous-time signal $x(t)$ of duration

FIGURE 6.27 Reconstruction of band-limited signals

T and that the data are taken with a sampling period of T_s and that sampling has been done without aliasing. Thus $T = NT_s$ where N is the total number of samples taken in the time duration $[0, T]$. Due to this process is implicit the assumption that the data being analysed have a fundamental frequency of $1/T$ Hz; i.e. the data are periodic in the time window of length T seconds. It is to be noted that this is not usually true with real data. The signal $x(t)$ can be represented as

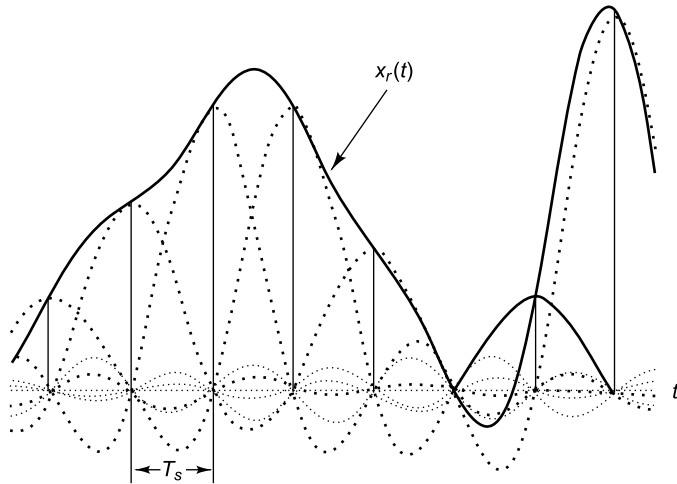$$x(t) = \sum_{n=0}^{N-1} x(nT_s)\delta(t - nT_s) \tag{6.73a}$$

The Fourier transform for this signal may be written as

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} dt$$

$$= \sum_{n=0}^{N-1} x(nT_s) e^{-i2\pi fnT_s} \tag{6.73b}$$

The computation is restricted to N frequencies: $0, \dfrac{1}{T_s}, \dfrac{2}{T_s}, \dfrac{3}{T_s}, ..., \dfrac{N-1}{T_s}$. Setting $f = \dfrac{k}{T} = \dfrac{k}{NT_s}$ in the above equation, we obtain the kth element in the DFT equation as

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-i(2\pi/N)kn}, \ k = 0, 1, 2, 3, ..., N-1$$

$$= \sum_{n=0}^{N-1} x[n] W_N^{kn} \tag{6.74a}$$

where $W_N = e^{-i(2\pi/N)}$ \hfill (6.74b)

Thus N samples of the sequence $x[n]$ are taken and transformed into N complex Fourier coefficients (at integral multiples of the base frequency $1/T$ Hz). Though $X[k]$ itself is approximate, the original sequence $x[n]$ can be got from it exactly, allowing the reconstruction of the original waveform.

The nth element in the inverse DFT (IDFT) is computed as

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{i(2\pi/N)kn}, \ n = 0, 1, 2, 3, ..., N-1$$

$$= \frac{1}{N} \sum_{k=0}^{N-1} X[k] W_N^{-kn} \tag{6.75}$$

Equation (6.74) may be simplified as

$$X[k] = \sum_{n=0}^{N-1} (\text{Re}\{x[n]\} + i\,\text{Im}\{x[n]\})(\text{Re}\{W_N^{kn}\} + i\,\text{Im}\{W_N^{kn}\}), \ k = 0, 1, 2, 3, ..., N-1$$

$$= \sum_{n=0}^{N-1} (\text{Re}\{x[n]\}\,\text{Re}\{W_N^{kn}\} - \text{Im}\{x[n]\}\,\text{Im}\{W_N^{kn}\} + i(\text{Im}\{x[n]\}\,\text{Re}\{W_N^{kn}\} + \text{Re}\{x[n]\}\,\text{Im}\{W_N^{kn}\})) \tag{6.76}$$

It is to be noted that here both $x[n]$ and $X[k]$ are periodic with period N. The 0 to $N-1$ values of the sequence in the time domain are called points and in the frequency domain they are known as lines. Inspection of the equation reveals that even for the computation of one line in the frequency domain, a very large number of complex additions and multiplications are involved; the larger the value of N, the larger are the two.

Cooley and Tukey in 1965 proposed an algorithm using the symmetry and periodicity properties of W_N^{kn}. This algorithm called the FFT led to a considerable reduction in computational effort in doing a DFT.

The complex conjugate symmetry of W_N^{kn} gives

$$W_N^{k(N-n)} = W_N^{-kn} = (W_N^{kn})^* \tag{6.77a}$$

The property of periodicity in n and k results in the equation

$$W_N^{kn} = W_N^{k(n+N)} = W_N^{n(k+N)} \tag{6.77b}$$

Using the above properties and what is called a decimation-in-time procedure, the number of complex multiplications and additions required for an N point DFT computation gets reduced to $N \log_2 N$ as compared to N^2 multiplications and additions by brute force. This has made real time analysis highly feasible.

It should be borne in mind that the original time series consisting of N points, when Fourier transformed, generates $N/2$ unique frequency coefficients only, from 0 to $(N/2)-1$. Curiously, one started with N points; now there are only $N/2$ points. Since each Fourier component is described by the equation

$$X[k] = A_k + iB_k \tag{6.78}$$

the total number of descriptive frequency components is N and thus no information has been lost, only transformed into a new domain.

6.8 FFT ANALYSER SETUP

6.8.1 Setup for Spectrum Collection

The spectrum option available in FFT analysers is such as to cater to a relatively wide range of frequencies; hence the analyser is forced to acquire a short time record. While deciding on the analyser settings, a good rule of thumb is that the time record length depends only on the line spacing of the FFT spectrum (Example 1) and the sampling rate depends only on the frequency range of the FFT spectrum (Example 2), and they are independently adjustable. The sampling frequency of the time record for many analysers is 2.56 times the highest frequency in the spectrum. Thus a frequency span of 1000 Hz requires a sampling rate of 2560 samples per second. As the size of the transform N is generally fixed (though not always) for a given analyser, the frequency range covered and the resolution are determined solely by the time length of each sample. Let us consider two examples.

Example 6.1: A 400-line spectrum extending from DC to 1000 Hz will have a line spacing of 1000/400, or 2.5 Hz. The length of the time record used to calculate this spectrum is 1/2.5, or 0.4 s. The frequency resolution or the line spacing of a spectrum is thus the reciprocal of the total time duration of the time domain record.

Example 6.2: In many analysers the size of the transform N is a power of 2 such as 512, 1024, etc. Let us consider the sampling of a continuous time function with a sampling frequency of 1 kHz or with $T_s = 0.001$ s, and number of data points in the analyser = 1024. The total duration of the time record is $T = NT_s = (1024)(0.001s) = 1.024s$. The spectral resolution is equal to the base frequency of the approximate Fourier series $1/T = 1/1.024 = 0.9766$ Hz.

6.8.2 Setup for Analysing Time Waveforms

When setting up an analyser to store time records or waveforms, an important point to be considered is that the frequency range normally convenient for looking at a spectrum is usually not suitable for looking at the waveform. Most FFT analysers, with the exception of PC based analysers, do not allow one to setup specific sampling rates or time domain record lengths; the setup is typically in terms of the frequency span and the frequency resolution.

Considering Example 6.1, the time record will show details that happen in a 0.4 s time span, but in practice, when looking at a machine vibration waveform, we often have to look at events occurring over a much longer duration. For instance if we have to look at beat phenomena in the vibration signature of a machine, or of the combined vibration of two machines running at slightly different speeds, the duration of the waveform required would be at least several seconds long and hence it has to be ensured that there is enough memory available in the data collector to store the waveform. It is best to use the lowest sampling rate and the shortest time record length that will provide the needed data. To acquire a waveform lasting 5 s, a setup with a line spacing of 1/5 Hz is required, and this can be done by adjusting the number of lines of resolution and the frequency span to convenient values.

If we want to observe beats in a waveform that occur only once in several seconds, the sampling rate needed will not be very high, something like 50 samples per second probably being sufficient. This corresponds to a frequency range of 50/2.56 (assuming a sampling frequency 2.56 times the highest frequency), or 19.53 Hz; thus the 20 Hz range in the frequency analyser could be chosen.

On the other hand, if it is required to capture a waveform that might have glitches at 50 times per second, sampling has to be done fast enough to resolve each glitch. Sampling could be done at 1000 samples per second, necessitating a frequency span of 1000/2.56, or about 390 Hz. Suppose as another example a pinion is rotating at 1800 RPM (30 Hz) and it is required to capture one revolution (which will take 1/30th of a second) for purposes of time synchronous averaging (please refer Section 8.6.1). To setup an FFT analyser to collect a record of length 1/30th second, its frequency resolution, or line spacing, must be 30 Hz. In order to acquire a little more than one revolution, 1/25th second of time domain data might be acquired. There are many combinations of frequency range and resolution that could be used to do this: a span of 5000 Hz with 200 lines resolution or a span of 2500 Hz and 100 lines of resolution, or a span of 250 Hz at 10 lines resolution.

6.9　Leakage and Windowing

One of the main limitations in the computation of the DFT is that the data record using which computation is done is of finite length. Besides, the assumption behind the computation of the DFT is that the signal is periodic in the finite length record. However in actual practice, we perform FFT on signals that are not repetitive or periodic in the window considered, violating the basis behind the DFT computation. Only in artificially contrived situations is it possible to have the data processed periodic in the data window. Due to the violation of this assumption, there is leakage of energy out of one resolution line of the spectrum to other lines, leading to significant broadening of the 'theoretical' line spectra. Figure 6.28(a) shows a sinusoidal wave that is exactly periodic in the window and one that is not periodic in the time record. The corresponding spectrum takes the form of a line for the former as seen in Fig. 6.28(b), while for the latter it is seen that the discontinuities near the ends of the window result in a high frequency content which is not present in the original signal, leading to smearing or spreading which can clearly be seen in the spectrum. This smearing of energy through the frequency domain is known as leakage.
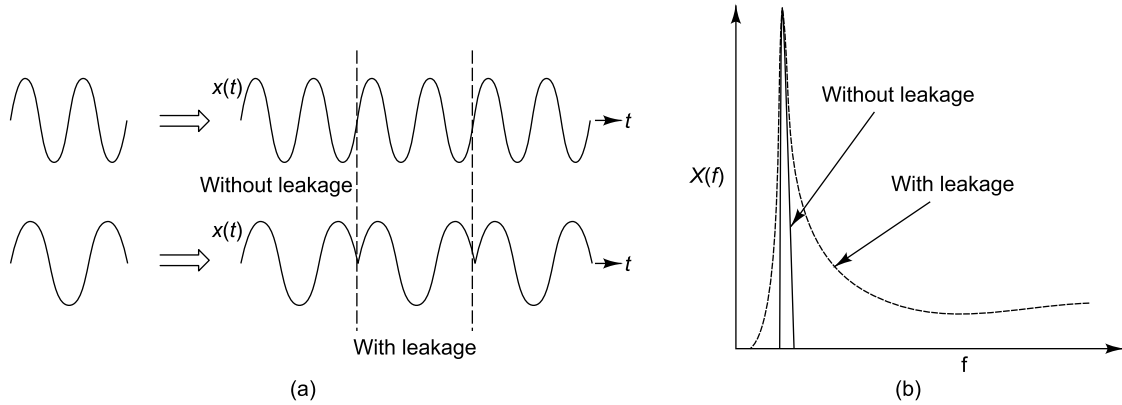
FIGURE 6.28 Phenomenon of leakage: (a) time domain data, (b) spectra

Let us recollect that leakage is due to the fact that a finite time record with a non-integral number of cycles has been taken for computation of the FFT. A sine wave will have a single line spectrum only if it exists for all time, from minus infinity to plus infinity, or if the time record considered for computation of the spectrum has an integral number of cycles in the window so that it would exactly match the actual input waveform. However, if we are measuring a continuous signal like a sine wave from a finite time record, leakage is caused if the input is not periodic in this record. The problem of leakage is severe enough to entirely mask small signals close to the sine wave under consideration. Obviously leakage is not a problem with transient signals that die down before the end of the time window as shown in Fig. 6.29. Any function like a transient which does not require a window because it occurs completely within the time record is called a self-windowing function. There are many examples of self-windowing functions, such as impacts, shock signals, damped vibration response, sine bursts, noise bursts, chirp bursts and pseudo-random noise. In vibration measurement, signals which are much more complex than a sinusoid of a single frequency occur and the effect of leakage can be quite severe. One would be tempted to conclude that the FFT is not a very useful spectrum analysis technique due to the associated problem of leakage. However there is a solution to this problem and this is known as windowing.
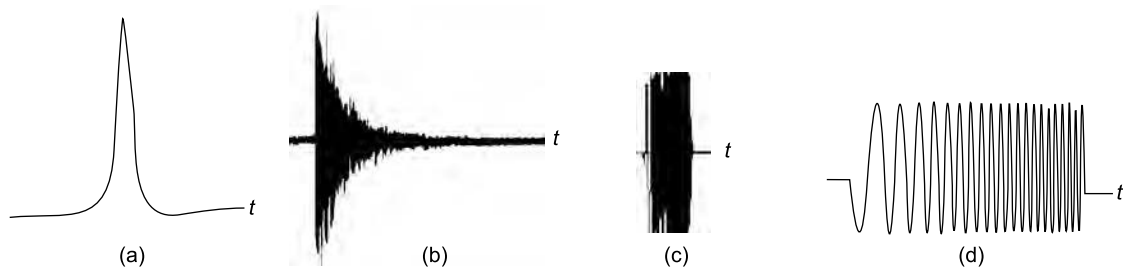

FIGURE 6.29 Self-windowing signals: (a) impact, (b) shock response, (c) noise burst, (d) chirp sine

Windowing involves the imposition of a prescribed profile on the time record before computing the Fourier transform and this is done by multiplying the time record by a window function of known shape and which is typically zero at the ends of the record. If the time record, the FFT of which is to be obtained, is denoted as $x(t)$ and the profile of the window function is depicted as $w(t)$, then the windowed

or analysed signal is $x_w(t) = x(t)w(t)$. Due to windowing, the signal is forced to be periodic in the analysis window as shown in Fig. 6.30. The envelope depicts the window function $w(t)$ and $x_w(t)$ is the windowed function.

Though the signal has been forced to be periodic as a result of windowing, the amplitudes, especially near the ends of the record, have been modified due to the modulation or multiplication process involved and the resultant FFT will yield improper spectral amplitudes. Hence, the FFT analyser has to rescale the FFT output to correct for the smaller spectral amplitudes obtained. Analysers
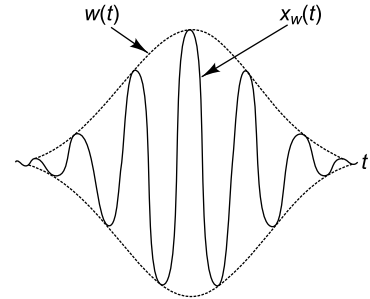
FIGURE 6.30   Windowed signal

are generally provided with a number of time domain windows such as Hanning, Hamming, tapered cosine, rectangular, flat-top, force-response windows, etc. One should select the window best suited to his/her application. Figure 6.31 shows the different types of windows available. The Hanning or Cosine taper windows are typically used for steady periodic or random signals, while the exponential window is used for transient vibration records which contain quite a bit of information concentrated in the initial part of the time record and would thus lose information by a choice of other windows.
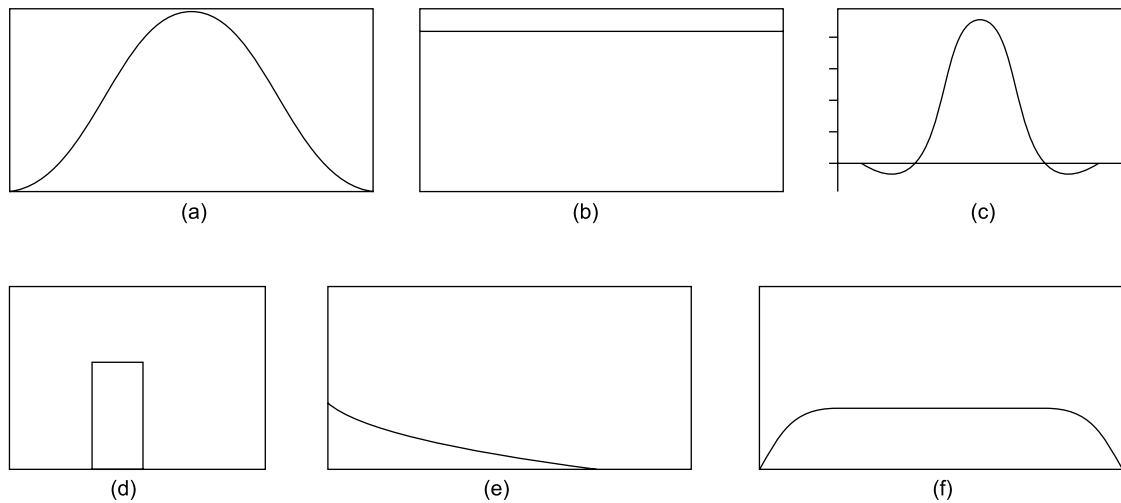

FIGURE 6.31   Different types of windows (as a function of time): (a) Hanning, (b) rectangular, (c) flat top, (d) force, (e) response (exponential), (f) tapered cosine.

6.9.1   Hanning Window

There are a large number of functions available to window data, but the most commonly used one is called the Hanning window (Fig. 6.31a). Generally the Hanning window does an exceptionally good job on random signals, and a reasonably good job on sinusoidal signals, irrespective of whether they are periodic or non-periodic in the time record. The problem with the Hanning window is that spectral amplitudes get changed. The window makes the transient look more like a sine wave than a transient. Therefore, for transients the Hanning window is not a good choice. For such signals it is preferable to

give equal weightage to all the data in the time record. Hence, it is preferable to use the uniform or rectangular window which weights the entire time record uniformly.

6.9.2 Rectangular/Uniform Window

The problem in using a Hanning window with a transient signal has just been discussed. Hence, for transients, we might prefer to use a uniform or rectangular window which does not attenuate the signal at its ends. The rectangular window (Fig. 6.31b) is thus used for self windowing functions, the values of which are zero near the ends of the window. Self-windowing functions are often used as excitation signals in measuring the FRFs of systems, especially if the system is lightly damped with a sharp peak at resonance, the reason being that such functions generate no leakage in the FFT. Without a self-windowing excitation, energy could leak from a lightly damped resonance into adjacent lines in the spectrum, with the result that the FRF obtained would show greater damping than actually exists. Generally a hammer with a built-in force transducer is used to excite a structure for the measurement of FRFs. The time record of the exciting force should just correspond to the transient impact imparted to the structure and should have the form of a half-sine wave, which is self-windowing and dies out within the time window of analysis. However, in practice it is seen that the movement of the hammer before and after the impact causes stray noise signals in the time record and one way of eliminating this noise is through the use of a rectangular force window shown in Fig. 6.31. This window is unity where the impact data is valid and zero everywhere else so that the analyser does not measure any stray noise.

6.9.3 Flattop Window

We have seen that the Hanning function gives the filter a very rounded top. In the signal being analysed, if most of the signal is in the middle of the time record, then it will be measured accurately using the Hanning window; if not, the modulating effect of the window will attenuate the signal by up to 1.5 dB (16%). This error is unacceptably large if we are trying to measure the amplitude of a signal accurately. This deficiency is overcome by choosing a window function which has a flatter pass band (in the frequency domain). The flat-top window has exactly this nature and helps avoid the distortion of spectral amplitudes caused by the Hanning window. The error in amplitude resulting from this window does not exceed 0.1 dB (1%), a reasonably good improvement over the Hanning window. The Hanning and flattop window functions enable the analyst to choose between increased accuracy or improved frequency resolution. Figure 6.31(c) shows the flattop window.

6.9.4 Force Window

A force window (Fig. 6.31d) is essentially a rectangular window of adjustable width which is used to window the impact force signal in experimental modal analysis. This window does not alter the actual force pulse, but helps minimize the noise in the rest of the window and thus improves the SNR. It also removes the pulse due to a double hit. Data capture may be effected by appropriate triggering.

6.9.5 Exponential/Response Window

The exponential window (Fig. 6.31e) is especially suited for processing free vibration responses. During impact excitation, the output of the force transducer is connected to one channel of the signal analyser

and the output of the response transducer to the second channel. It is known that the response of a structure is zero at the beginning of the time window before the impact excitation is imparted. Hence, the response window that is applied need not force the function there to zero. Besides, most of the structural response is obtained at the beginning of the time record; hence we have to ensure that the signal is heavily weighted by the response window function here. However the response has to be forced to zero near the end of the time record and an exponentially weighted window called the response window is used. Generally three or four decay rates are available and may be chosen depending on the inherent damping in the system. Unlike the Hanning window, the response window is thus not zero at both ends of the time record. A response window imposed on the response of a lightly damped structure which has not fully decayed by the end of the time record gives an impression that the system has more damping than it really does. The effect of this additional damping due to the exponential window should be corrected to get a true estimate of modal damping.

6.9.6 Tapered Cosine Window

The cosine tapered (Tukey) window (Fig. 6.31f) sets the data at the ends of the window smoothly to zero without significantly reducing the gain of the transform of the windowed signal.

6.10 AVERAGING

A very important feature of digital spectral analysis concerning the specific requirements for processing random signals is averaging. A pure random signal never repeats in the measurement window and therefore excites the shaker differently during each measurement period. When analysing random vibration signals, the Fourier transforms cannot be computed, since they do not exist for a random process and we must instead obtain the power spectral densities and correlation functions which are used to characterize such signals (these aspects are described in Chapter 7). In order to reduce the statistical variance of a measured power spectral density with a random excitation function such as pure random excitation and also to reduce the effects of non-linearities, it is necessary to employ an averaging process. Then the distortion, which shows up as random noise in each power spectrum, can be averaged out of the measurement, just like any other type of extraneous noise. When a pseudo random excitation is used, this distortion noise cannot be averaged out since all the power spectra are exactly alike, i.e. they contain the same distortion components at the same frequencies. The choice of the number of averages required is decided by the statistical reliability required and the degree to which spurious random noise is to be removed from the signals. For valid random signal processing, the following condition should be satisfied.

$$2\,BT \geq K \tag{6.79}$$

where B is the frequency bandwidth, T the total time encompassing the data and K a constant. If K is at least 10, then there is an 80% probability that the estimated spectrum lies between 0.5 and 1.5 times the true value. If K approaches 100, then reasonably reliable estimates are got, in the sense that, there is an 80% probability that the measured value is within 18% of the true value. The larger the averaging time, the smaller are the variations in the mean and the better the spectral estimates, but the longer is the processing time.

There are two types of averaging, sequential and overlap averaging. In the former, m samples, each of duration T, and mutually exclusive are averaged as shown in Fig. 6.32(a). However, since modern analysers

compute the DFT in an extremely short duration of time, a new transform could be computed faster than a new complete sample record has been collected. Under such conditions, it is convenient to perform a second FFT computation as soon as possible, using the most recent N data points, even though some of them may have been used for computation of the previous transform. This overlap averaging process is depicted in Fig. 6.32(b). It is evident that 100 averages performed using overlap averaging will not have the same statistical properties as would 100 completely independent samples.
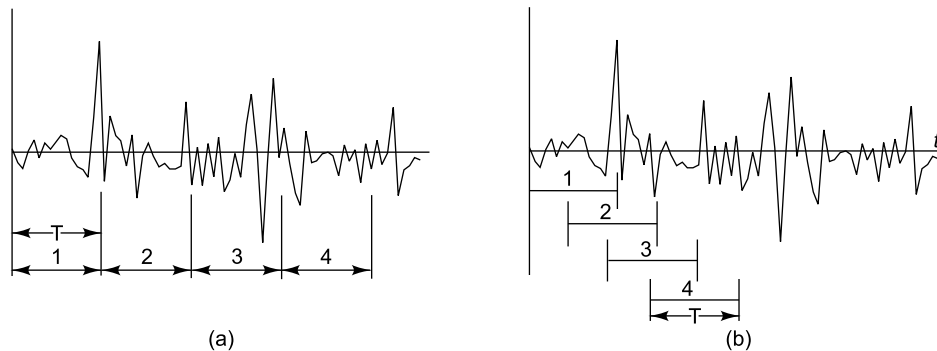
FIGURE 6.32   Averaging process: (a) sequential averaging, (b) overlap averaging

6.11   ZOOM

So far we have discussed the basic DFT; this however is often found to have limitations of inadequate frequency resolution, especially at the lower end of the frequency spectrum and especially, for lightly damped systems. This difficulty arises on account of the limited number of discrete points available (N) for FFT computation, the maximum frequency range to be covered and/or the length of time record available. One solution to this problem would be to use a larger N in the computation of the transform. Though this gives finer frequency resolution around the regions of interest, it has the drawback that it is forced to compute more information than is required. This problem can be overcome by using the zoom option available in spectrum analysers. This allows the user to utilize all the spectral lines available (1024 or 2048) in a narrow frequency range between f_{min} and f_{max} (instead of between 0 and f_{max} without zooming). Thus a zoom × 2 will give a resolution double that obtained without zooming and a zoom × 16, 16 times the resolution without zooming. Under such circumstances it is advisable to have both the response and excitation signals subjected to the same amount of zooming to compute FRFs. When using the zoom feature to measure the FRF in a narrow frequency range, it is to be ensured that there is practically no vibration energy outside the frequency range of interest. Hence, the excitation energy supplied to the structure should be band-limited to the analysis range. This feature is not available on all analysers.

FURTHER READINGS

1. Bendat, J.S., *Engineering Applications of Correlation and Spectral Analysis*, 2nd ed., Wiley-Interscience, New York, 1993.
2. Bendat, J.S. and Piersol, A.G., *Random Data: Analysis and Measurement Procedures*, 3rd ed., Wiley-Interscience, New York, 2000.

3. Carlson, G.E., *Signal and Linear System Analysis*, John Wiley and Sons Ltd., New York, U.S.A, 1998.

4. Ewins, D.J., *Modal Testing: Theory and Practice*, Research Studies Press Ltd., Somerset, England, 1984.

5. Haykin, S. and Kosko, B., *Intelligent Signal Processing*, IEEE Computer Society Press, New York, 2001.

6. Hsu, H.P., *Signals and Systems, Schaum's Outline Series*, McGraw-Hill, New York, 1993.

7. Lathi, B.P., *Linear Systems and Signals*, 2nd ed., Oxford University Press, USA, 2005.

8. Lindner, D.K., *Introduction to Signals and Systems*, McGraw-Hill Education, Asia, 1999.

9. Lyons, R.G., *Understanding Digital Signal Processing*, Addison-Wesley, Reading, MA, 1997.

10. Mitchell, L.D., Signal processing and the Fast Fourier Transform (FFT) analyzer, *International Journal of Analytical and Experimental Modal Analysis*, 1(4), 1986.

11. Northrop, R.B., *Introduction to Instrumentation and Measurements*, CRC Press, Boca Raton, FL, 1997.

12. Oppenheim, A.V. and Schafer, R.W., *Discrete-Time Signal Processing*, Prentice-Hall, Upper Saddle River, NJ, 1999.

13. Oppenheim, A.V., Hamid, S. and Willsky, A.S., *Signals and Systems*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1996.

14. Proakis, J.G. and Manolakis, D.G., *Digital Signal Processing: Principles, Algorithms, and Applications*, Prentice-Hall, Englewood Cliffs, NJ, 1996.

15. Rabiner, L.R. and Gold, B., *Theory and Application of Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

16. Rangan, C.S., Sarma, G.R. and Mani, V.S.V., *Instrumentation Devices and Systems*, Tata McGraw Hill Publishing Company Ltd., New Delhi, 1983.

17. Rao, R.P., *Signals and Systems*, Tata McGraw Hill Publishing Company Limited, New Delhi, 2008.

18. Robson, R.D., *An Introduction to Random Vibration*, Edinburgh University Press, Elsevier Publishing Company, Amsterdam (1963).

19. Smith, S.W., *The Scientist and Engineer's Guide to Digital Signal Processing*, www.dspguide.com.

20. Stanley, W.D., Dougherty, G.R. and Dougherty, R., *Digital Signal Processing*, Reston Pub. Co., Virginia, U.S.A, 1984.

21. Tocci, R.J. *Digital Systems: Principles and Applications*, Prentice Hall of India, New Delhi, 1988.

22. Ziemer, R.E., Tranter, W.H. and Ronald Fannin D., *Signals and Systems: Continuous and Discrete*, Prentice-Hall of India Pvt. Ltd., New Delhi, 2002.

Chapter
7

Dealing with Random Signals

7.1 INTRODUCTION

For deterministic processes, there are mathematical models establishing a functional relationship between different physical variables, such that for specific values of the relevant variables of the physical phenomenon, a particular state of the process is obtained. In the case of a random process, however, the different states cannot be predicted exactly, but can only be predicted with certain measures of likelihood. A deterministic signal is one which has a well defined waveform and which can be exactly resynthesized from frequency spectra, i.e. the magnitudes and relative phases of the synthesising sinusoidal waves. The value of a random signal or waveform, however, is not specified at every instant of time and it is not possible to predict its future values with certainty on the basis of the past values. Moreover, in some cases, the physical laws governing a phenomenon may be so complex that a detailed analysis becomes intractable, or there may be some inherent uncertainties associated with the phenomenon. A probabilistic model is therefore used to incorporate these uncertainties in a rational and systematic manner. In fact, it can be said that a probabilistic model is generally more realistic as compared to a deterministic model. While a deterministic model can be used to predict the definite state of a phenomenon, the probabilistic model helps investigate the inherent unavoidable and unexplained variations observed in almost all physical phenomena. It quantifies these variations and may also provide clues to the causes of the variations. This model may be validated by statistical studies on available observations of the random phenomenon.

7.2 MODELLING OF RANDOM PROCESSES

By definition, a random process is a family of random variables. The random processes occurring in most engineering applications are either functions of time, or space or both. The theory of random processes has been applied to many problems in physics and engineering involving uncertainty like kinetic theory of gases, statistical mechanics, Brownian motion, statistical communication theory, turbulent fluid flow and earthquake engineering. Some of the typical random variables that we come across are noise of a jet engine, height of waves in a choppy sea, road undulations, vibration at the driver's seat of a vehicle, pressure gusts encountered by an airplane in flight, ground motion during an earthquake, electrical output of a noise generator, acoustic pressures generated by air rushing through a pipe, microroughness of a machined part, etc.

There are essentially three approaches to stochastic modelling of random processes. The first approach is by modelling them as probabilistic processes defined by Gaussian or Poisson or any of the other

numerous distributions available and estimating the parameters of the distributions from the actual records available. In the second method, the sample functions of the random process are modelled as the outputs of appropriately designed filters to white noise excitation. The filter weights, which are the parameters of the model, are estimated by making use of the actual observations of the random phenomena. The third approach is non-parametric in nature, and here the process is described by higher order statistics. In some cases, the stochastic model is simply assumed to be a mathematically convenient process like the Gaussian process, with little regard to the information available about the actual process.

Random signals may be expressed as a function of any variable. For vibration or acoustic applications, they are expressed as a function of time t or space s, e.g. road roughness may be expressed as $x(t)$ in temporal form, or $x(s)$ in the spatial domain. The output of a linear time-invariant system to a random input is also random. Thus, for the road roughness described above, vehicle response may be expressed as $y(t)$ or $y(s)$. Figure 7.1 shows a typical 'sample record' or 'sample function' $x(t)$, which is nothing but a single time history representing a random phenomenon (observed over a finite time interval).
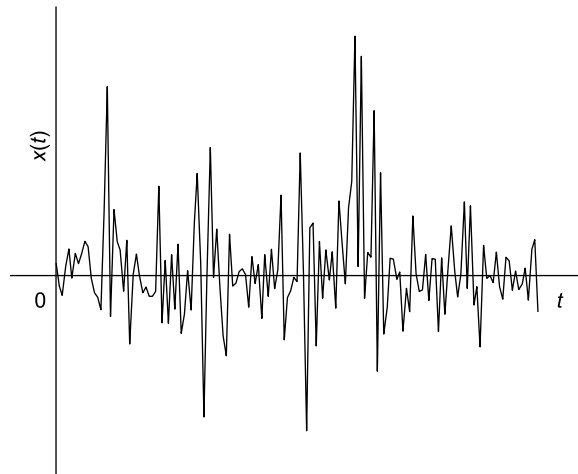
FIGURE 7.1   Sample function of a random process

7.3   PROBABILISTIC MODEL

For random signals, it is not possible to predict what the value of a signal is at a particular instant of time t_1, but these signals can very conveniently be described in probabilistic terms. There are various ways to characterize a probabilistic model. The most common is the probability density function (PDF), in which the probability of occurrence of a particular value from amongst the range of probable values is computed. Equivalent ways are the cumulative distribution function (CDF), the moments, the characteristic function, the moment-generating function, etc.

7.3.1   Probability Density Function (PDF) and Cumulative Distribution Function (CDF)

Random quantities can be described in a probabilistic sense. The probability of an event which is absolutely certain to occur may be taken as unity, while that of an event which is not at all likely to occur is taken

as zero, with the probability of any other event lying in between the two. We all know that in the tossing of a coin, the probability that the outcome will be a head is the same as the probability that the outcome will be a tail and this is 0.5. Mathematically this may be written as

$$\text{Prob } [H] = \text{Prob } [T] = 0.5 \tag{7.1}$$

This has been arrived at intuitively, based on the knowledge of the geometric symmetry of the coin and the resulting outcomes. We also know that the total probability, i.e. Prob $[H]$ + Prob $[T]$ = 1. This argument can be extended to the tossing of a symmetrical six-sided unloaded die. We can guess that the probability that the outcome of a toss would be 1, 2, 3, 4, 5 or 6 would all be the same, i.e. 1/6, again based on symmetry. Let the outcome of a particular toss be denoted by N. This outcome will be known only after the trial. Let n denote all possible previously known outcomes, i.e. 1, 2, 3, 4, 5 or 6. Now we can write

$$p\,[5] = \text{Prob } [N = 5] = 1/6 \tag{7.2}$$

and the total probability is

$$P\,[n] = \sum_{r=1}^{n} p(r) = 1 \tag{7.3}$$

The quantity $p(n)$ is known as probability density, while the quantity $P(n)$ is known as probability distribution or cumulative distribution.

Let us extend this argument to the case where the random variable takes on a large number of discrete values. Assuming that in this case there is no symmetry possible, a very large number of trials have to be conducted and the probability of the occurrence of an event has to be ascertained from the same. If a trial is repeated M times and if an event A occurs m times, the probability of occurrence of event A is defined as

$$p(A) = \lim_{M \to \infty} \left(\frac{m}{M} \right) \tag{7.4}$$

In general, a discrete random variable can take on values $x_1, x_2, x_3, \ldots$, etc. Then the probability that it takes on a value x_i is given by

$$\text{Prob } (x_i) = p_i \geq 0 \tag{7.5}$$

and total probability is

$$P\,[x] = \sum_{i} p_i = 1 \tag{7.6}$$

For a continuous random variable X, probability density $p(x)$ is defined as the probability that $X = x$ and is a non-negative function as before, i.e. $p(x) \geq 0$ for all x. Here x is any previously specified or known possible value of the function and X is the unknown result of a trial, which will be known only after the trial.

If a large number of samples of a continuous random variable are obtained to produce a histogram, or frequency diagram, depicting relative frequencies of occurrences of various values of the random signal, then this histogram will correspond to the PDF of the random variable, provided that a very large number of possible values are considered (i.e. they are sampled with very fine resolution Δx). A PDF is thus nothing but a smoothened out version of a histogram and is a convenient probabilistic quantity

when there are only a finite number of possible occurrences, since this will give rise to probabilities which are small fractional values of the total number of occurrences.

However, when there is an infinite number of occurrences possible, as in the case of a continuous variable $x(t)$ the probability of occurrence of one value out of the infinite possible values would tend to zero. Hence it would be convenient to use probability distribution or cumulative distribution instead of probability density in the case of continuous random variables, in contrast to the description of discrete random variables. For example, the probability that the height of an individual is exactly 165.249735 cm is practically zero, but the probability that the height is within the interval 165.1 and 165.3 cm is a measurable quantity. In such cases, it would be more realistic to find out the (total) probability $P(x)$ that the height of an individual would be less than or equal to a certain specified value, since this is a finite quantity, and can be obtained based on a sufficiently large number of trials. The probability that the value of x is less than or equal to some specified value X is the probability distribution or cumulative distribution and is

$$P(x) = \text{Prob } [X \le x] = \int_{-\infty}^{x} p(z)dz \tag{7.7}$$

For the discrete variable case the probability distribution function is given by

$$P(x) = \text{Prob } (X \le x) = \sum_{x_i \le x} p_i \tag{7.8}$$

The total area under any probability density curve corresponds to the total probability of the occurrences of all possible values of a variable. This implies that the total integral of $p(x)$ must be 1, i.e. the total area under the probability density graph is 1.

$$\int_{-\infty}^{+\infty} p(x)dx = 1 \tag{7.9}$$

The probability density $p(x)$ of a random variable $x(t)$ can also be computed from the time domain and is simply equal to the fraction of the time the variable takes on values between x and $(x + dx)$ to the total length of the time record T. The computation has to be done over a sufficiently long record of length T in the manner depicted in Fig. 7.2(a). Alternately $P(x_1)$ can be determined by finding out the total time that $x(t)$ is less than a value x_1, by summing all time intervals such as $\Delta t_1, \Delta t_2, \Delta t_3,...,$ etc. over a sufficiently long record of length T as shown in Fig. 7.2(b). This duration divided by the total time duration of the sample function represents the probability that $x(t)$ is less than x_1.
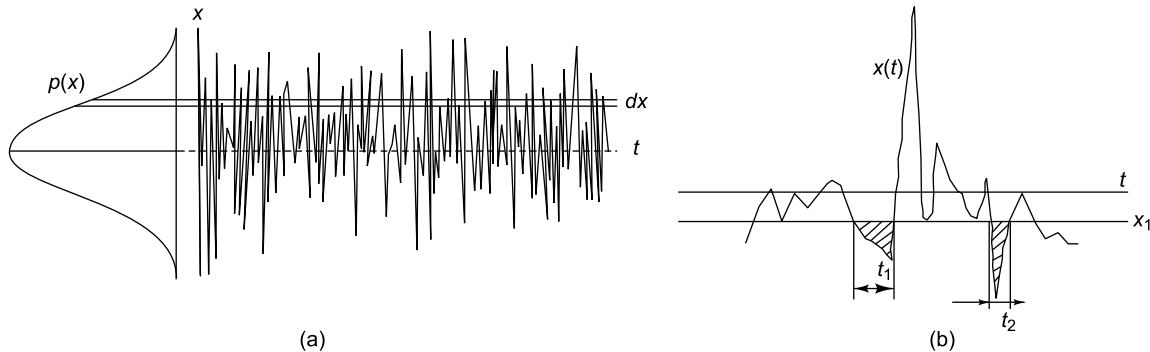
FIGURE 7.2   Computation of probability functions: (a) density function, (b) distribution function

If $p(x)$ is known, then the (total) probability $P(x)$ that the value of the variable $x(t)$ is within an interval $[x_1, x_2]$ is described by the integral of the function $p(x)$ over the integration interval of the variable.

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} p(x)dx \qquad (7.10)$$

This total probability that variable X will lie between two specific values x_1 and x_2 can be expressed mathematically as the difference of probabilities as shown on the right-hand side of the equation below.

$$\text{Prob}(x_1 \le X \le x_2) = \text{Prob}[X \le x_2] - \text{Prob}[X \le x_1]$$
$$= P(x_2) - P(x_1) \qquad (7.11)$$

Thus the probability that X lies in the very small interval $(x, x + dx)$ can be written as

$$dP(x) = \text{Prob}[X \le x + dx] - \text{Prob}[X \le x]$$
$$= P[x + dx] - P(x) \qquad (7.12)$$

For very small dx, this probability will be proportional to dx and may be denoted as $p(x)dx$. Therefore,

$$dP(x) = p(x)dx \qquad (7.13)$$

It is possible to say from intuition also that if a probability distribution has density $p(x)$, then in the infinitesimal interval $(x, x + dx)$, it has probability $p(x)dx$. In the limit $dP(x)$ and dx becoming infinitesimally small, probability density $p(x)$ can be obtained from probability distribution as shown below:

$$p(x) = \frac{dP(x)}{dx} \qquad (7.14)$$

where $p(x)$ and $P(x)$ are the probability density and distribution, respectively, as stated earlier. Therefore,

$$P(x) = \int_{-\infty}^{x} p(z)dz \qquad (7.15)$$

For practical continuous variables like the height or weight of a person, the lower limit of integration is set to 0.

Figure 7.3 shows typical probability density and distribution curves. The shapes of these curves give a qualitative indication of the nature of distribution. If a variable has values closely surrounding a mean value, the $p(x)$ curve would be a steep narrow curve, centred around the mean, while the $P(x)$ curve would show a steep rise in this area, increasing monotonically from 0 at $t = -\infty$ to 1 at $t = +\infty$.
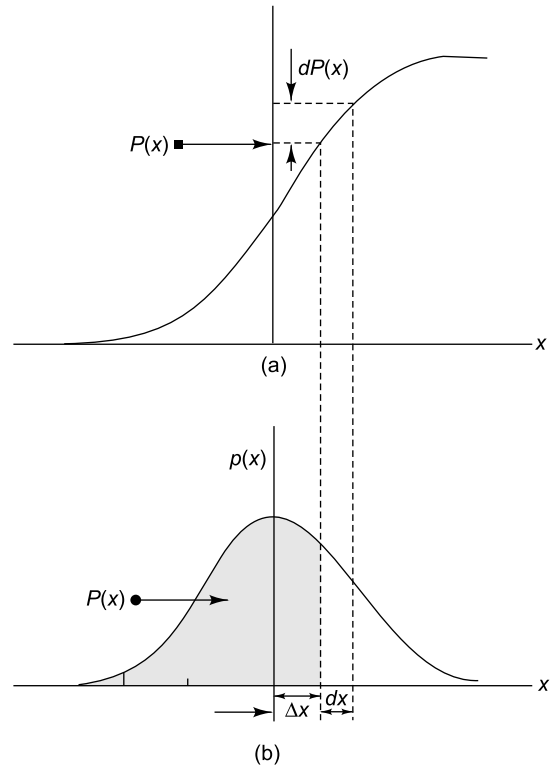
FIGURE 7.3   Probability functions: (a) probability distribution, (b) probability density

The word 'confidence' is often used when describing a statistical distribution. When we say that the declared life of a bulb is 180 days based on a 96% confidence level, it means that 96% of a very large number of bulbs tested has a life greater than 180 days. Or in other words, any sample taken would have at least 0.96 probability of exceeding the declared life of 180 days. Thus a confidence level is simply a measure of probability and can be obtained from the cumulative distribution curve.

Figures 7.4(a)–(c) show the time records, probability density and distribution curves, respectively for a few signals. The first row corresponds to a narrow-band random signal, the second row to a broad-band random signal and the third row to a sinusoidal signal.
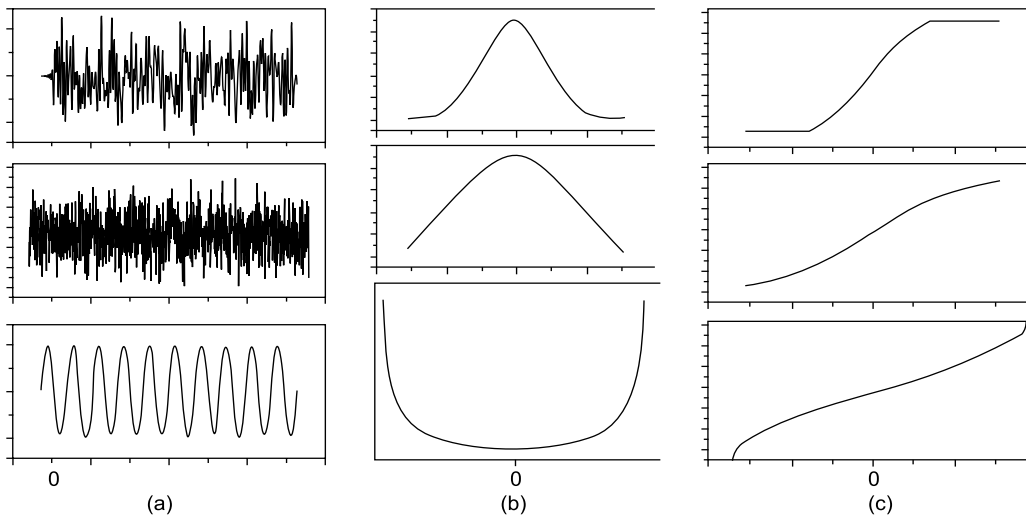
FIGURE 7.4   Probability for three typical signals: (a) $x(t)$ versus t, (b) $p(x)$ versus x, (c) $P(x)$ versus x

7.3.2   Probability of Joint Events

When considering probability of joint events, we define what are called independent events and dependent events.

7.3.2.1   *Independent events or variables*   Two discrete events, A and B are said to be independent or uncorrelated if occurrence of event A does not affect the occurrence of event B, and vice versa. The outcomes of two tosses of a die constitute a typical example of two independent events. The joint PDF, i.e. the probability of both events occurring is then

$$p(a, b) = p(a)p(b) \tag{7.16a}$$

When dealing with two continuous random variables, X and Y, the joint PDF is given by

$$p(x, y) = p(x)p(y) \tag{7.16b}$$

The joint CDF for this pair of random variables X and Y is given by

$$P(x, y) = \text{Prob } (X \leq x, Y \leq y)$$
$$= P(x)P(y) \tag{7.17}$$

where $P(x)$ and $P(y)$ are the marginal distributions of X and Y. The right-hand side of the above equation represents the probability that the random variable X takes on a value less than or equal to x and that Y takes on a value less than or equal to y. This can be extended to the case of N continuous random variables $X_1, X_2, X_3,..., X_N$. They are all independent of each other or uncorrelated if and only if

$$p(x_1, x_2, x_3,..., x_N) = p(x_1)p(x_2)p(x_3)...p(x_N) \tag{7.18}$$

If the joint PDF can thus be factored into a product of N functions of one variable each as shown above, the CDF can be defined as

$$P(x_1, x_2, x_3,..., x_N) = P(x_1)P(x_2)P(x_3)...P(x_N) \tag{7.19}$$

7.3.2.2 *Dependent events or variables*   Two variables are said to be dependent if they are related to each other, such as the load and deflection of a cantilever beam. Another example could be the humidity and rainfall at a particular location. Let us suppose that an experiment is conducted with two sets of possible outcomes. Suppose that out of N experiments performed, n produce outcome A; of these N experiments, let us suppose that m experiments also produce outcome B. The joint probability is the probability that the outcome A from one set occurs together with outcome B from the other set.

$$p(a, b) = \lim_{N \to \infty} \left(\frac{n}{N} \right)\left(\frac{m}{n} \right) p(a)p(b/a) \tag{7.20}$$

where $p(b/a)$ is the probability of B given that A has occurred.

For the case of two continuous random variables, the probability distribution and density functions are given by

$$P(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} p(\xi, \eta)\, d\xi d\eta \tag{7.21a}$$

$$p(x, y) = \frac{\partial^2 P(x, y)}{\partial x \partial y} \tag{7.21b}$$

For the case of N variables,

$$P(x_1, x_2, x_3,...) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2}\int_{-\infty}^{x_3} p(\xi, \eta, \varsigma,...)d\xi d\eta d\varsigma ... \tag{7.22a}$$

$$p(x_1, x_2, x_3,...) = \frac{\partial^N P(x_1, x_2, x_3,...)}{\partial x_1 \partial x_2 \partial x_3 ...} \tag{7.22b}$$

Here $p(x_1)$ is the PDF associated with the variable X_i alone and is called the marginal density function. It can be deduced from the joint PDF of the random variables $X_1, X_2, X_3,..., X_N$ by integrating with respect to all the other $N - 1$ variables as shown below

$$p(x_i) = \int \cdots \iint \cdots \int p(x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_N)\, dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_N \tag{7.23}$$

7.4 Some Common Distributions

Though there are a very large number of distributions that are in use, such as rectangular or uniform, Gaussian, binomial, Poisson, etc., we shall discuss a few distributions that come in handy while dealing with random variables in the context of random vibration.

7.4.1   Rectangular Distribution

A rectangular distribution is shown in Fig. 7.5(a). It is assumed that the probability density $p(x)$, of the variable, X, is constant between the limits a and b. The PDF for this distribution is given by

$$p(x) = \frac{1}{b-a} \tag{7.24}$$

7.4.2   Binomial Distribution

This distribution (Fig. 7.5(b)) may be used to determine the probability of failure of a component during its life. It is used to find the probability of occurrence of a single event from a very large number of trials when the event has only two possible results, say success and failure, which are mutually exclusive. Other assumptions made behind this distribution are that the results of repetitive trials are independent of each other and that the probabilities of occurrence are time invariant. Let us consider a series of trials in which the probability of occurrence of a particular event is a and that of its non-occurrence is $1 - a = b$. The probability that the event will not occur in any one trial is $1 - a$. The probability that it is not at all present in n trials is $(1 - a)^n$. Hence the probability that the event is not completely absent in n trials or in the other words that it will occur in n trials is

$$P = 1 - (1-a)^n \tag{7.25}$$

The above equation gives the probability of occurrence in a number of trials n, given the probability of occurrence in one trial.

Equation (7.25) can be expanded using the binomial theorem since $a < 1$,

i.e.
$$P = \left[1 - na + \frac{n(n-1)}{2!} p^2 ... \right] \tag{7.26}$$

When n is very large

$$P \backsim na \tag{7.27}$$

The mean value is na where n is the number of trials and a the probability of success. Let us consider a more general case of a sequence of n trials having k successes with probability a and $n - k$ failures with probability $b = 1 - a$. If we neglect the order in which these events occur, then the ways in which these events occur is given by

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{7.28}$$

Hence the probability of k successes in n trials is equal to

$$p(k \text{ successes, } n \text{ trials}) = \binom{n}{k} a^k b^{n-k} \tag{7.29}$$

This equation represents the binomial distribution with parameters a and n. A random variable X is said to have a binomial distribution if

$$p(X = k) = \binom{n}{k} a^k b^{n-k}, \quad k = 0, 1, ...n \tag{7.30}$$

The probability of success in all the other *n–k* trials is 0.

The sum of the probabilities is given by the binomial theorem as

$$\sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k} \;=\; (a+b)^n = 1 \tag{7.31}$$
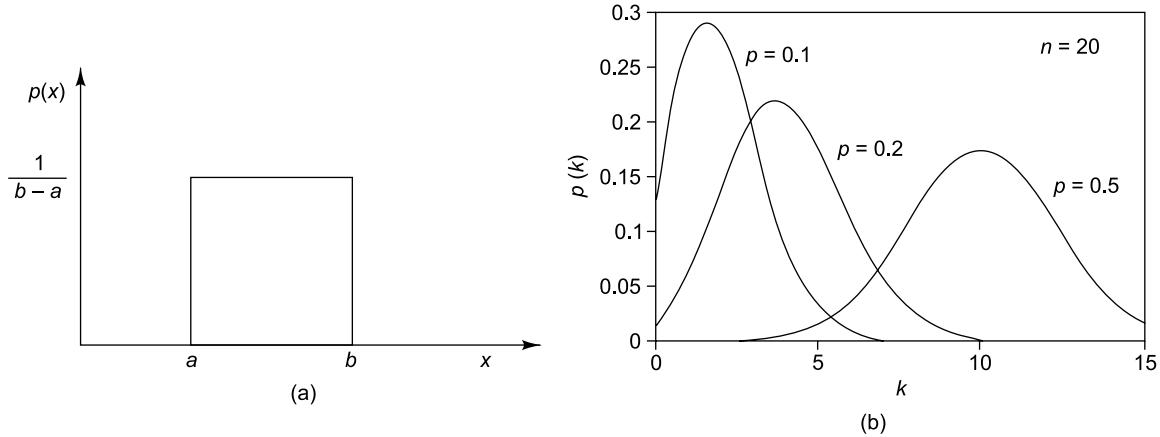
FIGURE 7.5   Probability density function: (a) rectangular distribution, (b) binomial distribution

7.4.3 Gaussian Distribution/Normal Distribution

The normal distribution, also called the Gaussian distribution, is a very important form of continuous probability distribution widely used in science and engineering, for the reason that many real life signals can be idealized as Gaussian. It is also called the bell curve because its probability density is a bell shaped curve. Many measurements of physical phenomena, such as radio noise, jet engine pressure fluctuation, atmospheric turbulence, thermal noise, etc. can be approximated by the normal distribution. This distribution is the most extensively used family of distributions in statistics to find significance levels in many hypothesis tests and confidence intervals. In modelling applications, such as linear and non-linear regression, the error term is often assumed to follow a normal distribution. The advantage of being able to model a random process as a Gaussian distribution is that the output of a linear time-invariant system to Gaussian random input is also Gaussian. Under certain conditions, such as for independent and identically-distributed variables with finite variance, the sum of a large number of random variables is approximately normally distributed and this is the essence of the central limit theorem. The practical importance of this is that the normal CDF can be used as an approximation to some other CDFs. For example, a binomial distribution with parameters *n* and *a* is approximately normal for large *n* and *a* not too close to 1 or 0. Thus for most distributions, if the number of events or trials is very large, then the Gaussian distribution function may be used to describe them.

The probably density function for a Gaussian distribution shown in Fig. 7.6 is given by

$$p(x) \;=\; \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty \tag{7.32}$$

where μ and σ totally characterize the Gaussian signal, μ being the mean and σ being the standard deviation. To indicate that a real-valued random variable X is normally distributed with mean μ and

variance $\sigma^2 \geq 0$, we write $X \sim N(\mu, \sigma^2)$. The mean, or the expected value corresponds to the centroid of the PDF; it is also the point at which the PDF is maximum. The variance σ^2 is a measure of the dispersion of the random variable around the mean. Thus this distribution is highly mathematically tractable, being characterized by only two parameters, the mean and the variance (square of standard deviation) and if these are known, the PDF is defined. Figures 7.6(a) and (b) show the probability density and distribution functions for various values of μ and σ.
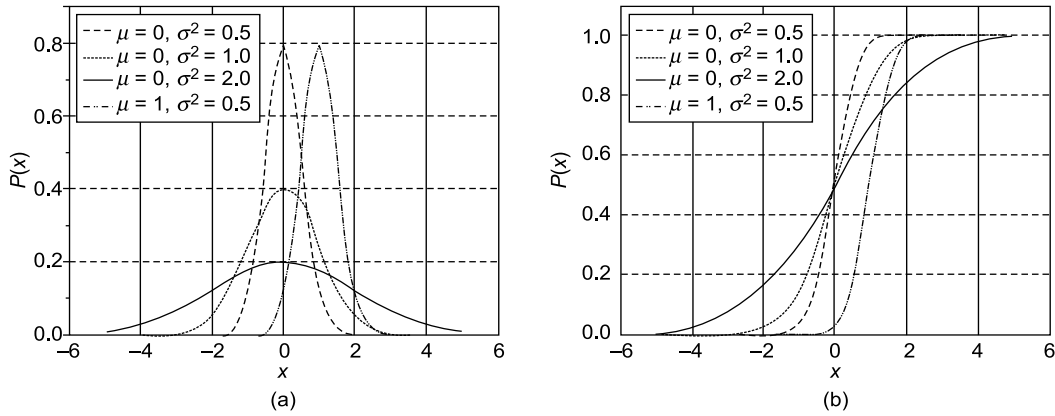
FIGURE 7.6   Normal distribution for different μ and σ: (a) probability density function, (b) probability distribution function

If the mean value μ is zero, then

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x^2/2\sigma^2)} \tag{7.33}$$

The formula for the CDF of the normal distribution does not exist in the form of a simple closed form expression. It may be expressed in terms of the density function as follows:

$$P(x) = \text{Prob}\,(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(y-\mu)^2/2\sigma^2}\,dy = \Phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \tag{7.34}$$

where

$$\text{erf}\,(z) = \frac{2}{\sqrt{\pi}} \int_{0}^{z} e^{-y^2}\,dy \tag{7.35}$$

The Gaussian distribution can be normalized so that $\mu = 0$ and $\sigma = 1$ and this is called the standard normal distribution. This is done by a change of variables such that $z = [(x - \mu)/\sigma]$. It may be noted that the case corresponding to $\mu = 0$ and $\sigma = 1$ in Fig. 7.6 represents the standard normal distribution. The equation for the PDF of the standard normal distribution is given by

$$p(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right), -\infty < z < \infty \tag{7.36}$$

i.e. $Z \sim N(0,1)$. Conversely, if Z has a standard normal distribution, $Z \sim N(0,1)$, then X is a normal random variable with mean μ and variance σ^2 as shown below

$$X = \sigma Z + \mu \qquad (7.37)$$

Since the general form of the Gaussian probability function can be expressed in terms of the standard normal distribution, all subsequent formulae in this section are given for the standard form of the function. To evaluate the probability in Eq. (7.34), the error function erf (z), has to be evaluated. The standard normal distribution is tabulated (usually in the form of values of the CDF) and is available readily in most text books, and the other normal distributions are simple transformations, as described above, of the standard one. Therefore, one can use tabulated values of the CDF of the standard normal distribution to find values of the CDF of a general normal distribution. The cumulative distributions computed numerically for $\mu = 0$ and $\sigma = 1$ give the cumulative probability that $z = z_1$, which is nothing but the area under the normal curve $p(z)$ up to z_1, computed as shown in the following equation and as depicted in Fig. 7.7. Table 7.1 gives $P(z)$ corresponding to a few often used values for the area under the bell curve. These values are the areas computed under the normal curve from 0 to z, i.e. erf(z) and are useful in determining confidence intervals of the specified levels based on standard normal distribution. Figure 7.8 depicts a plot of the standard normal PDF.

$$P(z = z_1) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{-(z^2/2)} \, dz \qquad (7.38)$$
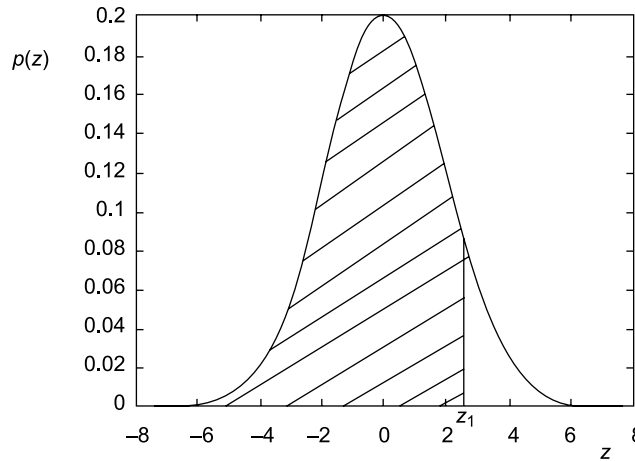
FIGURE 7.7   Computation of CDF from standard normal distribution

From this density curve, it is observed that values of z less than one standard deviation from the mean, account for about 68% of the total set of values, or in other words the probability that z is within one standard deviation of the mean is 0.683. Similarly values of z within two standard deviations from the mean account for about 95%. Values of z that are less than three standard deviations from the mean account for about 99.7% of the entire population. This can be represented mathematically as

$$\left.\begin{array}{l} \text{Prob}\{|X-\mu| \leq \sigma\} = 2\,\text{erf}\,(1) = 0.6827 \\ \text{Prob}\{|X-\mu| \leq 2\sigma\} = 2\,\text{erf}\,(2) = 0.9545 \\ \text{Prob}\{|X-\mu| \leq 3\sigma\} = 2\,\text{erf}\,(3) = 0.9973 \end{array}\right\} \qquad (7.39)$$

TABLE 7.1   Gaussian table for computation of CDF

z	Area	z	Area	z	Area	z	Area
0.05	0.01994	0.80	0.28814	1.55	0.43943	2.30	0.48928
0.10	0.03983	0.85	0.30234	1.60	0.44520	2.35	0.49061
0.15	0.05962	0.90	0.31594	1.65	0.45053	2.40	0.49180
0.20	0.07926	0.95	0.32894	1.70	0.45543	2.45	0.49286
0.25	0.09871	1.00	0.34134	1.75	0.45994	2.50	0.49379
0.30	0.11791	1.05	0.35314	1.80	0.46407	2.55	0.49461
0.35	0.13683	1.10	0.36433	1.85	0.46784	2.60	0.49534
0.40	0.15542	1.15	0.37493	1.90	0.47128	2.65	0.49597
0.45	0.17365	1.20	0.38493	1.95	0.47441	2.70	0.49653
0.50	0.19146	1.25	0.39435	2.00	0.47726	2.75	0.49702
0.55	0.20884	1.30	0.40320	2.05	0.47982	2.80	0.49744
0.60	0.22575	1.35	0.41149	2.10	0.48214	2.85	0.49781
0.65	0.24215	1.40	0.41924	2.15	0.48422	2.90	0.49813
0.70	0.25804	1.45	0.42647	2.20	0.48610	2.95	0.49841
0.75	0.27337	1.50	0.43319	2.25	0.48778	3.00	0.49865
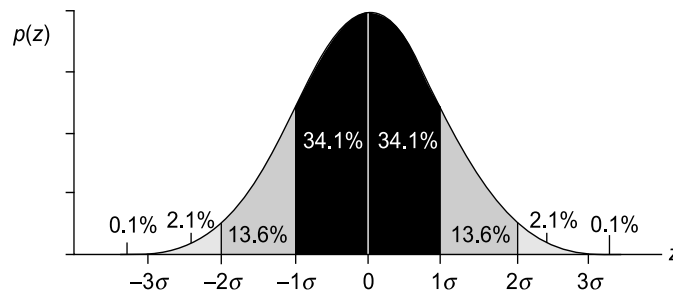
FIGURE 7.8   Standard normal probability density function

7.5   STATISTICAL DESCRIPTORS FOR RANDOM SIGNALS

In any statistical method, a large amount of data is required to establish reliability. For example, to characterise road undulations over a certain stretch, a large number of records may have to be collected. The values typically used to characterize random signals in a statistical sense are mean, mean square value, skewness, kurtosis and other higher moments and central moments.

7.5.1   Expected Value, Time Averaging and Moments

Expectation or expected value is defined as the expected result in any trial, and is assumed to be the mean value of a very large number of trials. In the tossing of a symmetrical die, it can be assumed that all numbers from 1 to 6 will occur with equal frequency, provided that the number of throws is very large. Therefore, the expected value in this case is the mean of all the outcomes, i.e. 1, 2, 3, 4, 5 and 6.

$$E[N] = \left(\frac{1+2+3+4+5+6}{6}\right) = 3\frac{1}{2} \tag{7.40}$$

Here E represents the expectation operator. It can also be written as

$$E[N] = \sum_1^6 n\, p(n) \tag{7.41}$$

Thus for the case of discrete random variables, the expected value, which is the average value, can be generalized as

$$E[X] = \bar{x} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} x_i \tag{7.42a}$$

The concept of time averaging over a sufficiently long period of time T (ideally tending to infinity) may also be used and is often done so in the treatment of continuous-time random variables. The expected value of $x(t)$ or expectation or mean value denoted by $E[X] = \langle x(t)\rangle = \overline{x(t)}$ is defined as the average or mean value of the variable sampled over time T tending to infinity.

Here $\langle\ \rangle$ denotes the expectation operator.

$$\text{Mean} = E[X] = \langle x(t)\rangle = \overline{x(t)} = \mu_x = \lim_{T\to\infty}\frac{1}{T}\int_0^T x(t)\,dt \tag{7.42b}$$

The mean square value, designated by $\overline{x^2(t)}$ or expectation of $x^2(t)$, i.e. $E[X^2]$, may also be obtained by integrating $x^2(t)$ over a sufficiently long time T and taking its average value. This quantity is an indication of the energy present in the signal. The mean square value of a vibration signal is quite often used in machinery condition monitoring to give an indication of the severity of a fault. Mean square value can be applied to a single variable such as $x^2(t)$ or to a product of variables such as $x(t).y(t)$.

$$E\left[x^2(t)\right] = \overline{x^2(t)} = \lim_{T\to\infty}\frac{1}{T}\int_0^T x^2(t)\,dt \tag{7.43}$$

Covariance is a quantitative measure of the degree to which two random variables $x(t)$ and $y(t)$ are correlated and is defined as

$$\sigma_{xy}^2 = \langle x(t).y(t)\rangle \tag{7.44}$$

If $x(t)$ and $y(t)$ are completely uncorrelated, σ_{xy}^2 will be zero; if they are not independent of each other, σ_{xy}^2 will be other than zero.

When one is interested in characterising a signal in terms of its mean value and its fluctuation about the mean, a quantity called the variance σ^2, which is the mean squared value about the mean is used. This is a quantity of great importance when dealing with random signals and is obtained by doing a time averaging as shown below. σ is referred to as the standard deviation. σ^2 may be obtained from time integration and is given by the equation below.

$$\text{Variance} = \sigma^2 = \left\langle \left[x(t) - \langle x(t) \rangle \right]^2 \right\rangle = \lim_{T \to \infty} \frac{1}{T} \int_0^T (x - \bar{x})^2 dt \tag{7.45}$$

This can be simplified to obtain an equation relating variance and mean as shown below.

$$\begin{aligned} \sigma^2 &= \left\langle [x^2(t) - 2x(t)\langle x(t) \rangle + \langle x(t) \rangle^2 \right\rangle \\ &= \left\langle x^2(t) \right\rangle - 2\langle x(t) \rangle^2 + \langle x(t) \rangle^2 \\ &= \left\langle x^2(t) \right\rangle - \langle x(t) \rangle^2 \end{aligned} \tag{7.46}$$

i.e.
$$\sigma^2 = \overline{x^2} - (\bar{x})^2 = \overline{x^2} - \mu_x^2$$

From this equation it is easily seen that variance σ^2 equals mean squared value minus square of the mean. For zero mean signals, the variance is the same as the mean square value and the square root of the variance, i.e. the standard deviation σ is the same as the root mean square value. This fact may come in handy when trying to validate experimentally obtained results. In practice 'AC coupling mode' gives rise to a zero mean signal.

The mean and mean square values defined in terms of the time averages are related to the PDF as explained below. If a random variable X is given, and its distribution admits a PDF $p(x)$, then the first moment or expectation or expected value of X (if it exists) can be determined by computing the first moment of the area under the probability density curve about the y axis, that is the vertical axis passing through $x = 0$ (Fig. 7.9). First moment of the probability density curve about the $p(x)$ axis gives

$$E(X) = \bar{x} = \int_{-\infty}^{+\infty} x\, p(x) dx \tag{7.47}$$

The mean value $\bar{x}$ or the average value coincides with the centroid of the area under the probability density curve $p(x)$. One can recall from theory relating to properties of surfaces that the following equations give the location of the centroid denoted by $(\bar{x}, \bar{y})$. First moments of an area A about the y and x axes give

$$\bar{x} = \frac{1}{A} \int_0^A x\, dA \tag{7.48a}$$

$$\bar{y} = \frac{1}{A} \int_0^A y\, dA \tag{7.48b}$$

The area A for a probability density histogram is 1.
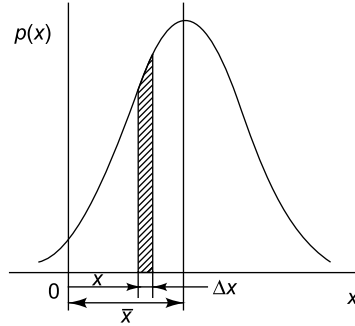
FIGURE 7.9   Computation of moments of area of PDH about $p(x)$ axis

Similarly the mean square value defined in terms of a time average is related to the PDF and may be determined from the second moment of the area under the probability density curve about the vertical axis $x = 0$.

$$\text{Mean square value} = E(X^2) = \left\langle x^2(t) \right\rangle = \overline{x^2} = \int_{-\infty}^{+\infty} x^2\, p(x)\, dx \tag{7.49}$$

This is analogous to the computation of the second moment of an area about $x = 0$ which gives the radius of gyration. The radii of gyration got from computation of the second moments of an area about the y and x axes are given in the following equations.

$$k_y^2 = \frac{1}{A}\int_0^A x^2\, dA \tag{7.50a}$$

$$k_x^2 = \frac{1}{A}\int_0^A y^2\, dA \tag{7.50b}$$

The variance σ^2 which is the mean squared value about the mean, is also known as the second central moment. It may be defined as the second moment of the area under the probability density curve about the vertical axis through $x = \mu_x$ or second moment about the mean.

$$E[(x - \mu_x)^2] = \sigma^2 = \overline{(x - \overline{x})^2} = \int_{-\infty}^{+\infty} (x - \overline{x})^2\, p(x)\, dx = \int_{-\infty}^{+\infty} \left(x - \mu_x\right)^2 p(x)\, dx \tag{7.51}$$

In a similar fashion, higher order moments and central moments as well as cross-moments may be computed. In particular, the nth moment $E(X^n)$ of the probability distribution of a random variable X is given by

$$E(X^n) = \int_{-\infty}^{+\infty} x^n\, p(x)\, dx \tag{7.52}$$

Similarly, the higher order central moments may also be computed as

$$E\left[\left(X - \mu_x\right)^n\right] = \int_{-\infty}^{\infty} \left(x - \mu_x\right)^n p(x)\, dx \tag{7.53}$$

Amongst the higher order moments, the quantities of interest are skewness and kurtosis which correspond to order 3 and 4, respectively. These quantities find use in machinery diagnostics.

7.5.2   Auto- and Cross-Correlations

Correlation is a measure of the similarity between two quantities. As applied to vibration or acoustic signals, correlation is a time domain analysis procedure useful for detecting hidden periodicities buried in a measured signal and for the measurement of the propagation time through a structure. Correlation gives information regarding a structure's spectral characteristics.

7.5.2.1   Autocorrelation   The autocorrelation function is a quantity which in spite of being time-parametered, gives a lot of information about the spectral content of a signal. It is a measure of the general dependence of the value of a signal at some instant of time to values at other instants of time. Autocorrelation is the expected value of the product $x(t)x(t + \tau)$. It is a quantity that can be computed from a single record of sufficiently long duration T by multiplying the ordinates at time t and time $t + \tau$ (Fig. 7.10) and determining the average over all t.
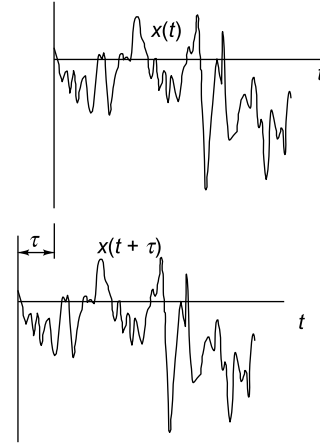
FIGURE 7.10   Computation of autocorrelation

$$
\left.
\begin{aligned}
R_{xx}(\tau) &= \lim_{T \to \infty} \frac{1}{T} \int_0^T \left[x(t)x(t + \tau) \right] dt \\
&= \left\langle x(t)x(t + \tau) \right\rangle
\end{aligned}
\right\}
\tag{7.54}
$$

If averaging is done over a sufficiently long record, $R_{xx}(\tau)$ is independent of t and is a function of τ alone. When τ equals zero, the autocorrelation reduces to the mean squared value as shown below.

$$
R_{xx}(0) = \sigma^2 = \overline{x^2} \tag{7.55}
$$

$R_{xx}(0)$ is the maximum value of the autocorrelation function. As all the products are squares, $R_{xx}(0)$ must be positive. When τ is very small, it can be expected that $x(t)$ and $x(t + \tau)$ are not much different from each other, and hence the number of positive products is larger than that of the negative products, resulting in a positive value of $R_{xx}(\tau)$, which will nevertheless be less than $R_{xx}(0)$. For increasing values of τ, the factors of each product become less and less related to each other and $R_{xx}(\tau)$ goes on oscillating between positive and negative values, before finally settling to zero. If fluctuations in $x(t)$ are very rapid, then $x(t)$ and $x(t + \tau)$ may have very little relationship. If on the other hand, the fluctuations are slow, then $x(t)$ and $x(t + \tau)$ are closely connected. Thus the nature of the autocorrelation curve is dependent on the spectral content of the signal. When τ equals infinity, in the products $x(t) \, x(t + \tau)$ encompassing all τ, it can be expected that there are almost an equal number of positive and negative values in the products, cancelling each other out and therefore $R_{xx}(\infty)$ tends to zero. The autocorrelation function of highly random signals, such as wide-band noise, therefore shows a sharp spike at τ equal to zero that drops off rapidly with $\pm \tau$. Therefore, except at τ equal to zero, wide-band random records have little or no correlation. If we assume that we are dealing with a member function $x(t)$ of a stationary ergodic random process, $R_{xx}(\tau)$ will also correspond to the autocorrelation function of any other member function

(Section 7.7).

Figure 7.11 shows the autocorrelation of some common functions. The first, second and third rows depict narrow-band random, broad-band random and sine wave signals, respectively. The autocorrelation function of a sine wave is a cosine wave. For a narrow-band signal, the autocorrelation is similar to that of a sine wave with the envelope decaying slowly. The autocorrelation function of a wide-band random signal is a sinc function. Figure 7.12 shows the setup for the computation of autocorrelation.
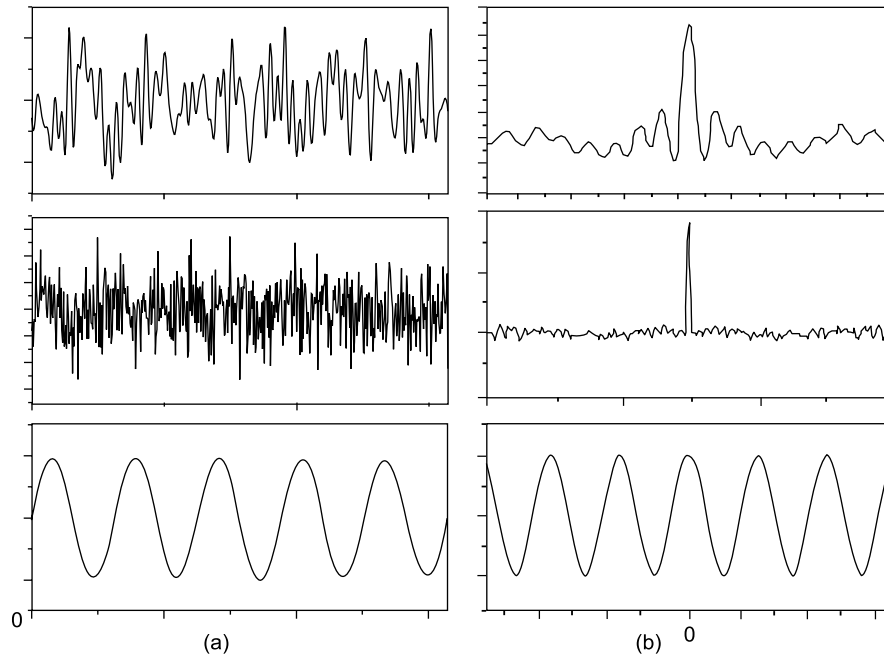

FIGURE 7.11 Autocorrelation functions of different signals: (a) $x(t)$ versus t, (b) $R_{xx}(\tau)$ versus τ
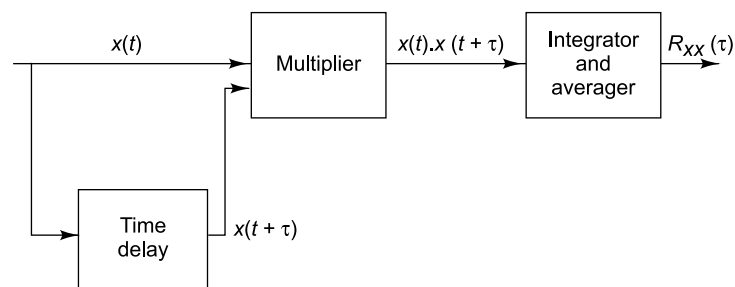
FIGURE 7.12 Computation of autocorrelation

Let us obtain the autocorrelation function of a non-zero mean signal.

Let
$$x(t) = \xi(t) + \langle x(t) \rangle \tag{7.56}$$

Here ξ is the fluctuating part of $x(t)$ and $\langle x(t) \rangle = \mu_x$ = mean value of $x(t)$.

$$R_{xx}(\tau) = \langle x(t) x(t+\tau) \rangle$$

$$= \left\langle \left[\left\langle x(t) \right\rangle + \xi(t) \right] \left[\left\langle x(t+\tau) \right\rangle + \xi(t+\tau) \right] \right\rangle \qquad (7.57)$$

$$= \left\langle x(t) \right\rangle^2 + R_{\xi\xi}(\tau)$$

This reduces to the above form for a stationary ergodic random process since $\left\langle x(t+\tau) \right\rangle = \left\langle x(t) \right\rangle$

and $\left\langle \xi(t+\tau) \right\rangle = \left\langle \xi(t) \right\rangle = 0.$

7.5.2.2 *Cross-correlation* The correlation between two records called the cross-correlation, is a measure of the similarity between the two records. A typical example could be $x(t)$ being the deflection at a point on a beam due to a load $F(t)$ at some other point on the same beam. Cross-correlation between two records $x(t)$ and $y(t)$ can be obtained by multiplying the ordinates of the two records at time t and time $t + \tau$ for all time t and determining the average value or expectation $\left\langle x(t)y(t+\tau) \right\rangle$ by dividing the sum of the products by the number of products. It is clear that cross-correlation will be highest if the two records are identical or similar. For dissimilar records, some products will be positive and some negative, so that this sum will be smaller.

Cross-correlation between two quantities $x(t)$ and $y(t)$ may be calculated as

$$R_{xy}(\tau) = \left\langle x(t)y(t+\tau) \right\rangle \qquad (7.58)$$

Correlation is related to spectral density through a Fourier relationship as shown later.

7.6 CLASSIFICATION OF RANDOM DATA

Random processes may be classified as stationary and non-stationary as shown in Fig. 7.13. Ergodic and non-ergodic random processes form subsets of stationary random processes.
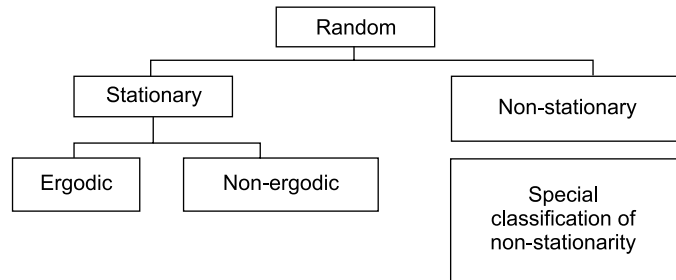
FIGURE 7.13 Classification of random data

7.6.1 Stationarity

Let $x(t)$ be a random variable, say the random displacement of a vehicle traversing a rough road. This signal could be any other random quantity as well, such as random road undulation or jet noise. This quantity varies with time, but if the statistical characteristics of this time record are invariant with respect to time, then we may say that it is stationary. Consider a random process $\{x(t)\}$ consisting of N member

functions or sample records $x_1(t)$ to $x_N(t)$ as shown in Fig. 7.14. The total collection of sample functions or sample records is called an ensemble and constitutes the random process $\{x(t)\}$.

$x_1(t)$

$x_2(t)$

$x_3(t)$
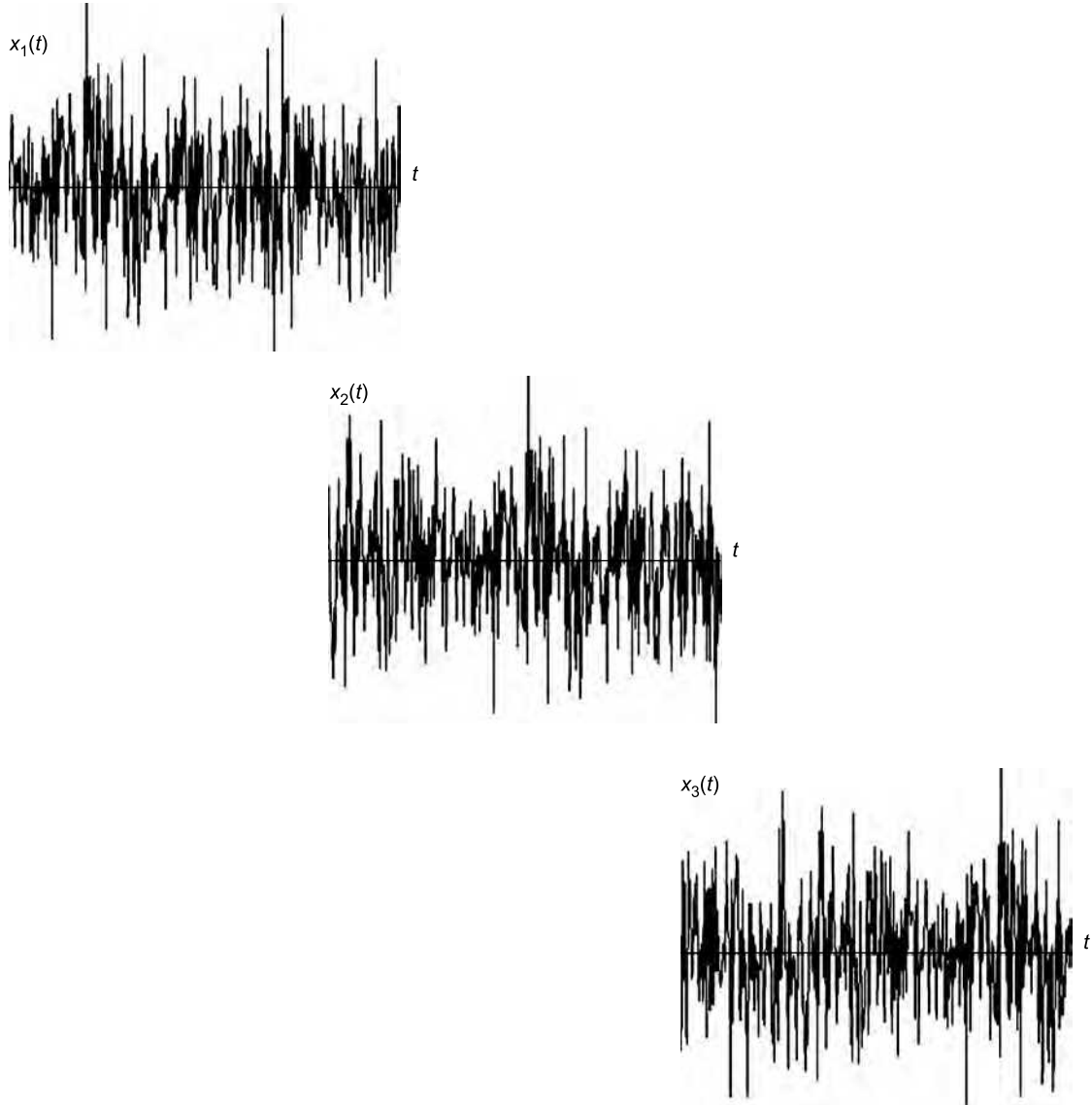
FIGURE 7.14   Ensemble of records of a random phenomenon

The mean value of the signal across the ensemble at different instants of time t_1, t_2, t_3,..., t_n may be represented as $\mu_x(t_1)$, $\mu_x(t_2)$, $\mu_x(t_3)$,..., $\mu_x(t_n)$. Here $\mu_x(t_1)$ is computed as

$$\mu_x(t_1) = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} x_N(t_1) \tag{7.59}$$

Consider the scenario where $\mu_x(t_1) = \mu_x(t_2) = \mu_x(t_3) = \dots \mu_x(t_n) = \mu_x$ (7.60)

If these ensemble mean values are equal, i.e. if ensemble averages at time $t = t_1$, $t = t_2$, $t = t_3, \dots t = t_n$ for all the N sample functions are the same, then the random process is said to be invariant in the mean, or in other words, the ensemble average is independent of time t. It is possible to compute the ensemble autocorrelations at different instants of time as before. Ensemble autocorrelation at instant of time t_1 is

$$R_{xx}(t_1,\ \tau)\ =\ \lim_{N\to\infty}\sum_{k=1}^{N}x_k\left(t_1\right)x_k\left(t_1+\tau\right) \tag{7.61}$$

Consider the case

$$R_{xx}(t_1,\ \tau) = R_{xx}(t_2,\ \tau) = R_{xx}(t_3,\ \tau) = \dots = R_{xx}(t_n,\ \tau) = R_{xx}(\tau) \tag{7.62}$$

The random process is said to be invariant with respect to the autocorrelation function, or the ensemble autocorrelation is independent of time t. If conditions represented by Eqs (7.60) and (7.62) are satisfied, then $\{x(t)\}$ is said to be a weakly stationary random process. If all possible higher order moments and joint moments are also time invariant, then the random process is said to have strong stationarity.

7.6.2 Ergodicity

This is a subset of stationarity and requires that the time averages of all sample functions of duration T are equal. If $\mu_x(k)$ and $R_{xx}(\tau, k)$ which are the temporal averages over each sample function, do not differ when computed over different sample functions in the ensemble, then the random process is said to be ergodic. The sample averages as computed for the kth sample are shown below:

$$\mu_x(k)\ =\ \lim_{T\to\infty}\frac{1}{T}\int_0^T x_k\left(t\right)dt \tag{7.63}$$

$$R_{xx}(t,\ k)\ =\ \lim_{T\to\infty}\frac{1}{T}\int_0^T x_k\left(t\right)x_k\left(t+\tau\right)dt \tag{7.64}$$

Consider the situation where the ensemble averages are equal to the sample averages, i.e.

$$\mu_x(1) = \mu_x(2) = \mu_x(3) = \dots\mu_x(N) = \mu_x \tag{7.65}$$
$$= \mu_x(t_1) = \mu_x(t_2) = \mu_x(t_3) = \dots\mu_x(t_n)$$

and
$$R_{xx}(\tau,\ 1) = R_{xx}(\tau,\ 2) = R_{xx}(\tau,\ 3) = \dots R_{xx}(\tau,\ N) = R_{xx}(\tau) \tag{7.66}$$
$$= R_{xx}(t_1,\ \tau) = R_{xx}(t_2,\ \tau) = R_{xx}(t_3,\ \tau) = \dots R_{xx}(t_n,\ \tau)$$

This is a case of the random process having both stationarity and ergodicity. A stationary, ergodic random signal is very convenient to handle, since one member function is sufficient to describe the entire random process and quantities like mean, variance, probability density, cumulative distribution, autocorrelation, power spectral density (PSD), etc. of one sample function are sufficient. Thus for a stationary ergodic random process, the statistical properties may be got from a single time function of sufficiently long duration. Though the assumption of stationarity and ergodicity may be a hypothetical one, it simplifies to a large extent the task of dealing with random variables. It is also to be mentioned

that verification of weak stationarity can be taken as a justification of an assumption of strong stationarity for most practical problems.

7.7 FREQUENCY DOMAIN REPRESENTATION OF RANDOM SIGNALS

For periodic signals, we have seen in Section 1.2.3 that we can have a Fourier series representation. A non-periodic signal, a transient signal, for example, can be expressed in the form of a Fourier series if it is considered to be periodic with infinite period. In this case the discrete spectrum becomes a continuous curve as the Fourier series becomes the Fourier integral. Thus the equations relating the time and frequency domain representations for a transient signal are given by the Fourier transform pair

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-i\omega t} dt$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) e^{i\omega t} d\omega$$

(7.67)

$X(\omega)$ is called the Fourier transform of $x(t)$.

The two equations show near symmetry and they can be made more symmetric by writing the equations in terms of frequency f in Hz instead of circular frequency $\omega = 2\pi f$ in rad/s. Thus,

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft} dt$$

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{i2\pi ft} df$$

(7.68)

Application of Parseval's theorem to these equations gives some very useful results. Let the complex conjugate of $X(f)$ be denoted as $X^*(f)$.

$$X^*(f) = \int_{-\infty}^{\infty} x(t) e^{i2\pi ft} dt$$

(7.69)

We know that

$$X(f).X^*(f) = |X(f)|^2$$

(7.70)

Thus

$$\int_{-\infty}^{\infty} x^2(t) dt = \int_{-\infty}^{\infty} x(t) \left[\int_{-\infty}^{\infty} X(f) e^{i2\pi ft} df \right] dt$$

$$= \int_{-\infty}^{\infty} X(f) \left[\int_{-\infty}^{\infty} x(t) e^{i2\pi ft} dt \right] df$$

(7.71)

by changing the order of the integration. Therefore,

$$\int_{-\infty}^{\infty} x^2(t)\,dt = \int_{-\infty}^{\infty} X(f).X^*(f)\,df$$

$$= \int_{-\infty}^{\infty} |X(f)|^2\,df \tag{7.72}$$

Since negative frequencies have no physical meaning, this equation can be rewritten as

$$\int_{-\infty}^{\infty} x^2(t)\,dt = 2\int_{0}^{\infty} |X(f)|^2\,df \tag{7.73}$$

This result comes in handy when trying to define the PSD, which is the frequency domain representation of a random signal. A random signal cannot be expressed in the form of a Fourier transform since we have convergence issues in the computation. In order to have stationary properties, a random signal must be assumed to have an infinite length, in which case, the Fourier transform does not converge to a steady value. It is possible to define a new quantity called the PSD which has no convergence issues and which is applicable to all sample functions of a stationary, ergodic random process.

Let us consider a sample function $x(t)$ of a stationary random process. Since this exists from $t = -\infty$ to $t = +\infty$, its Fourier transform does not exist. We can, however, define a signal $x_T(t)$ identical to $x(t)$ over the interval $-(T/2)<t<(T/2)$ and zero elsewhere as shown in Fig. 7.15.
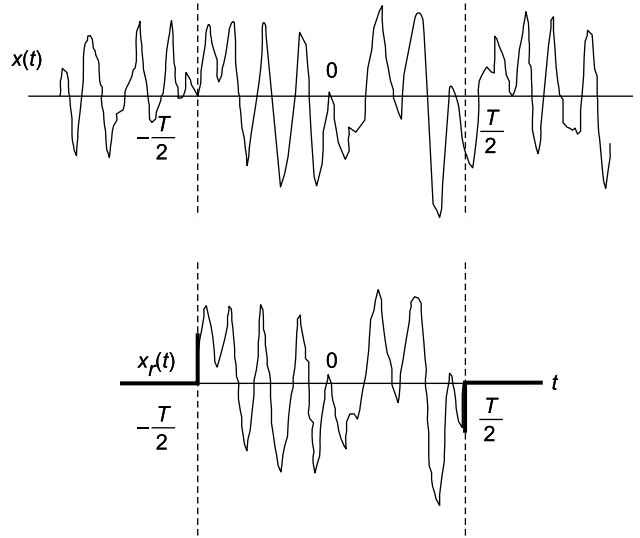
FIGURE 7.15   Description of $x_T(t)$

The mean square value of $x_T(t)$ is given by

$$\langle x^2(t) \rangle = \frac{1}{T}\int_{-(T/2)}^{T/2} x_T^2(t)\,dt = \frac{1}{T}\int_{-\infty}^{\infty} x_T^2(t)\,dt = \frac{2}{T}\int_{0}^{\infty} |X_T(f)|^2\,df \tag{7.74}$$

Letting $T \rightarrow \infty$, we can get an expression for the mean square value of $x(t)$.

$$\left\langle x^2(t) \right\rangle = \int_0^\infty \lim_{T \to \infty} \left[\frac{2}{T} \left| X_T(f) \right|^2 \right] df = \int_0^\infty S_{xx}(f) df$$

$$(7.75)$$

where
$$S_{xx}(f) = \lim_{T \to \infty} \left[\frac{2}{T} \left| X_T(f) \right|^2 \right]$$

Here $S_{xx}(f)$ is called the PSD or simply (auto) spectral density of $x(t)$ and represents the distribution of the harmonic content of the signal. It is represented in terms of energy/Hz (Fig. 7.16).

Thus for instance, voltage PSD would be represented in terms of V^2/Hz. The units for PSDs of current, vibration acceleration and pressure would be A^2/Hz, g^2/Hz and Pa^2/Hz, respectively. $S_{xx}(f)\Delta f$ is the mean square value of the signal passed through a narrow band-pass filter of bandwidth Δf. The area under the PSD curve gives the root mean square value of the signal.

FIGURE 7.16   Typical PSD plot

We have already seen that the latter quantity may be obtained from a time domain description $x(t)$ or from a probabilistic description $p(x)$. A third method is now seen to exist in terms of PSD $S_{xx}(f)$,

i.e.
$$\sigma^2 = \left\langle x^2(t) \right\rangle = \int_{-\infty}^\infty x^2 p(x) dx = \int_0^\infty S_{xx}(f) df \qquad (7.76)$$

One salient aspect of the PSD is that since it is obtained from the modulus $|X(f)|$, it does not contain any phase information and is therefore applicable to a whole range of functions $x(t)$ which might have been generated by a physical system. Hence it can be expected that different member functions $x(t)$ of a stationary ergodic random process $\{x(t)\}$ will have a common spectral density $S_{xx}(f)$. Therefore, the spectral content of all member functions $x(t)$ of a stationary ergodic random process may be defined by a single PSD $S_{xx}(f)$. However a knowledge of $S_{xx}(f)$ alone is not sufficient to define a random process.

A random signal which has a constant spectral content over a range of frequencies is called white noise, through an imperfect analogy with white light. Though the concept of a white noise signal extending to infinite frequency is not physically meaningful, in practice we do have what are called band-limited white noise signals extending over a limited range of frequencies and with a definite cut-off frequency. Thus band-limited white noise is quite often used in vibration testing to determine the first few modes of a system with the knowledge that the response of the system at higher modes is very small. $S_{xx}(f)$ can be measured using a bandpass filter. Figure 7.17 shows typical PSDs of some common signals: they are from the top, narrow band noise, broadband noise and a harmonic signal.

In general, knowing the PSD, the mean square value can be determined, but nothing can be said about the PDF or cumulative distribution. If, however, a random process is known to be Gaussian, with PDF in the form given in Eq. (7.32), then the PDF can be written knowing the value of mean μ and standard deviation σ. This is the reason why scientists and engineers grab the opportunity to model a random process as Gaussian, with the process displaying the slightest hint of Gaussian properties.
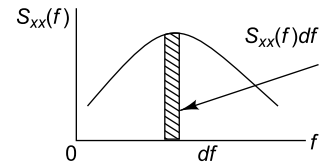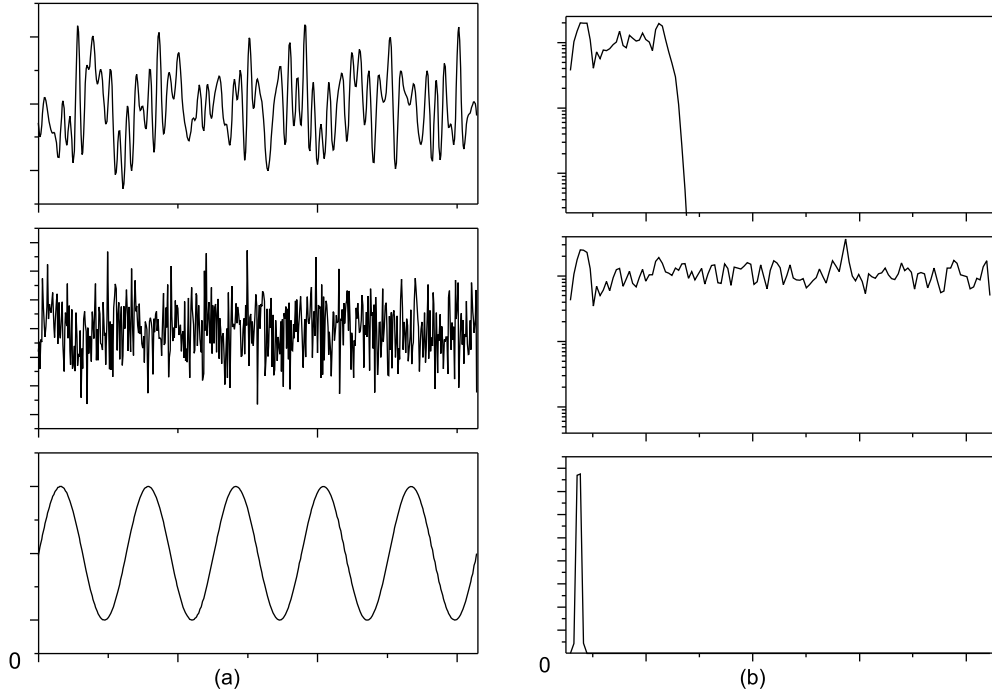
FIGURE 7.17 PSDs of common signals: (a) $x(t)$ versus t, (b) $S_{xx}(f)$ versus f

7.7.1 Relationship between Autocorrelation $R_{xx}(\tau)$ and Power Spectral Density $S_{xx}(f)$

We have intuitively deduced in Section 7.5.2.1 that autocorrelation is related to spectral content in a qualitative manner. The mathematical expression relating the two may be obtained as described below. We shall start with the function $x_T(t)$ coinciding with $x(t)$ over the range $-(T/2) < t < (T/2)$ and zero elsewhere as before (Fig. 7.15).

Autocorrelation $R_{xx}(\tau)$ can be defined as shown below, as the autocorrelation of function $x_T(t)$, mean value being taken over interval T.

$$R_{xx}(\tau) = \lim_{T\to\infty} R_{x_T x_T}(\tau) = \lim_{T\to\infty} \langle x_T(t) x_T(t+\tau) \rangle \tag{7.77}$$

Taking Fourier transform of $R_{x_T x_T}(\tau)$ and making the simplification $s = t + \tau$, we get

$$\int_{-\infty}^{\infty} R_{x_T x_T}(\tau) e^{-i2\pi f\tau} d\tau = \int_{-\infty}^{\infty} \frac{1}{T}\left[\int_{-\infty}^{\infty} x_T(t) x_T(t+\tau) dt\right] e^{-i2\pi f\tau} d\tau$$

$$= \frac{1}{T}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} x_T(t) x_T(t+\tau) e^{-i2\pi f\tau} dt\right] d\tau$$

$$= \frac{1}{T} \int\limits_{-\infty}^{\infty} \left[\int\limits_{-\infty}^{\infty} x_T(t) x_T(t+\tau) e^{i2\pi ft} e^{-i2\pi f(t+\tau)} dt \right] d\tau$$

$$= \frac{1}{T} \int\limits_{-\infty}^{\infty} \left[\int\limits_{-\infty}^{\infty} x_T(t) x_T(s) e^{i2\pi ft} e^{-i2\pi fs} dt \right] ds$$

$$= \frac{1}{T} \int\limits_{-\infty}^{\infty} x_T(t) e^{i2\pi ft} dt \int\limits_{-\infty}^{\infty} x_T(s) e^{-i2\pi fs} ds \qquad (7.78)$$

$$= \frac{1}{T} X_T^*(f) . X_T(f)$$

$$= \frac{1}{T} |X_T(f)|^2$$

Letting $T \to \infty$ so that $R_{x_T x_T}(\tau)$ becomes $R_{xx}(\tau)$ and using Eqs. (7.75) and (7.78), we have

$$R_{xx}(\tau) = \int\limits_{-\infty}^{\infty} \frac{1}{2} S_{xx}(f) e^{i2\pi f \tau} \, df \qquad (7.79)$$

i.e. the autocorrelation function and the PSD are related through a Fourier transform relationship. This equation is called the Wiener–Khinchin equation. Spectral density may be easier to measure, while autocorrelation may be more convenient for time domain analysis.

7.8   RESPONSE OF AN SDOF SYSTEM

In the following discussions we shall try to derive the response of an SDOF system to transient and random excitations. These concepts will come in handy in understanding the modal analysis procedures that will be discussed in Chapter 10. However, before we do so, it is imperative to understand how the SDOF system responds to a harmonically varying load. Then we may graduate to transient and random loads. The assumptions made in all the following analyses are that the system is linear, time-invariant and stable.

7.8.1   Response to Harmonic Excitation

This has been discussed at length in Section 2.2.6.2 for an SDOF system and in Section 2.4.5 for an MDOF system. Let us refer to Fig. 2.9 which represents a damped SDOF system. Let $\alpha(\omega)$ be its complex frequency response function (FRF), or receptance, to be more specific. We shall derive an expression for this letting the harmonic loading to be represented by $f(t) = F_0 e^{i\omega t}$. The dynamic equation of the system is

$$m\ddot{x} + c\dot{x} + kx = F_0 e^{i\omega t} \qquad (7.80)$$

Since the response of an LTI system to harmonic loading is harmonic at the same frequency as the excitation, Eq. (7.80) may be simplified as

$$\left(-m\omega^2 + i\omega c + k\right)X_0 = F_0 \tag{7.81}$$

or

$$X_0 = \frac{F_0}{\left(k - m\omega^2\right) + i\omega c} \tag{7.82}$$

$$= \alpha(\omega)F_0$$

The above equation is a statement of the fact that a harmonic function is an eigenfunction of an LTI system and $\alpha(\omega)$ is the eigenvalue or complex FRF of the system.

$$\alpha(\omega) = \frac{1}{\left(k - m\omega^2\right) + i\omega c} \tag{7.83}$$

The response $x(t)$ of the system is described by the equations

$$x(t) = \frac{F_0}{\left(k - m\omega^2\right) + i\omega c} e^{i\omega t}$$

$$= \frac{F_0\left(k - m\omega^2 - i\omega c\right)}{\left(k - m\omega^2\right)^2 + \omega^2 c^2} e^{i\omega t}$$

$$= \frac{F_0 e^{-i\phi}}{\sqrt{\left(k - m\omega^2\right)^2 + \omega^2 c^2}} e^{i\omega t}, \quad \text{where } \tan\phi = \frac{\omega c}{k - m\omega^2} \tag{7.84}$$

Working in terms of actual frequency f instead of circular frequency ω, we get

$$\alpha(f) = \frac{1}{k - m 4\pi^2 f^2 + i 2\pi fc} \tag{7.85}$$

This quantity $\alpha(f)$ is the response of the SDOF system to a complex force of unit modulus and proportional to $e^{i\omega t}$ and is called the receptance of the system. The receptance provides a complete description of the response of the LTI system to harmonic excitation.

The above result may be obtained in an alternate fashion as described below. We know that the response to a force $f(t) = F_0 e^{i2\pi ft}$ is given by the convolution integral

$$x(t) = \int_0^\infty h(\tau) F_0 e^{i2\pi f(t-\tau)} d\tau$$

$$= F_0 e^{i2\pi ft} \int_0^\infty h(\tau) e^{-i2\pi f\tau} d\tau \tag{7.86}$$

where $h(t)$ is called the Impulse Response Function as discussed in Chapter 6 and is the response to a unit impulse. However due to the eigenfunction property of $e^{i2\pi ft}$, we know from Eq. (7.82) that

$$x(t) = \alpha(f) e^{i2\pi ft} \tag{7.87}$$

Therefore,
$$\alpha(f) = \int_{0}^{\infty} h(\tau)e^{-i2\pi f \tau} d\tau \tag{7.88}$$

i.e. $\alpha(f)$ is the Fourier transform of $h(t)$ if $h(t)$ is assumed to be 0 for $\tau < 0$, i.e. for a causal system.

For any other arbitrary signal, which can be represented as the superposition of a number of harmonic signals, i.e. for an excitation which can be expressed in the form of a Fourier series, the total response can be obtained by superimposing the individual responses to all harmonic excitations, taking phase information into account.

7.8.2 Response to Transient Excitation

We often come across situations where a system is subjected to impulsive disturbances (large amplitude, short duration forces of the order of a few seconds). Typical examples of such excitations are excitations during forging, punching and blasting operations. Duhamel's integral or convolution integral is used for the solution of transient vibration problems in linear systems and is based on the principle of superposition of responses to a sequence of impulses. Any impulsive loading may be expressed as

$$F(t) = I\delta(t) \tag{7.89}$$

where I is the magnitude of the impulse and $\delta(t)$ is the Dirac δ function described in Eqs (6.13) and (6.14). For an arbitary loading of the form shown in Fig. 7.18, it is possible to compute the response of an LTI system as discussed below.

The loading itself may be expressed as a superposition of magnitude-scaled and time-shifted impulses in the form of a convolution integral as shown below.

$$F(t) = \int_{0}^{\infty} \delta(\tau) F(t - \tau) d\tau \tag{7.90}$$
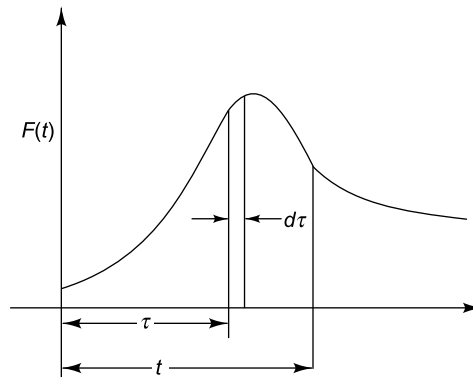
FIGURE 7.18   Arbitrary load

The total response to this loading is obtained using the convolution integral

$$x(t) = \int_{0}^{\infty} h(\tau) F(t - \tau) d\tau \tag{7.91}$$

i.e. the response is expressed as the superposition of amplitude scaled, time shifted IRFs. It is also to be remembered that $h(t)$ is the inverse Fourier transform of $H(\omega)$; $h(t)$ and $H(\omega)$ represent the system parameters in the time and frequency domains, respectively.

7.8.3   Response to Random Loading

Any random excitation that we may input will be idealized as a stationary, ergodic random process, making the response also stationary and ergodic. It is to be borne in mind that the random response which is calculated will have no more details than that present in the mathematical description of the excitation function. Let us consider a randomly varying force $F(t)$. Let autocorrelation function of the force be denoted by $R_{FF}(\tau)$ and its autospectral density by $S_{FF}(f)$. Let the displacement output of the SDOF system to this force be denoted by $x(t)$ with autocorrelation $R_{xx}(\tau)$ and spectral density $S_{xx}(f)$. The assumption behind this analysis is that the system is a causal, stable, LTI system and that the input and output random signals are stationary, ergodic random processes with zero mean. It is also assumed that the input is applied at a single point and we are concerned with the output displacement at a single point. The aim of the exercise is to determine $R_{xx}(\tau)$ in terms of $R_{FF}(\tau)$ and $S_{xx}(f)$ in terms of $S_{FF}(f)$. The response $x(t)$ of the system may be written in terms of the convolution integral

$$x(t) = \int_0^\infty h(\tau_1) F(t - \tau_1) d\tau_1$$

$$x(t + \tau) = \int_0^\infty h(\tau_2) F(t + \tau - \tau_2) d\tau_2$$

$$(7.92)$$

The autocorrelation of the response is given by

$$R_{xx}(\tau) = \langle x(t) x(t + \tau) \rangle$$

$$= \lim_{T \to \infty} \frac{1}{T} \int_0^T \left[\int_0^\infty h(\tau_1) F(t - \tau_1) d\tau_1 \right] \left[\int_0^\infty h(\tau_2) F(t - \tau + \tau_2) d\tau_2 \right] dt$$

$$= \int_0^\infty h(\tau_1) \int_0^\infty h(\tau_2) \langle F(t - \tau_1) F(t + \tau - \tau_2) \rangle d\tau_2 \, d\tau_1$$

$$= \int_0^\infty h(\tau_1) \int_0^\infty h(\tau_2) \langle F(t) F(t + \tau_1 - \tau_2 + \tau) \rangle d\tau_2 \, d\tau_1$$

$$= \int_0^\infty h(\tau_1) \int_0^\infty h(\tau_2) R_{FF}(\tau_1 - \tau_2 + \tau) d\tau_2 \, d\tau_1 \qquad (7.93)$$

The last of the above equations follows from the fact that a change in the origin of t makes no difference to the averaged quantity if the process is stationary. The spectral density may be obtained by taking the Fourier transform of $R_{xx}(\tau)$.

$$S_{xx}(f) = 2 \int_{-\infty}^\infty R_{xx}(\tau) e^{-i2\pi f \tau} d\tau$$

$$= 2\int_{-\infty}^{\infty}\left[\int_0^{\infty}h(\tau_1)\int_0^{\infty}h(\tau_2)R_{FF}(\tau_1-\tau_2+\tau)d\tau_2\,d\tau_1\right]e^{-i2\pi f\tau}d\tau$$

$$= 2\int_0^{\infty}h(\tau_1)\int_0^{\infty}h(\tau_2)\left[\int_{-\infty}^{\infty}R_{FF}(\tau_1-\tau_2+\tau)e^{-i2\pi f\tau}d\tau\right]d\tau_2\,d\tau_1 \tag{7.94}$$

$$= 2\int_0^{\infty}h(\tau_1)e^{i2\pi f\tau_1}\int_0^{\infty}h(\tau_2)e^{-i2\pi f\tau_2}\int_{-\infty}^{\infty}R_{FF}(\tau_1-\tau_2+\tau)e^{-i2\pi f(\tau_1-\tau_2+\tau)}d(\tau_1-\tau_2+\tau)d\tau_2\,d\tau_1$$

$$= 2\int_0^{\infty}h(\tau_1)e^{i2\pi f\tau_1}d\tau_1\int_0^{\infty}h(\tau_2)e^{-i2\pi f\tau_2}d\tau_2\int_{-\infty}^{\infty}R_{FF}(\tau_1-\tau_2+\tau)e^{-i2\pi f(\tau_1-\tau_2+\tau)}d(\tau_1-\tau_2+\tau)$$

Therefore,

$$S_{xx}(f) = 2\alpha^*(f)\alpha(f)\frac{1}{2}S_{FF}(f) \tag{7.95}$$

$$= |\alpha(f)|^2 S_{FF}(f)$$

This is an elegant and simple equation relating the PSD of the response to the PSD of excitation through the transfer function. The equation states that the PSD of the response at any frequency is the product of the PSD of the exciting force at that frequency and square of the modulus of the receptance at that frequency. It is worthwhile noting that phase information is lacking in this equation and hence obtaining $|\alpha(f)|$ is not difficult. Once the PSD of the response is obtained, the mean square value of the response can also be got.

7.8.3.1 *Response involving cross-correlations*   Figure 7.19 shows a system with two random excitations $P(t)$ and $Q(t)$ and response $x(t)$. Let $\alpha_{xP}(f)$ and $\alpha_{xQ}(f)$ (denoted simply by α_{xP} and α_{xQ}) be the receptances for harmonic excitation at locations P and Q and let the corresponding IRFs be $h_{xP}(t)$ and $h_{xQ}(t)$, respectively.

Let autocorrelations of $x(t)$, $P(t)$ and $Q(t)$ be $R_{xx}(\tau)$, $R_{PP}(\tau)$ and $R_{QQ}(\tau)$, respectively. Then the autocorrelation of the response may be expressed as

FIGURE 7.19   Response of a system to random inputs

$$R_{xx}(\tau) = \left\langle x(t)x(t+\tau)\right\rangle$$

$$= \left\langle \int_0^{\infty}\left[h_{xP}(\tau_1)P(t-\tau_1)+h_{xQ}(\tau_1)Q(t-\tau_1)\right]d\tau_1 \right.$$

$$\left. \times \int_0^{\infty}\left[h_{xP}(\tau_2)P(t-\tau_2+\tau)+h_{xQ}(\tau_2)Q(t-\tau_2+\tau)\right]d\tau_2 \right\rangle$$

$$= \left\langle \int_0^{\infty}h_{xP}(\tau_1)P(t-\tau_1)d\tau_1\int_0^{\infty}h_{xP}(\tau_2)P(t-\tau_2+\tau)d\tau_2 \right\rangle$$

$$+ \left\langle \int_0^\infty h_{xP}(\tau_1) P(t - \tau_1) d\tau_1 \int_0^\infty h_{xQ}(\tau_2) Q(t - \tau_2 + \tau) d\tau_2 \right\rangle$$

$$+ \left\langle \int_0^\infty h_{xQ}(\tau_1) Q(t - \tau_1) d\tau_1 \int_0^\infty h_{xP}(\tau_2) P(t - \tau_2 + \tau) d\tau_2 \right\rangle$$

$$+ \left\langle \int_0^\infty h_{xQ}(\tau_1) Q(t - \tau_1) d\tau_1 \int_0^\infty h_{xQ}(\tau_2) Q(t - \tau_2 + \tau) d\tau_2 \right\rangle$$

$$= \int_0^\infty h_{xP}(\tau_1) \left[\int_0^\infty h_{xP}(\tau_2) \langle P(t) P(t + \tau_1 - \tau_2 + \tau) \rangle d\tau_2 \right] d\tau_1$$

$$+ \int_0^\infty h_{xP}(\tau_1) \left[\int_0^\infty h_{xQ}(\tau_2) \langle P(t) Q(t + \tau_1 - \tau_2 + \tau) \rangle d\tau_2 \right] d\tau_1$$

$$+ \int_0^\infty h_{xQ}(\tau_1) \left[\int_0^\infty h_{xP}(\tau_2) \langle Q(t) P(t + \tau_1 - \tau_2 + \tau) \rangle d\tau_2 \right] d\tau_1$$

$$+ \int_0^\infty h_{xQ}(\tau_1) \left[\int_0^\infty h_{xQ}(\tau_2) \langle Q(t) Q(t + \tau_1 - \tau_2 + \tau) \rangle d\tau_2 \right] d\tau_1 \qquad (7.96)$$

The first and the last terms in the expansion within <> are autocorrelation functions and may be written as $R_{PP}(\tau_1 - \tau_2 + \tau)$ and $R_{QQ}(\tau_1 - \tau_2 + \tau)$. The corresponding quantities in the second and third terms are cross-correlation functions $R_{PQ}(\tau_1 - \tau_2 + \tau)$ and $R_{QP}(\tau_1 - \tau_2 + \tau)$. Hence Eq. (7.96) may be simplified as

$$R_{xx}(\tau) = \int_0^\infty h_{xP}(\tau_1) \left[\int_0^\infty h_{xP}(\tau_2) R_{PP}(\tau_1 - \tau_2 + \tau) d\tau_2 \right] d\tau_1 \text{ m}$$

$$+ \int_0^\infty h_{xP}(\tau_1) \left[\int_0^\infty h_{xQ}(\tau_2) R_{PQ}(\tau_1 - \tau_2 + \tau) d\tau_2 \right] d\tau_1$$

$$+ \int_0^\infty h_{xQ}(\tau_1) \left[\int_0^\infty h_{xP}(\tau_2) R_{QP}(\tau_1 - \tau_2 + \tau) d\tau_2 \right] d\tau_1$$

$$+ \int_0^\infty h_{xQ}(\tau_1) \left[\int_0^\infty h_{xQ}(\tau_2) R_{QQ}(\tau_1 - \tau_2 + \tau) d\tau_2 \right] d\tau_1 \qquad (7.97)$$

To determine the PSD of the response, we use the relationship

$$S_{xx}(f) = 2 \int_{-\infty}^\infty R_{xx}(\tau) e^{-i2\pi f \tau} d\tau$$

$$= 2 \Im \{ R_{xx}(\tau) \} \qquad (7.98)$$

$$= \alpha_{xP}^{*} \alpha_{xP} S_{PP}(f) + \alpha_{xP}^{*} \alpha_{xQ} S_{PQ}(f) + \alpha_{xQ}^{*} \alpha_{xP} S_{QP}(f) + \alpha_{xQ}^{*} \alpha_{xQ} S_{QQ}(f)$$

Here the cross-PSD terms are of the form

$$S_{PQ}(f) = 2 \int_{-\infty}^{\infty} R_{PQ}(\tau) e^{-i2\pi f\tau} d\tau \tag{7.99}$$

Thus it is seen that the response PSD is a function of auto and cross-PSDs of input.

Case (i) If $P(t)$ and $Q(t)$ are independent, then $S_{PQ}(f) = S_{QP}(f) = 0$ and therefore $S_{xx}(f)$ reduces to the form

$$S_{xx}(f) = |\alpha_{xP}|^{2} S_{PP}(f) + |\alpha_{xQ}|^{2} S_{QQ}(f) \tag{7.100}$$

Case (ii) If $P(t)$ and $Q(t)$ are directly correlated so that $Q(t) = kP(t)$ where k is a constant, then the auto and cross-correlations can be written as

$$R_{PQ}(\tau) = \langle P(t) Q(t+\tau) \rangle$$
$$= \langle P(t) kP(t+\tau) \rangle = kR_{PP}(\tau)$$
$$R_{QP}(\tau) = \langle kP(t).P(t+\tau) \rangle = kR_{PP}(\tau)$$
$$R_{QQ}(\tau) = \langle kP(t).kP(t+\tau) \rangle = k^{2} R_{PP}(\tau) \tag{7.101}$$

The PSD of the response is therefore

$$S_{xx}(f) = \alpha_{xP}^{*} \alpha_{xP} S_{PP}(f) + \alpha_{xP}^{*} kS_{PP}(f) + \alpha_{xQ}^{*} \alpha_{xP} kS_{PP}(f) + \alpha_{xQ}^{*} \alpha_{xQ} k^{2} S_{PP}(f)$$

$$= \left(\alpha_{xP}^{*} + k\alpha_{xQ}^{*} \right) \left(\alpha_{xP} + k\alpha_{xQ} \right) S_{PP}(f) \tag{7.102}$$

$$= \left| \alpha_{xP} + k\alpha_{xQ} \right|^{2} S_{PP}(f)$$

Case (iii) If the two forces are directly correlated, in which case $S_{PP}(f) = S_{QQ}(f) = S(f)$, the response PSD is

$$S_{xx}(f) = \left| \alpha_{xP} + \alpha_{xQ} \right|^{2} S(f)$$

$$= \left[|\alpha_{xP}|^{2} + |\alpha_{xQ}|^{2} + 2|\alpha_{xP}||\alpha_{xQ}| \cos\phi \right] S(f) \tag{7.103}$$

where ϕ is the phase difference between the two receptances at a particular frequency f. When $\phi = \pm \dfrac{\pi}{2}$, $\cos\phi = 0$ and this amounts to uncorrelated loadings.

Case (iv) If $Q(t)$ reproduces $P(t)$ after a lag τ_0 so that $Q(t) = P(t + \tau_0)$, the cross-correlation is

$$R_{PQ}(\tau) = \langle P(t) Q(t+\tau) \rangle$$
$$= \langle P(t) P(t+\tau_0+\tau) \rangle = R_{PP}(\tau_0+\tau) \tag{7.104}$$

The corresponding cross-spectral densities are given by the expressions

$$S_{PQ}(f) = 2\int_{-\infty}^{\infty} R_{PP}(\tau_0 + \tau)e^{-i2\pi f\tau}d\tau$$

$$= 2e^{i2\pi f\tau_0}\int_{-\infty}^{\infty} R_{PP}(\tau_0 + \tau)e^{-i2\pi f(\tau_0+\tau)}d(\tau_0 + \tau) \tag{7.105}$$

$$= e^{i2\pi f\tau_0}S_{PP}(f)$$

Obviously

$$S_{QQ}(f) = S_{PP}(f) \tag{7.106}$$

The PSD of the response is given by

$$S_{xx}(f) = \left(\alpha_{xP}{}^*\alpha_{xP} + e^{i2\pi f\tau_0}\alpha_{xP}{}^*\alpha_{xQ} + e^{-i2\pi f\tau_0}\alpha_{xQ}{}^*\alpha_{xP} + \alpha_{xQ}{}^*\alpha_{xQ}\right)S_{PP}(f) \tag{7.107}$$

Case (v) We often have to find the response to more than two loadings. Let us consider the case when n random loads act simultaneously, as shown in Fig. 7.20.
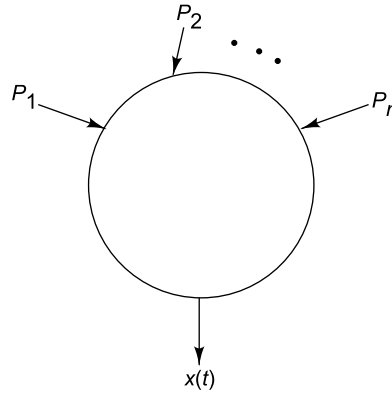
FIGURE 7.20 Multiple random loadings

In general for n forces P_1, P_2,..., P_n acting on a system to produce responses $x_1(t)$, $x_2(t)$,..., $x_n(t)$ at a particular location, the total response $z(t)$ at that location is given by

$$z(t) = x_1(t) + x_2(t) + \ldots + x_n(t) \tag{7.108}$$

The autocorrelation of the response takes on the form shown below

$$R_{zz}(\tau) = \sum_{r=1}^{n}\sum_{s=1}^{n} R_{x_r x_s}(\tau) \tag{7.109}$$

The response PSD is of the form

$$S_{xx}(f) = \sum_{r=1}^{n}\sum_{s=1}^{n} \alpha_{xP_r}{}^*\alpha_{xP_s}S_{P_r P_s}(f) \tag{7.110}$$

It is seen that the autocorrelation of the response has terms containing the autocorrelations of the input forces, as well as all possible combinations of cross-correlations between the inputs. This is true also of the autospectral density of the output.

FURTHER READINGS

1. Bendat, J.S. and Piersol, A.G., *Random Data: Analysis and Measurement Procedures*, 3rd ed., Wiley-Interscience, New York, 2000.
2. Bendat, J.S., *Engineering Applications of Correlation and Spectral Analysis*, Wiley-Interscience, New York, 1993.
3. Bertsekas, D.P., *Introduction to Probability*, Athena Scientific, Belmont, Massachusetts, 2002.
4. Billingsley, P., *Probability and Measure*, John Wiley and Sons, New York, Toronto, London, 1979.
5. Chung, K.L., *A Course in Probability Theory Revised*, Academic Press, Boston, 2000.
6. Crandall, S.H. and Mark, W.D., *Random Vibration in Mechanical Systems*, Academic Press, New York, 1963.
7. DeGroot, M.H. and Schervish, M.J., *Probability and Statistics*, Addison Wesley, Boston, 2001.
8. Elishakoff, I., *Probabilistic Theory of Structures*, Dover Publications, New York, 1999.
9. Feller, W., *An Introduction to Probability Theory and Its Applications*, Vol. 2, Wiley, New York, 1991.
10. Gallager, R.G., *Discrete Stochastic Processes*, Kluwer Academic, Boston, MA, 1996.
11. Garcia, A.L., *Probability, Statistics and Random Processes for Electrical Engineering*, Addison-Wesley, Reading, MA, 1994.
12. Hsu, H., *Schaum's Outline of Probability, Random Variables and Random Processes*, 1st ed., McGraw-Hill, New York, 1997.
13. Ibrahim, R.A., *Parametric Random Vibration*, Dover Publications, New York, 2008.
14. Karlin, S. and Taylor, H.M., *A First Course in Stochastic Processes*, Academic Press, New York, 1975.
15. Karlin, S. and Taylor, H.M., *A Second Course in Stochastic Processes*, Academic Press, New York, 1981.
16. Kolmogorov, A.N., *Foundations of the Theory of Probability*, Chelsea Publishing Company, New York, 1956.
17. Lin, Y. K., *Probabilistic Theory of Structural Dynamics*, Krieger Pub. Co., New York, 1976.
18. Lipschutz, S. and Lipson, M., *Schaum's Outline of Theory and Problems of Probability*, McGraw-Hill, New York, 2000.
19. Lutes, L.D. and Shahram Sarkani, S., *Random Vibrations: Analysis of Structural and Mechanical Systems*, Elsevier Butterworth-Heinemann, Burlington MA, 2004.
20. Miller, S.L. and Childers, D.G., *Probability and Random Processes: With Applications to Signal Processing and Communications*, Elsevier Science, Amsterdam, 2004.
21. Newland, D.E., *An Introduction to Random Vibrations, Spectral & Wavelet Analysis*, Dover Publications, New York, 2005.
22. Nigam, N.C., *Introduction to Random Vibrations*, MIT Press, Cambridge, USA, 1983.

23. Nigam, N.C. and Narayanan, S., *Applications of Random Vibrations*, Narosa Publishing House, New Delhi, India, 1994.

24. Papoulis, A. and Pillai, S.U., *Probability, Random Variables and Stochastic Processes*, McGraw-Hill Science/Engineering/Math, New York, 2002.

25. Preumont, A., *Random Vibration and Spectral Analysis*, Kluwer, Boston, 1994.

26. Roberts. J.B. and Spanos, P.D., *Random Vibration and Statistical Linearization*, Dover Publications, New York, 2003.

27. Robson. J.D., *An Introduction to Random Vibration*, Edinburgh University Press, Edinburgh, 1964.

28. Ross, S.M., *First Course in Probability*, Prentice Hall, Upper Saddle River, New Jersey, 2005.

29. Ross, S.M., *Stochastic Processes*, Wiley, New York, 1996.

30. Solnes, J., *Theory of Stochastic Processes and Random Vibrations*, John Wiley & Sons Inc., New York, 1997.

31. Soong, T.T. and Grigoriu, M., *Random Vibration of Mechanical and Structural Systems*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.

32. Spiegel, M.R. and Stephens, L., *Schaum's Outline of Statistics (Schaum's Outline Series)*, McGraw-Hill, New York, 2007.

33. Spiegel, M.R., Schiller, J. J. and Srinivasan, R.A., *Schaum's Outline of Probability and Statistics*, McGraw-Hill, New York, 2000.

34. Wirsching, P.H., Paez, T.L. and Ortiz, K., *Random Vibrations*: *Theory and Practice*, Dover Publications, New York, 2006.

35. Yang, C.Y., *Random Vibrations of Structures*, Wiley Press, New York, 1986.

Chapter

8

Common Vibration Monitoring and Analysis Techniques

8.1 INTRODUCTION

Vibration signature analysis is the process by which intelligent information is extracted about machinery condition from the vibration data obtained during machine operation (operational characteristics) or from the response induced in the machine by an external stimulus (forced vibration). Historically speaking, vibration data have been obtained in the time domain using oscilloscopes, or recorders and the vibration engineer has had to use his/her observation/intuition to extract meaningful information from this data. Detection of machinery faults from raw time domain data is extremely difficult in view of the fact that there are generally a very large number of frequency components present in these signals. The use of conventional time domain analysis techniques was followed by autocorrelation and cross-correlation techniques. But these methods result in equally complex functions when the vibration signal consists of a superposition of a very large number of frequency components. This resulted in frequency domain techniques using narrow-band analysis methods, with constant bandwidth narrow-band analysers. These methods turned out to be a boon to engineers involved in diagnostics of machinery faults, since it was evident that the frequencies seen in a signal's narrow-band spectrum correlated with the various operating frequencies in the machine being monitored. It was found that shaft rotational rates and their harmonics, fractions of rotational frequencies, gear mesh frequencies, blade passing frequencies, system resonance frequencies, operational cycle rates, amplitude modulation (AM) processes, frequency modulation (FM) processes, etc. could be correlated to specific faults in machinery. With the proposal of the fast Fourier transform (FFT) algorithm by Cooley and Tukey, the current computer based FFT techniques caught on. With the boom in the electronics and signal processing industries in the late 1960s and early 1970s, real time analysers (analysers which process and display a spectrum while the next data set is being collected) and FFT-based spectrum analysers became popular. The main advantage of the FFT analysers was that they inherently had the phase information available. This is essential in the computation of FRFs and sometimes throws light on faults which are not diagnosed from the vibration magnitude alone. There soon followed software for modal analysis which allowed the animation of the dynamic motion of a structure on a computer monitor, offering the designer scope for redesign and determination of the new dynamic properties of the system.

This chapter gives an insight into the common faults in machines, their symptoms, various vibration monitoring and test methods available for structural dynamics, techniques for analysis and interpretation

of results, with greater emphasis on recent trends. Starting with the conventional time and frequency domain techniques, there is a progression to more sophisticated techniques like short-time Fourier transform, wavelet technique and Artificial Neural Networks (ANNs) for fault classification.

8.2 Sources of Vibration in Rotating Machines

Dynamic forces in machines are responsible for producing vibration and noise and lead to defects; these in turn change the features of the forces and hence the vibration and noise characteristics. Depending on the nature of the excitation forces, the vibration may be periodic or random.

8.2.1 Classification of Forces

The various dynamic forces in a machine that cause vibration and noise are classified as described below.

8.2.1.1 Mechanical forces The main forces of mechanical nature are:
- **Centrifugal forces** due to imbalances in rotating parts
- **Parametric forces** due to the fluctuation in rigidity of the rotating parts such as shafts, bearings and their supports
- **Frictional forces** that are the result of the contacting micro-roughness of rolling and slide friction surfaces
- **Shock forces** excited by the interaction of friction elements accompanied by their compliances

8.2.1.2 Electrical and/or magnetic forces The forces which are electrical and/or magnetic in nature are:
- **Magnetic forces** due to the variations in magnetic energy caused by the air gap fluctuation
- **Electrodynamic forces** due to fluctuations in the interaction of an electric current with a magnetic field
- **Magnetostrictive forces** caused by magnetostrictive effect, viz., the variation of linear dimensions of the magnetic material under the influence of magnetic field

8.2.1.3 Flow induced forces The dynamic forces due to flow are:
- **Lifting forces** associated with the movement of a blade or group of blades of an impeller in flow
- **Friction forces** on the boundary between the flow and fixed machine parts (e.g. the inner walls of a pipe)
- **Pressure pulsation forces** in a flow caused by turbulence or cavitation

8.2.2 Common Machinery Faults Requiring Diagnosis

Some of the common faults that show up in vibration spectra or can be diagnosed from orbits are described below. The description is based on the excitation forces causing these faults.

8.2.2.1 Unbalance This is one of the most common contributors to vibration in rotating machines and has been discussed in detail in Section 2.2.7. Unbalance, if left uncorrected, results in fatigue in support structures, excessive bearing wear and power losses. In variable speed machines or machines that must be coasted up to operational speed gradually, the effects of unbalance will depend on the shaft rotational speed. Unbalance results in a 1 × rpm vibration component with a steady amplitude. This problem may be

corrected by a procedure called dynamic balancing. A bent shaft or bowed rotor leads to a special form of unbalance, the spectrum due to which is similar to that resulting from regular unbalance.

8.2.2.2 Misalignment Vibration due to misalignment causes excessive radial loads on bearings, leading to their premature failure. The most common forms of misalignment are: (i) parallel or offset misalignment, (ii) angular misalignment and (iii) combination of both, this being the most common type. Parallel misalignment occurs when shaft centre lines are parallel, but offset from one another in the horizontal or vertical or both directions. Angular misalignment is due to shaft centre lines meeting at an angle. Though flexible couplings can tolerate some misalignment, it should be minimized. Misalignment can also occur between shafts and bearings or between gears.

8.2.2.3 Mechanical looseness Mechanical looseness may appear in different forms, the main types being: (i) a bearing loose on a shaft and (ii) a bearing loose in its housing. The former will give rise to a modulated time waveform with many harmonics, the period of modulation varying and the time signal also often being clipped. The latter will display a strong fourth harmonic. The spectrum due to looseness may also resemble that from a faulty rolling element bearing.

8.2.2.4 Rubs When there is excessive mechanical looseness or oil whirl, moving parts of machines come into contact with stationary parts, causing rubs. The resulting vibration signature is similar to that of looseness, but is drowned in wideband noise created by impacts. If the impacts are repetitive in nature, prominent spectral components show up at the impacting frequency. Quite often, rubs are due to a rotor pressing too hard against a seal. This causes the rotor to heat up asymmetrically and develop a bowed shape, generating a 1 × rpm component in the spectrum. This may be distinguished from an unbalance condition, since a peak due to rubbing is absent till such time the machine reaches normal operating temperature.

8.2.2.5 Soft foot Soft foot is a form of mechanical looseness, which may manifest as misalignment, unbalance, or a bent shaft. It occurs when one of a machine's bolts is not tight enough to restrain its motion, causing that part of the machine to cyclically lift from and rest on the foundation, depending on the forces acting on it. All the symptoms associated with mechanical looseness are present in this case also. If the foundation bolts are not in a plane, then tightening them will cause the casing and/or rotor to be distorted, producing vibration signatures corresponding to misalignment, unbalance and bent shaft. In order to ascertain if there is a soft foot, the vibration level must be checked while each bolt is loosened and again retightened and the appearance and/or disappearance of the above frequency components should be monitored. If the vibration levels of a machine cannot be reduced by realignment or balancing, soft foot could be the culprit.

8.2.2.6 Resonances Machines or their components are sometimes set into resonance by excitations from an adjacent machine. The characteristic feature of resonance is that it gives rise to a strong peak, the frequency of which is independent of the rotating speed of the machine. Resonance can be clearly seen during a coast up or coast down since it causes a drastic increase in the response when the exciting speed coincides with the running speed. The amplitude of response at resonance is limited only by the damping in the system. A highly damped system will not show large amplitudes, while a lightly damped one will show considerable dynamic magnification. The term 'critical speed' is used to refer to the resonant condition when the rotating speed of a machine is equal to its natural frequency. Most machines are

designed to operate well away from their critical speeds, but these frequencies can change due to changes in a machine's support condition.

8.2.2.7   Foundation problems   Any foundation is to be designed such that its natural frequency does not coincide with that of the machine it supports; if it does, a lot of energy is transmitted through the foundation to the earth and possibly into adjacent machines. All vibration frequencies below the natural frequency will also be transmitted to the earth. Besides, a foundation is also expected to be sufficiently rigid so as to keep the machine properly aligned. Care should be taken while changing foundations.

8.2.2.8   Oil whirl   Oil whirl is a common phenomenon occurring in journal bearings. It occurs due to the fluid film in a lightly loaded journal bearing exerting a non-uniform force on the supported shaft, resulting in an unstable operating condition. Oil whirl can be induced by light dynamic and preload forces, excessive bearing wear or clearance, a change in oil shear viscosity, oil pressure or oil temperature, change in internal damping, gyroscopic effects, etc. During oil whirl, the shaft pushes a wedge of oil in front of itself and moves with a circular motion within the bearing clearance at speeds slightly less than one half the shaft rotational speed (0.42–0.48 times the rpm). The rotor actually revolves around inside the bearing in a direction opposite to that of shaft rotation. This condition can be prevented by using tilt pads to ensure that the oil pressure gradient is sufficient to support the shaft. Due to the inherent instability of this phenomenon, there are a few revolutions during which oil whirl occurs and a few during which it does not, revealing an intermittent pattern or 'beating' effect in the time record. This phenomenon can be seen well in an orbit plot (Section 8.8) as well as in a spectrum cascade plot (Section 8.12.1). If there is persistent oil whirl, the bearing needs to be replaced.

8.2.2.9   Oil whip   Oil whip occurs when the frequency of oil whirl present in a machine coincides with its natural frequency and excites a resonant condition. This causes the system to remain locked to the constant natural frequency, irrespective of machine operating speed. Oil whip often occurs at twice oil whirl frequency, since at that frequency, oil whirl matches the critical speed. Dry whip occurs due to a lack of lubrication causing excessive friction between the stationary bearing and rotating journal. This friction can excite vibration in the bearing and other components. It can also result from the use of the wrong lubricant in journal bearings or due to excessive clearance as well as insufficient clearance.

8.2.2.10   Faults in rolling element bearings   This section deals with the fault diagnosis of rolling element bearings since they are very important components of a machine and since their degradation has detrimental effects on machine reliability. Bearing defects may be distributed as in surface roughness, waviness, misaligned races and off-size rolling elements. They may also be local such as cracks, pits and spalls on the rolling surfaces. Whereas in a fluid-film bearing, the shaft can experience motion relative to the bearing, a rolling element bearing has extremely small clearances which do not allow a significant amount of shaft motion relative to the bearing. Therefore, shaft relative measurements made for the diagnosis of fluid-film bearings with proximitors will not work for rolling element bearings. In the latter, forces from the shaft are transferred through the rolling elements to the bearing outer race and then ultimately to the bearing housing. Hence, a casing (bearing housing) measurement is normally acceptable for condition monitoring of machines with rolling element bearings. When using bearing housing measurements, two important points are to be considered: (i) signal amplitude versus transmission distance and (ii) susceptibility of measurements to noise. Therefore, it is preferable to make the

measurements as close to the machine vibration source as possible, the ideal distance being within 2.5–5 cm from the bearing.

Rolling Element Bearing Spectrum: The vibrations produced by machines with rolling element bearings occur in three frequency regions. These are described below.

Rotor Vibration Region: Rotor-related vibrations normally occur from 1/4 to 3 times shaft rotational speed (1/4 × to 3 × rpm) and are best measured in terms of velocity or displacement. Many rolling element bearing failures are the direct result of rotor-related faults such as unbalance, misalignment or rotor instability. Field studies reported in literature show that 90% of all bearing failures are related to either inner or outer race flaws and the remaining 10% to either rolling element or cage flaws, all of which generate frequencies appearing in the rotor vibration region. Most general purpose equipment with running speeds from 1200 to 3600 rpm generate rotor-related vibration signals between 20 and 600 Hz.

Prime Spike Region: The second frequency region is the Prime Spike (element passage) region. Local defects in rolling element bearings produce a series of impacts which repeat periodically at a rate dependent on bearing geometry, number of rolling elements and the speed at which the bearing is rotating. The magnitude of impact depends on the load and defect and these repetition rates are known as the bearing defect frequencies. These frequencies, typically in the range from one to seven times the element passage rate (Ep×) are generated even by a new bearing, but the amplitudes are small. Element passage rate is defined as the rate at which the rolling elements pass a point on either the inner or outer bearing race. Observation of these bearing-induced signals is used to diagnose rolling element bearing-related vibration problems and to determine what has failed in the bearing. Vibrations in this frequency range can be measured effectively in terms of acceleration, velocity or displacement.

High Frequency Region: The third region is the high frequency region covering frequencies from 5 to ~25 kHz and is measured in terms of acceleration. In this region we have the resonant frequencies due to the inner and outer races and rolling elements which are excited by the impacts. If high frequency region measurements are used for bearing failure detection, they should be used as supplement to measurements made in the rotor-related and prime spike regions.

Rolling Element Bearing Characteristic Frequencies: Regardless of type (ball, cylindrical, spherical, tapered or needle), rolling element bearings consist of an inner and outer race separated by the rolling elements, which are held in the cage. Defects may develop on any of these components. When the bearing rotates, each type of defect generates a vibration component of a particular frequency. By identifying the bearing characteristic frequency, the cause of the defect can be determined. These characteristic frequencies are in the same range as the low frequency vibrations (0–2 kHz) caused by the normal operation of a machine making it difficult to distinguish these peaks from the normal machinery noise. The bearing characteristic or defect frequencies are: ball pass frequency-outer race (BPFO), Ball pass frequency-inner race (BPFI), ball/roller spin frequency (BSF) and fundamental train frequency (FTF)/cage frequency. They are defined in terms of the following parameters.

d = roller diameter
D = pitch diameter
n = number of rollers
β = angle of contact
N_r = rotational speed (rpm)

Ball Pass Frequency-Outer Race (BPFO)

The balls or rollers passing the outer race generate the ball pass frequency-outer race given by the expression

$$\text{BPFO} = \frac{n}{2} \times \frac{N_r}{60} \times \left[1 - \left(\frac{d}{D} \right) \cos \beta \right] \tag{8.1}$$

Ball Pass Frequency-Inner Race (BPFI)

The speed of the ball/roller rotating relative to the inner race generates the ball pass frequency-inner race. This is described by the equation

$$\text{BPFI} = \frac{n}{2} \times \frac{N_r}{60} \times \left[1 + \left(\frac{d}{D} \right) \cos \beta \right] \tag{8.2}$$

Ball Spin Frequency/Rolling Element Frequency

Each of the balls or rollers within the bearing rotates around its own axis as it rolls around the bearing race. This spinning motion generates a vibration component at a frequency known as ball spin frequency.

$$\text{BSF} = \frac{D}{d} \times \frac{N_r}{60} \times \left[1 - \left(\frac{d}{D} \right)^2 \cos^2 \beta \right] \tag{8.3}$$

Fundamental Train Frequency (FTF)/Cage Frequency

Some friction normally exists between the rolling elements and the races, even with perfect lubrication. This friction is transmitted to the cage, causing it to rotate around the bearing races. Because this is a friction driven motion, the cage turns much slower than the inner race of the bearing and generates the FTF as it rotates around the bearing race. The FTF is less than one half of the shaft speed and may be expressed as

$$\text{FTF} = \frac{1}{2} \times \frac{N_r}{60} \times \left[1 - \left(\frac{d}{D} \right) \cos \beta \right] \tag{8.4}$$

8.3   TRANSDUCER CONSIDERATIONS

8.3.1   Choice of Vibration Transducer

Vibration may be measured in terms of displacement, velocity or acceleration as has been discussed in Section 3.2.1. Since we have large displacements at low frequencies, measurement of displacement is used for low frequency vibration components. Conversely acceleration is the most appropriate parameter for high frequency components. In applications where we are constrained to make measurements with a single transducer over a wide frequency band, vibration velocity measurement is preferred, since vibration velocities at low and high frequencies have equal weightage from a vibration energy point of view. In fact, the overall root mean square (RMS) value of vibration velocity measured over the frequency range 10–1000 Hz has been found to be a very good indicator of the severity of vibration of a machine. Most modern vibration meters come with electronic integrators, facilitating conversion of the acceleration signal to velocity and displacement.

Another consideration is the mass of the transducer, which becomes an important factor when measurements are to be made on smaller structures. A heavy transducer mounted on a small structure will cause mass loading and alter its dynamics. As a thumb rule, the transducer mass should preferably be less than one tenth the mass of the structure. Many other considerations such as location and orientation of the transducer ensure obtaining proper data. A transducer with the appropriate frequency range, dynamic range and sensitivity must be used. Temperature, dust, humidity and extraneous magnetic fields are other aspects to be looked into. All these are discussed at length in Chapter 3. Care regarding shielding and movement of cables should be exercised when using long cables. Improper instrumentation grounding is another factor that limits good data collection. The choice of transducer mounting method is important since some mounting conditions transfer the data more effectively than others. Convenience mounts such as magnetic mounts and wax mounting attenuate the high frequency information reaching the accelerometer and the user should be aware of this. This is discussed in Section 8.3.3. Proper mounting however cannot compensate for poor instrumentation practice; therefore appropriate signal conditioning amplifiers and data acquisition techniques should be used.

8.3.2 Location and Orientation of Transducer

The location of a vibration transducer is of utmost importance in machinery fault diagnosis. It is required to ensure that the site chosen for the transducer is clean and flat for an area at least as large as the base of the transducer. When selecting sensor locations, the vibration transmission path and the locations where vibrations corresponding to the natural modes may be excited are to be taken into account. It is recommended that vibrations be measured at several locations and directions since no amount of analysis will be able to extract the required features from a signal if the data is incomplete and if the vibration components relating to a given fault condition are not recorded in the first place. One of the preferred locations for the collection of vibration data is the bearing housing since the loads from the rotating components of a machine are transmitted to it through the bearings. Thus vibrations measured at the bearings give us information about rotational defects. Another factor that may affect the transducer location is its transmission sensitivity to high frequency vibrations; measurements made with a velocity or displacement transducer a few mm away from the chosen location will not have much effect on a large casing. However this may not be the case when making measurements with acceleration or ultrasonic transducers, as high frequency vibrations get attenuated even at small distances from the source, and especially so through housing interfaces. In the event that it is not possible to measure vibration directly on the bearing housing, the transducer location should be chosen for the best, most direct transmission path.

Regarding orientation of the transducer, it is a standard practice to measure vibration in three directions at each bearing housing: vertical-radial, horizontal-radial and axial. Most machines are usually better supported in the vertical direction than in the horizontal direction, leading to lower stiffness and hence higher vibrations in the horizontal direction. Hence if a single measurement is to be made, the direction chosen is usually horizontal. If the machine is less stiff vertically due to flexible supports, then this direction should be selected. At times, access may decide the location and direction of mounting a vibration transducer. Axial measurement is usually taken at the coupling end of the shaft.

8.3.3 Mounting the Transducer

For machinery condition monitoring applications, transducers may be fixed on the vibrating surface in a variety of ways, depending on the machine being monitored and the monitoring strategy employed. They are (i) permanent bonding, (ii) stud mounting, (iii) fixing with magnet, (iv) fixing using bee's wax, (v) fixing using double sided tape and (vi) holding with hand or with probe extension. Mounting accessories such as adhesive pads, mounting bases, magnetic bases, triaxial mounting cubes, mounting studs and stud adaptors are used for this. Ideally speaking transducers should be fixed with as rigid a mount as possible. The mounting of a transducer affects its frequency response, the mounted natural frequency being directly dependent on the stiffness of the mounting. The higher the stiffness, the higher is the mounted natural frequency.

$$\frac{f_1}{f_0} = \sqrt{\frac{m_0}{m_0 + m_1}} \tag{8.5}$$

where f_1 and f_0 are the resonant frequencies of the structure with and without the transducer mounted and m_1 and m_0 are the masses of the transducer and the structure, respectively. Figure 8.1 shows the response curves for an accelerometer with different mounts. The least stiff mounting technique is magnetic mounting and the one with the highest stiffness is stud mounting, with other mounting methods falling in between.
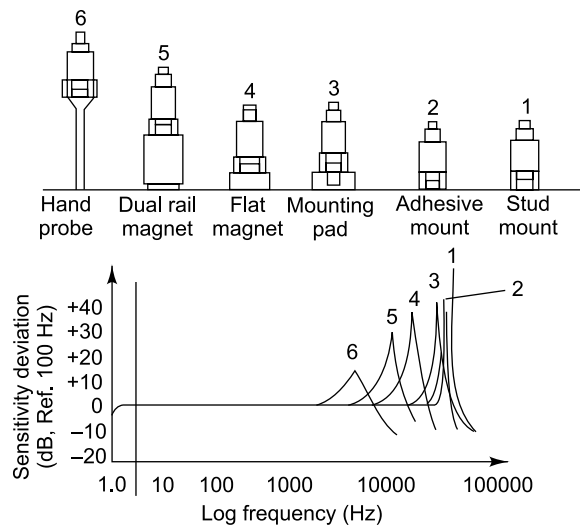
FIGURE 8.1 Accelerometer mounting considerations (Courtesy of http://www.kistler.com)

It is advisable to have a film of oil or Silicone grease between the accelerometer and the machine surface since it greatly enhances the coupling characteristics, thus improving the working frequency range. Sometimes an isolation stud and an isolation washer are used so that the measurement system can be grounded at one point only, preferably at the analyser. An additional ground at the accelerometer will provide another closed ground loop which may cause the accelerometer output to be noisy. It is advisable to apply a sealing compound in the region where the cable enters the accelerometer to protect the system from moisture-induced errors. The accelerometer end of the cable should be glued or strapped to the

vibrating machine, while the other end of the cable connected to the preamplifier, should have minimum vibration. This will minimize the noise in the measurement caused by dynamic bending, compression, or tension in the cable and also minimize cable and connector fatigue failures. If it is required to bend the cable, an adequate radius should be given to ensure proper strain relief.

8.3.3.1   Stud mounting   Permanent installations with studs or bolts offer the best form of sensor coupling on high speed machinery where high frequency measurements are required. The close coupling between the machine and the sensor in such mounting allows for best transmission of the vibrations to be measured to the sensor. This type of mounting is expensive and may not be practical if a number of measurements are to be made in the field with a portable instrument or in modal analysis involving roving response measurements for obtaining FRFs. Many accelerometers are specifically designed for stud mounting; they also allow the studs to be removed to facilitate both stud and adhesive mounting. The accelerometer should be tightened to the torque specified on the transducer's calibration certificate using a torque wrench. It should be ensured that the stud does not bottom out at the base of the accelerometer or the drilled hole. If the accelerometer does not have a tapped hole for stud mounting, an adhesive mounting pad may be used; this also provides electrical isolation at the cost of a lowered frequency range of measurement.

8.3.3.2   Direct adhesive mounting   Some accelerometers are specifically designed for adhesive mounting and hence do not require any special mounting adapters. A cyano-acrylate type adhesive provides an extremely thin bond with good frequency response and is therefore preferred to epoxies. This adhesive provides good, strong coupling of the sensor (especially large ones) to the specimen and allows use at higher temperatures than wax mounting. Accelerometers provided with tapped holes for stud mounting can also be used with adhesives, taking care to ensure that no adhesive enters the hole, thus making stud mounting still possible at some future time. Once fixed, it is difficult to remove the sensor and this is usually done using a wrench of proper size after the adhesive has been sufficiently softened by the manufacturer's recommended adhesive solvent or acetone.

8.3.3.3   Mounting with wax   Wax mounting is highly suitable for lightweight transducers in temporary installations where temperatures are not high. Petro-wax/bee's wax has long been used as a mounting agent for high frequency measurements. The former has been formulated such as to provide a better frequency response than the latter.

8.3.3.4   Adhesive double side tape   This offers a very convenient mounting solution for accelerometers with low acceleration rating and low weight only; it is not suitable for larger accelerometers or for long durations of time.

8.3.3.5   Magnetic mounting   Magnetic mounts are suitable for making measurements on ferromagnetic surfaces; however they produce mass loading and result in a smaller usable frequency range than the stud-mounted pickups. Magnetic mounts have the advantage of being easily movable and are highly suitable for experimental modal analysis requiring a roving pickup.

8.3.3.6   Holding by hand   The hand-held mount is the least expensive, but has the drawback that it offers a frequency response up to around 1 kHz only. Holding the transducer by hand or with probe extensions is useful when conducting general vibration measurements, but it usually does not give consistent results.

8.4 VIBRATION DATA COLLECTION ERRORS AND THEIR EFFECTS

It is to be remembered that the analysis results will be only as good as the data collected. Since the electrical signals produced by transducers are typically quite small (of the order of mV), it is of utmost importance that the data collected are noise-free and accurate. Some problems resulting in errors and ways of avoiding them are listed below.

Damaged transducer: If a transducer consistently measures vibrations considerably different from those felt by hand, it could be suspected that it is damaged internally. This can be verified by making measurements with an alternate transducer.

Poor cable connections and cable movement: These result in changes in the electrical resistance of connecting cables and adaptors and irregular variation in vibration amplitude. Damaged cables, damaged connectors, moisture or dirt entry and ineffective shielding also result in poor data. Low noise coaxial cables prevent pick up of extraneous noise by providing an electrostatic shield around the high impedance input lead of a piezoelectric accelerometer. Triboelectric noise generated by the physical movement of cables should also be minimized.

Poor accelerometer to magnet and magnet to machine fixing: It is imperative that the surface of the specimen be machined properly. The surfaces of the accelerometer, the magnet and the machine should be free of damage and should be cleaned, grease should be applied between the surfaces and the stud tightly coupled; if not, the transmission of vibration from the machine to the transducer will be poor.

Inadequate stabilization of readings: After fixing the probe on the measuring surface, it is required to ensure that the reading has stabilized before recording it, else there will be errors.

Fluctuating vibrations: In certain situations, vibrations may drift with time or may vary in a slow cyclic or unknown pattern. In such cases, vibrations are to be recorded when they are at their largest values.

Change in background vibrations: If there are background vibrations from nearby machines, they can affect the vibration levels on the machine being logged. The level and source of the background vibrations are to be noted and if these vary, measurements are to be made when the background interference is lowest.

Transducer temperature: If the equipment on which vibrations are being measured is very hot, a high temperature transducer should be used; if not, a thermally insulating pad can be placed between the transducer and the mounting pad/magnet to minimize the effect of temperature on the electrical characteristics of the transducer. The sensor contact time should also be minimized.

Vibration signature clipping: If the amplitude of the vibration signal is very high, there is a possibility that the signal conditioning amplifier will saturate and become overloaded, resulting in clipped waveforms. Hence the gain of the amplifier should be properly chosen or auto-ranging should be used.

Effects of cable length: The portable accelerometers typically used with data collectors have internal amplifiers and so cable length should not be a major factor. Where very long cables (length > 50 m) are used for permanent accelerometers, there may be an influence on outputs, so also in accelerometers with short cables, but external amplifiers.

Electrical interference: If the transducer cable shielding is damaged, then the measured signal may be affected by magnetic and electric fields and the cable may require replacement. Improper cable shielding and grounding of the instrumentation will result in a line frequency component at 50 Hz and its harmonics. Cables should not be run alongside high voltage power lines.

Accelerometer mounting: If the contact force between the transducer and the machine is insufficient, it may have a significant effect on the measured values, especially at high frequencies.

8.5   WAYS OF DISPLAYING VIBRATION DATA

Vibration signals may be processed to obtain relevant information and hence may be displayed in a variety of ways. Each display format is associated with a specific analysis tool and has its pros and cons. Some of the common ways of monitoring and displaying vibration data are listed below.

- Vibration amplitude versus time (time domain analysis)
- Trends of features extracted, i.e. analysis of RMS, kurtosis and other statistical values over time
- Shock pulse/spike energy versus time
- Orbit/Lissajous plots
- Vibration amplitude versus frequency (frequency domain or spectrum analysis)
- Bodé and Nyquist plots
- Envelope spectrum
- Cepstrum (quefrency domain)
- Vibration amplitude versus frequency versus time: 3-D plot called spectrum cascade or waterfall
- PSDs from parametric spectra
- Modal parameters, i.e. natural frequencies, damping ratios and mode shapes (modal domain)
- Spectrogram obtained from short-time Fourier transform (STFT)
- Spectrogram obtained from wavelet analysis

Besides, there are also a number of data classification techniques such as those using ANNs and support vector machines (SVMs).

8.6   TIME DOMAIN ANALYSIS

Most of the signals obtained from a machine using a recorder/oscilloscope/data acquisition system (Fig. 8.2) are time domain signals in their raw format, i.e. the signals have time as one axis (independent variable) and vibration amplitude as the other (dependent variable). Time domain analysis involves visually inspecting the measured time record to identify machinery faults. The time history reveals the general nature of the vibration signal, i.e. whether it is sinusoidal, random, repetitive or transient in nature. The main advantage of collecting and storing the time domain signal is that a lot of detailed analysis using any other descriptor may be done at a later stage using the raw data since no data have been lost prior to analysis. The flip side to it, however, is that there is often too much data (much of which is likely to be redundant and of little use) being collected for easy and clear fault diagnosis. This representation is not always the best for most diagnostics related applications since most of the time the information characterising the faults is hidden in the spectral content of the signal.

Though it is difficult to make out individual frequency components from a time record (unless it is a harmonic signal), such a signal is useful for studying instantaneous or transient types of vibration signals. Some mechanical problems, in spite of having different dynamic behaviour, may result in identical frequencies. Hence the spectrum may not provide all the information needed to make a complete diagnosis and it is therefore a good idea to examine the time waveform. For example, the vibration due to unbalance results in a 1 × rpm vibration component, as also does a gear with a defective tooth (Fig. 8.3a and b).

From a frequency domain analysis, we cannot differentiate between the two. However, the differences are clear in the time domain, unbalance producing a sinusoidal waveform and the defective tooth showing a distinctive spike-like appearance owing to a short duration contact. Likewise, the spectra of an impulse or transient and of a random signal may look almost exactly alike, even though their time domain representations are very different in nature. Transient phenomena such as coast-up and coast-down, are most easily captured and analysed using time waveforms. Some common faults captured by time domain plots are as described below.

(a)   (b)
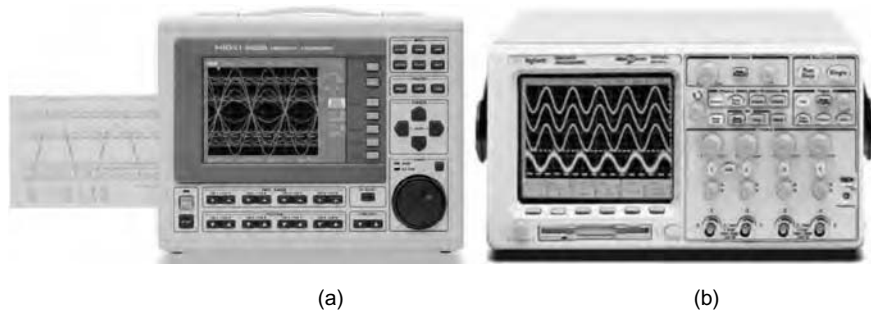
FIGURE 8.2   Instrumentation for time domain analysis: (a) data recorder (Courtesy: http://www.hioki.com), (b) oscilloscope (Courtesy: http://www.home.agilent.com)

Periodic impacts: They are caused in rolling element bearings when the rollers encounter a crack or small spall in a race. In the presence of external noise, the spectrum of this signal may not show a well-defined peak, however, the acceleration record will usually exhibit repetitive peaks with their repetition rate equal to the period of the race defective frequency or bearing ball pass frequency.

Low-frequency events: When there are two or more adjacent machines running at almost the same speed, it results in a beat phenomenon with the sum and difference frequencies present. The latter may not be clearly seen in the spectrum since it corresponds to a very low frequency. However it is clearly seen in the time record as an amplitude modulated signal (Fig. 8.3c). Sometimes the vibration signal has a discontinuity once in a while, resulting in so low a frequency that it will not be clearly seen in the spectrum. A waveform recorded over a long time will however show the discontinuities. An example of this is a low speed gearbox which has one broken or cracked tooth on the large gear.

(a)   (b)   (c)   (d)
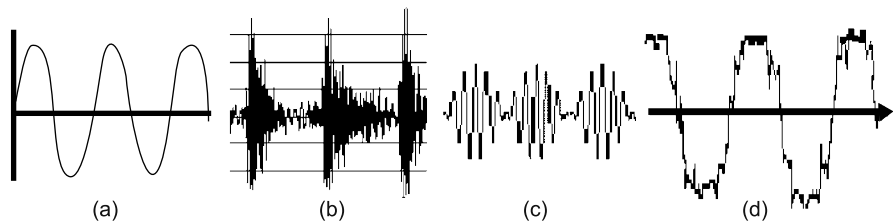
FIGURE 8.3   Typical machinery time records: (a) unbalance, (b) bearing outer race fault, (c) beat phenomenon, (d) looseness

Truncation or flattening of the signal: In many cases of looseness, such as that of a bearing pillow block that rises up a little during part of the rotation and then contacts the base for the rest of the cycle, the waveform will get flattened on one side as in Fig. 8.3(d). This will manifest in the spectrum as harmonics

which are indistinguishable from other types of waveform distortion which also produce harmonics. From the time waveform a quick identification of this kind of looseness where motion is restricted in one direction may be made.

Random or non-periodic impacts: When a loose machine component impacts something at a rate unrelated to machine speed, it generates impacts that are usually non-periodic or essentially random. The vibration spectrum corresponding to such a condition has a continuous spectrum extending to very high frequencies. Though this spectrum resembles that of other sources of wideband noise, the impacts are very clearly seen in the time domain waveform.

8.6.1   Time-Synchronous Averaging (TSA)

Averaging of the time-domain vibration signal in synchronization with the running speed of the machinery being monitored is called time-synchronous averaging. Many present-day FFT/signal analysers have this capability. Synchronous averaging is different from the usual spectrum averaging done in FFT analysis in that the waveform itself is averaged in time. In this mode, a tachometer input is used to trigger or provide a reference signal corresponding to one pulse per revolution of the machine and a synchronous averaging process is done over many machine cycles. Alternately the reference signal may be obtained from a proximitor sensing key phasor information to give out pulses on sensing protrusions or depressions in the shaft/rotor. This way, the time samples are synchronized in the sense that they all begin at exactly the same angular position of the machine. This process removes the contribution of asynchronous vibration signals such as background noise, structural resonances, random transients and any other components from other rotating parts of the machine (with a different speed) present in a time domain average, since they are not synchronized with the repetition rate of the trigger. The averaged time record gradually enhances the contribution of the rotating components that are synchronous with the trigger. The setup parameters to be used for TSA are generally quite different from those for spectrum analysis. The two important considerations in setting up TSA are the length (duration) and resolution of the time record. The length typically considered is slightly longer than one revolution of the shaft, or one cycle of the event being studied; any duration longer than this results in loss of resolution in the time domain.

For a sampling frequency f_s, the time interval between two samples is $\Delta t = 1/f_s$. If the obtained vibration signal $y(t)$ consists of R records (revolutions) and each record has L data samples, then the TSA signal $x(t)$ can be obtained as

$$x(t) = \frac{1}{R} \sum_{r=0}^{R-1} y(t + rL\Delta t), \quad t = k\Delta t, k = 0, 1, \ldots L-1 \qquad (8.6)$$

The number of records R is chosen depending on the noise reduction requirement and computer memory limitations. Generally, it should be taken as large as possible. Besides, if the number of samples per record is not exactly equal from one record to another, a resampling process should be done in each record to eliminate shaft speed variances. Figure 8.4 shows the effect of time-synchronous averaging.

An important application of time-synchronous averaging is in the waveform analysis of machinery vibration, especially in the case of gear drives. If the output of a tachometer connected to the pinion shaft of a gearbox is used to trigger a time-synchronous averaging procedure, it will make the events related to the pinion revolution rate stand out, removing all other vibration components. Vibrations caused by the driven gear will also be removed, and the resulting time record will show the vibration

caused by each individual tooth on the pinion. This technique is also extremely useful for monitoring machines where there are multiple shafts operating at only slightly different speeds and in close proximity
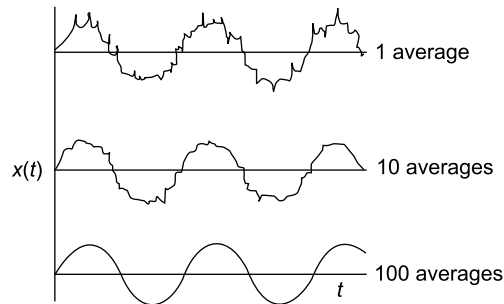
FIGURE 8.4   Effect of time-synchronous averaging

to one another. This is the only type of averaging that actually does reduce noise. However, it must be emphasized that synchronous averaging is not suitable for use in finding modulation effects such as sidebands and other conditions where a high resolution is required in the frequency domain.

8.7   STATISTICAL/PROBABILISTIC DESCRIPTORS OF VIBRATION SIGNALS

From the time record of a vibration signal, some indices which are useful in diagnostics can be extracted and are described below and also in Section 7.5. The most commonly used indices are the statistical parameters which can be computed from the raw vibration signal and which highlight differences between records, making them useful for fault diagnosis and trending. Unfortunately these parameters cannot pinpoint the faulty component in a machine, since they are affected by vibrations of all the components of the machine. Some of these parameters are: peak, mean, RMS, crest factor, skewness, kurtosis, K-factor, impulse factor, shape factor, probability density histogram and cumulative distribution. It is customary to plot these features at regular intervals of time (trend monitoring); significant variations of these features from their reference or baseline values (measured in normal condition) will indicate that the system has faults.

Vibration meters (Fig. 8.5a) are small, portable, inexpensive, simple to use instruments and may be used for measuring overall vibration levels or many of these other descriptors. They are usually equipped with a velocity and/or acceleration transducer. Some of them have built-in diagnostics capability using FFT analysis; in others, there is provision for taking the measured time record out to an analyser. A spike energy/shock pulse meter (Fig. 8.5b) is a device with an accelerometer tuned to have its resonant frequency (typically 32 kHz) excited by impacts produced by defective bearings, thus giving early warnings of deteriorating bearings. High speed rolling element bearings, in the early stages of failure (while producing pitting) emit vibration energy in a rather high, but well defined, frequency band. These accelerometers are designed to have their natural frequencies much higher than the expected frequency components of the response to be measured. The only drawback is that response from this meter is load-dependent and care has to be taken to ensure that measurement conditions are constant; if not, spurious results may be got.

FIGURE 8.5 Instruments for measuring vibration: (a) portable vibration meter (Courtesy of http://www.rion.co.jp), (b) shock pulse meter (Courtesy of http://www.bdi-usa.com)

8.7.1 Peak (Maximum) Value

The peak value of a signal is often used as one of the features for diagnosis. This value only indicates the maximum value, without any consideration of the time history of the wave. It is nevertheless a significant parameter for quantifying the level of short-duration shocks, which are typically half-sine in nature. The peak-to-peak value indicates the maximum excursion of the vibration signal and comes in handy where, say, the displacement of a machine part is important from the point of view of mechanical clearance or maximum stress. The peak value increases with the growth of faults.

8.7.2 Mean Value

Large variations of the mean from the reference value may indicate that the system has entered an abnormal state. It is however to be noted that the mean value of a signal can be obtained in the DC coupled mode only. Most signal conditioning amplifiers offer the AC coupled mode, in which case the mean value is always zero. The mean value of a time record may be computed using the following expression:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{8.7a}$$

in the case of discrete-time signals. Here, x_i represents the ith value of the discrete-time signal and N the number of measurement samples. For continuous-time signals it is given by

$$\bar{x} = \frac{1}{T} \int_{0}^{T} x(t)\, dt \tag{8.7b}$$

8.7.3 Root Mean Square (RMS) Value

The RMS value is computed taking into account the time history of the wave. It is a measure of the energy content in the vibration signature and hence is one of the most relevant statistical parameters for

machinery fault severity. This feature is good for tracking the overall vibration level, but it does not provide any information on which component is failing. It can be very effective in detecting a major out-of-balance in rotating systems. RMS value also increases with the appearance of shock pulses (bearing related signal). The equations below may be used for calculating RMS values of discrete and continuous-time signals.

$$\text{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{8.8a}$$

$$\text{RMS} = \sqrt{\frac{1}{T}\int_{0}^{T} x^2(t)dt} \tag{8.8b}$$

It is of interest to note that the RMS value of a signal is equal to the standard deviation for 0-mean signals.

8.7.4 Crest Factor

The crest factor also sometimes called the 'peak-to-RMS-ratio', is defined as the ratio of the peak value of a waveform to its RMS value and is therefore a dimensionless quantity. The following expression defines the crest factor.

$$\text{CF} = \frac{x_{\max}}{x_{\text{RMS}}} \tag{8.9}$$

The crest factor of a sine wave is $\sqrt{2}$, or 1.414. A typical vibration signal from a machine with a large imbalance, but no other problems will have a crest factor of about 1.5, but as the bearings begin to wear, resulting in impacts, the crest factor becomes much larger than this. The reason why the crest factor is so sensitive to the existence of sharp peaks in a waveform is that the peaks are short-lived and therefore do not contain much energy. This quantity is used extensively in bearing diagnostics and is effective for the simple reason that as a bearing deteriorates, the peak levels of acceleration increase more rapidly than the RMS levels due to impacts. Crest factor is easily measured and is relatively insensitive to bearing speed and load. Though it increases under bearing faults, it may decrease in the final stage of the fault due to a considerable increase in the RMS value with progressive wear. Examples of actual crest factors measured on an off-road vehicle have shown values as large as 12. In other words, the peak value is 12 times the RMS value.

8.7.5 K, Impulse and Shape Factors

K factor is defined as the product of the maximum (peak) value and the RMS value of the time signal. It increases under bearing faults and may be found from the following equation

$$\text{KF} = x_{\max} \times x_{\text{RMS}} \tag{8.10}$$

The impulse factor which is also found to be an indicator of bearing faults is defined as the ratio of the peak value to the mean value of the time signal and may be computed as

$$\text{IF} = \frac{x_{\max}}{\overline{x}} \tag{8.11}$$

Shape factor is defined as the ratio of the RMS value to the mean value of the time signal. It presents changes under unbalance and misalignment. The following expression is used for calculating the shape factor.

$$\text{SF} = \frac{x_{\text{RMS}}}{\overline{x}} \tag{8.12}$$

8.7.6   Probability Density Functions

As described in Section 7.3.1 this represents the probability that the instantaneous amplitude of a vibration signal lies within a certain amplitude range. Typically, the shape of the probability density functions got from random vibration records on a machine without fault are similar to a Gaussian (or normal) probability distribution. Fault conditions give rise to different shapes, the shapes depending on the mean value and the standard deviation of the vibration record. Figure 8.6 shows two probability density functions, one for a good bearing and one for a faulty bearing. An impulsive signal such as that got for rolling element bearing faults gives rise to a probability function with a high probability around the mean and a wide spread about it with low probabilities. This type of display format can be used for condition trending and fault diagnostics.
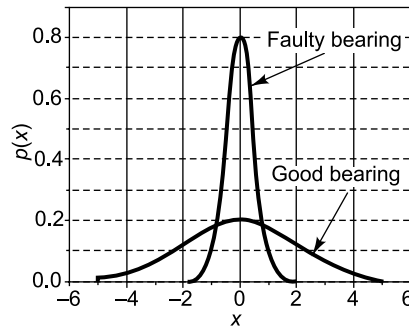
FIGURE 8.6   Probability density functions of bearing signals

8.7.7   Probability Density Moments

They are computed using the probability density function. The first moment of the probability density function about the $p(x) = 0$ axis gives the mean, while the third moment about the mean gives skewness which is a measure of symmetry, or to be more precise, the lack of symmetry. The second and fourth moments about the mean yield standard deviation and kurtosis, respectively. Both these values are indicative of the spread of the distribution about the mean. Perhaps the most useful of these indices is kurtosis which is a measure of whether the data are peaked or flat relative to a normal distribution. Thus, data sets with high kurtosis have a distinct peak around the mean, decline rather rapidly and have long tails; data sets with low kurtosis on the other hand, tend to have a rather flat top about the mean. A uniform distribution would be the extreme case. The probability density function can effectively show both the skewness and kurtosis of a data set.

8.7.8   Skewness

Skewness is the degree of symmetry of a probability distribution. It may be defined as

$$\beta_2 = \frac{1}{N\sigma^3} \sum_{i=1}^{N} \left(|x_i| - \overline{x} \right)^3 \tag{8.13}$$

where $\bar{x}$ is the mean, σ is the standard deviation and N is the number of data points. The skewness for a standard normal distribution is zero, and any symmetric data should have a skewness near zero. Data that are skewed left (having a left tail which is long relative to the right tail) have negative values of skewness, while data that are skewed right have positive values of skewness.

8.7.9 Kurtosis

Kurtosis is obtained from the fourth order central moment (moment about the mean) of amplitude probability distribution and is defined as

$$\beta_1 = \frac{1}{N\sigma^4} \sum_{i=1}^{N} \left(|x_i| - \bar{x} \right)^4 \tag{8.14}$$

It has been found that kurtosis is sensitive to the spiky character or peakiness of vibration signatures and therefore to the kind of vibration signal generated in the early stages of a rolling element bearing fault. It has therefore been used extensively for bearing vibration diagnostics. The principal advantage of this technique is that the measurement is insensitive to loading as well as the size of the bearing and the shaft speed. In general, vibration signals from a good bearing have a Gaussian distribution. Analysis of the broad frequency band of vibration signals has shown that the kurtosis value for an undamaged bearing is always $3 \pm 8\%$ and increases in value during the early stages of a fault. However, as a rolling-element-bearing degrades, the vibration signal becomes more random in nature and the impulsiveness reduces, giving rise to a fall in kurtosis value.

8.7.10 Autocorrelation Analysis

This is used to measure how well a signal correlates with a delayed version of itself. The autocorrelation function (also discussed in Section 7.5.2) can be used to enhance periodicity since it separates this type of signal from broadband noise. In the context of machinery diagnostics, the cross-correlation function is used for studying the correlation between two vibration signals or between acoustic and vibration signals and to correlate and identify the time-varying spectral properties common to the two sources. Cross time-frequency distributions, in particular, the cross spectrogram can be obtained from the cross-correlation for this.

8.8 LISSAJOUS PATTERN/ORBIT PLOTS

Lissajous or orbit plots are display formats which are very convenient for observing relative motion of journal bearings. Such plots throw light on bearing wear, shaft misalignment, shaft unbalance, lubrication instabilities (oil whirl, oil whip) and seal rub and hence constitute a very powerful monitoring and diagnostic tool, especially for slow speed machinery. These plots depict the locus of the centre of a rotating shaft and can be obtained on an oscilloscope screen/recorder by connecting the outputs of two radial non-contact pickups (typically eddy current probes/proximity probes/proximitors) to the horizontal and vertical inputs, which are 90° apart. The signal from the horizontal pickup is applied to the horizontal input of the oscilloscope moving the dot horizontally, while the output of the vertical pickup is applied to the vertical input, moving it vertically. These transducers are mounted in a sleeve concentric to the shaft or rotor, the vibrations of which are being sensed. This results in a plot of the total motion or orbit

of the rotor or shaft within the bearings. Diagnostic information is obtained from the shape of these lissajous figures or orbits. Thus, when the shaft rotates, the dot remains stationary if the centre of the shaft is stationary in the bearing. If the shaft moves around in the bearing, the dot follows it, tracing out the motion in real time. It is a common practice to use four transducers around the shaft as shown in Fig. 8.7. The two horizontal transducers are connected to the X-plates and generally used in differential connection to produce an additive effect; so also the two vertical transducers are connected to the Y-plates in a differential configuration. This arrangement helps to improve the sensitivity and to reduce distortion.
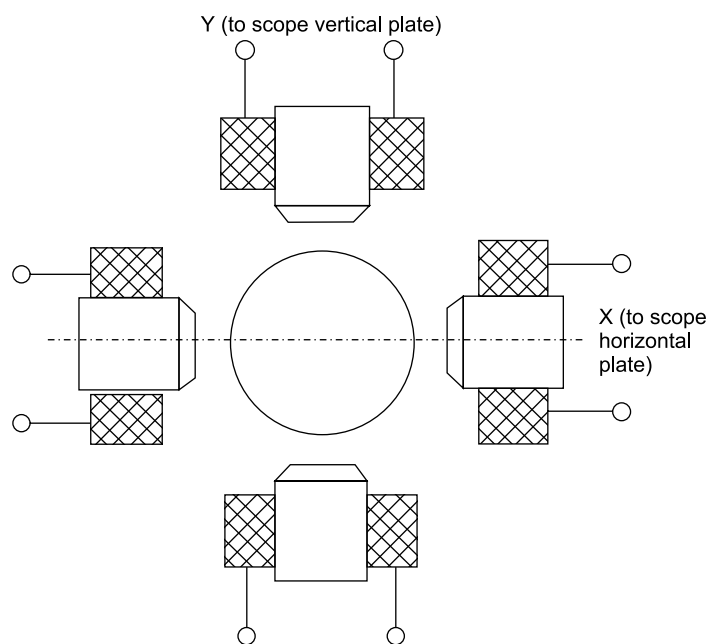
FIGURE 8.7   Proximity probe setup for orbit analysis

8.8.1   Diagnosis Based on the Shape of the Orbit

If the horizontal and vertical vibration magnitudes are equal and are at the same frequency, the orbit is a circle, indicating that there is some imbalance, causing the dot to move in a circle. Figure 8.8(a) depicts the orbit without any phase information. Figure 8.8(b) shows a phase marker superimposed on the orbit. There are two dark spots seen in the figure indicating that the frequency corresponds to a 1/2 × rpm component. Figure 8.8(c) shows an orbit in ideal condition of the shaft in the bearing with the horizontal and vertical vibration amplitudes being equal, while the orbit in Fig. 8.8(d) indicates that the vibration in the vertical direction is less than that in the horizontal direction. In the case of bearing vibration, this may mean that the bearing is worn in an oval pattern, with less vertical clearance than horizontal clearance. In the case of a rotor, it indicates that the rotor has better support (with higher stiffness) in the vertical direction than in the horizontal direction. Figure 8.8(e) shows a highly elliptical orbit which could be due to misalignment, wear or unbalance. The banana-shaped orbit in Fig. 8.8(f) could be indicative of misalignment. Figures 8.8(g) and (h) are typical patterns resulting from coupling

misalignment. The former orbit indicates shaft motion moving in the form of number 8 meaning that the shaft is vibrating twice as fast in one direction as in the other. This can be caused by excessive clearance in one direction, or a bearing worn into an oval shape. A pattern like this indicates a serious condition since the journal is likely to develop metal-to-metal contact with the bearing, causing extensive damage to both. Figure 8.8 (i) indicates the orbit corresponding to an oil whirl pattern, with the inner circle moving slowly within the outer circle. Figures 8.9(j)–(l) indicate rubbing, the last being very severe.

(a)   (b)   (c)   (d)

(e)   (f)   (g)   (h)

(i)   (j)   (k)   (l)
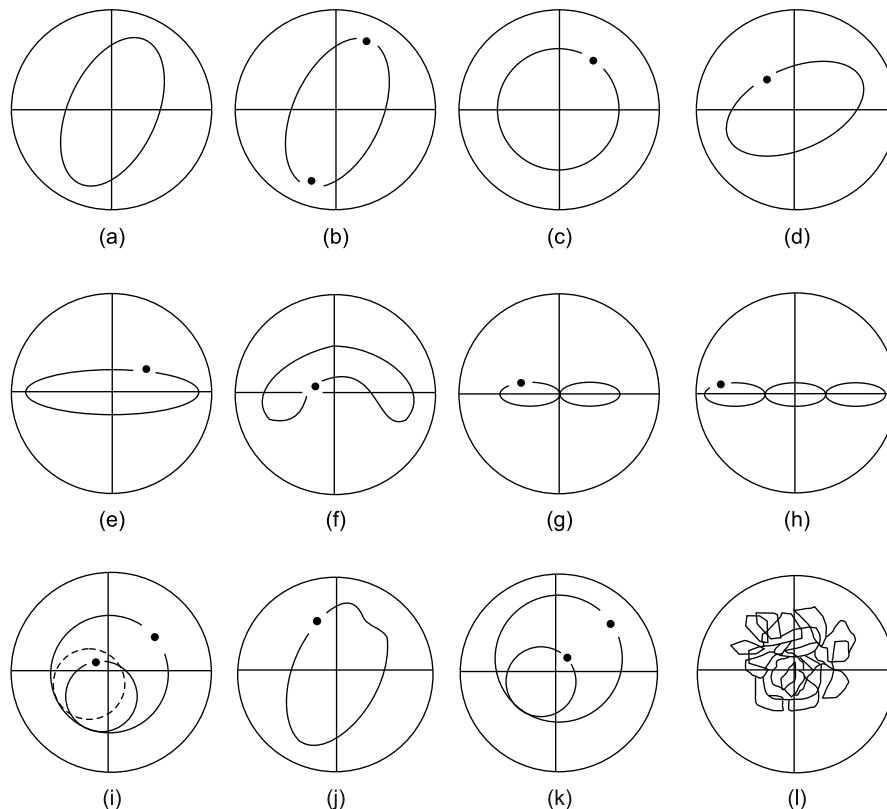
FIGURE 8.8 Typical orbits: (a) orbit alone, (b) orbit with superimposed reference signal, (c) and (d) unbalance, (e) and (f) misalignment, (g) and (h) coupling misalignment, (i) oil whirl, (j), (k) and (l) rubbing

8.9 PHASE (RELATIVE MOTION) ANALYSIS IN THE TIME DOMAIN

Phase is nothing but an indication of the time difference between a reference event and an event being studied. The concept of phase is interesting in that it can be considered a property of the time domain waveform or a property of the frequency spectrum. It can be depicted well in both time and frequency domain plots, the time axis information in an oscilloscope being used for the former and the phase from the complex spectrum in the latter. It is usually represented in milliseconds or microseconds in the time signal, and as degrees or radians in the spectrum with one complete rotation corresponding to 360° of

phase angle. There are many different machinery and structural problems, where phase information can provide more insight than magnitude, dynamic balancing being one of them. Phase analysis can help distinguish bending vibration from torsional vibration. In such cases, phase difference between two points on a machine reveals useful information. For instance, if a machine with an overhung rotor has a strong frequency component corresponding to its rotational rate, the cause of vibration could be imbalance, wobble, or misalignment. Observing the magnitude will not help, whereas measurement of phase at two points on opposite sides of the coupling can help identify the actual problem. Imbalance and/or wobble will result in in-phase vibrations at 1 × rpm at the two measurement points, while misalignment will cause out-of-phase vibration. Comparative phase readings are obtained by the use of two transducers, keeping one fixed at the same location and moving the other; alternately, they can be obtained by simply moving the vibration transducer one location after another, while maintaining the same angular reference. When doing dynamic balancing, the vibration of a rotating machine is recorded with a tachometer signal or key phasor providing the reference phase information. A reference mark is placed on the rotor at the end of a shaft or at some other location, or an existing key or keyway may be used. A high intensity stroboscopic light triggered by the measured vibration can also be used to provide a quick and convenient means of obtaining phase readings as discussed in Section 5.6.1. Phase information is also used extensively in modal analysis for reconstruction of mode shapes from measured FRF data.

8.10 Mode Shape Analysis

Experimental modal analysis involves finding out the natural frequencies, damping ratios and mode shapes through experiments called modal tests. It does not fall under traditional condition monitoring, but quite often it is found that unless the natural frequencies and mode shapes of machines and their foundations are fully understood, a complete and accurate assessment of machinery condition is not possible. A complete overview of modal analysis is provided in Chapter 10. Typically, in modal analysis a force is imparted to a structure through impact or shaker excitation and the response is measured. Operational deflection shape (ODS) analysis is a specific form of modal analysis in which an excitation force is provided to a structure or machine due to its own normal operation; only the response then has to be measured in this case and not the forcing function. Vibration sensors may be placed at or moved around critical locations and with a reference phase signal synchronizing all the recorded signals, animation of the structure's deflection under normal operation can be generated. These animations, along with spectral information can provide useful information for the diagnosis of structural or machinery problems. This is a very convenient method for in-situ applications or for huge structures like bridges, which are difficult to excite through conventional means.

8.11 Frequency Domain Analysis/Spectrum Analysis (Amplitude versus Frequency)

A plethora of mathematical transformations exist to 'transform' time signals to obtain 'processed' signals which reveal information that is not easily seen in the raw signal, transformation to the frequency domain being the most common amongst these. The time-domain vibration signal is transformed into

the frequency domain by applying a Fourier transform, using an FFT algorithm. In this method, the energy in the original signal is separated into its various frequency components and amplitude versus frequency representation of that signal is obtained. The principal advantage of this format is that any periodicities in the vibration signal are clearly displayed as peaks in the spectrum at the corresponding frequencies. This allows for faults, which usually generate specific characteristic frequency components in the vibration signal, to be detected early, diagnosed accurately, and trended over time as the condition deteriorates. However, the disadvantage of frequency domain analysis is that a significant amount of information (transients, non-repetitive signal components) may be lost during the transformation process. This information is non-retrievable, unless a permanent record of the raw vibration signal has been made. This technique is the most commonly used one in machinery diagnostics and 85% of mechanical problems in rotating machines can be identified by it. Spectrum averaging may be required if there is a fluctuation in the vibration spectrum.

There are two types of spectra which are generally observed in frequency analysis. The first has all the power at very clearly defined frequencies, and is referred to as line spectrum (see Fig. 1.5); this is typical of faults producing discrete frequency components. The second has all possible frequencies within the range of interest and is called a continuous or distributed power spectrum (Fig. 7.17); this is typical of jet noise or a seismic record.

Another important point to be considered is whether to use a linear or logarithmic scale for the spectral amplitude. If all the values are nearly equal, then a linear scale could be used. If the values however are spread over several orders of magnitude, then a logarithmic scale has to be used. This scale has the advantage of boosting the lower amplitudes. Regarding the appropriate scale for the frequency axis, a linear scale may be used for vibration analysis and a logarithmic scale to mimic the human ear for applications in acoustics (see Section 8.11.1.2).

8.11.1   Frequency Analyser

This is one of the key instruments for diagnostics. Different faults in machines, such as unbalance, misalignment, looseness, bearing flaws, etc. generate characteristic frequency components in the spectrum which are captured by the frequency analyser. To obtain a spectrum, a set of bandpass filters is used; the spectrum analyser displays the RMS value of energy present in each band as a function of band centre frequency. Every bandpass filter has a specific central frequency f_0 and lower and upper cutoff frequencies f_1 and f_2 on either side. Ideally, it passes all the frequencies in the band of interest without attenuation, i.e. it has a flat frequency response and completely eliminates all other frequencies with infinite attenuation at the cutoff frequencies. Practical filters however do not have a flat top, owing to a certain amount of ripple which is present. Besides, they do not have sharp cutoffs or vertical sides, but a gradual roll off. The bandwidth B of the filter is defined as the difference between upper and lower cutoff frequencies, i.e. $B = f_2 - f_1$. The 3 dB or half-power bandwidth is the difference between the frequencies where the passband level is 3 dB below or 0.707 times the value in the flat region. Noise bandwidth is another quantity used to characterize a filter and is defined as the bandwidth of an ideal filter with the same passband level and the same area (area under the passband curve) as the real filter. Figures 8.9(a)–(c) show the passband characteristics of an ideal and practical filter. Figure 8.9(c) shows the concept of noise bandwidth.
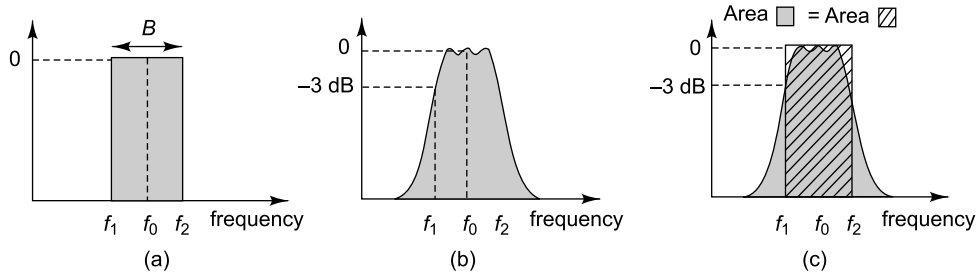
FIGURE 8.9 Bandpass filter: (a) ideal, (b) practical, (c) definition of noise bandwidth

8.11.1.1 *Sequential and parallel analysers* Two types of analysers have conventionally been used: (i) sequential analysers and (ii) parallel analysers. In the first method, a tunable bandpass filter is slowly swept through the entire frequency range of interest and the RMS value of the signal in each band is found using an RMS detector. One has to ascertain that the sweep is done sufficiently slowly to enable all frequency components to be captured. This type of sequential analysis is very time consuming, especially when one has to analyse over a very wide frequency range. For the analysis of transient signals, one has to ensure that the signals have been recorded using a suitable recorder; the signal will have to be replayed a large number of times to get the various spectral amplitudes. In the second method, namely the one using parallel analysers, the signal to be analysed is fed simultaneously to a large number of parallel bandpass filters (either hardware or software) with different centre frequencies covering the frequency range of interest. Though this method ensures that the analysis time is minimal, the hardware becomes bulky and expensive.

8.11.1.2 *Constant bandwidth and constant percentage bandwidth analysers* The resolution of the spectrum is dependent on the bandwidth of the filters used for analysis. There are two types of band pass filters available for analysis: (i) constant bandwidth filters and (ii) constant percentage bandwidth filters. In the first type, the bandwidth of the filter is independent of the selected centre frequency. It gives a uniform resolution on a linear frequency scale. Depending on the frequency resolution Δf required, a user may choose from a wide set of Δf values provided in an analyser. Bandwidths typically used in the early sequential analysers are 3.16, 10, 31.6, 100, 316 and 1000 Hz or a constant percentage of the frequency range. This type of constant bandwidth analysis is useful when one is trying to pinpoint a fault in a machine or when one requires spectral information for the sake of design modifications. In the second method, the bandwidth is a constant percentage of the centre frequency. Hence the bandwidth is smallest corresponding to the lowest centre frequency and goes on increasing as one proceeds up the frequency range. It is highest for the largest centre frequency; such filters are not suitable for machinery diagnostics, but find applications in the estimation of the subjective response of human beings to, say, airport noise, industrial noise or traffic noise. They are also used for estimating the perception of vibration by a human being, i.e. ride comfort in a vehicle since a human being's response to such external stimuli is logarithmic, i.e. the human being is more sensitive to ratio changes rather than changes in absolute value. The use of constant percentage bandwidth filters gives uniform resolution on a logarithmic scale. The octave and one-third filters which are extensively used in acoustics and measurement of human body vibration are constant percentage bandwidth filters. Figures 8.10(a) and (b) illustrate the concepts of constant bandwidth and constant % bandwidth analysers.
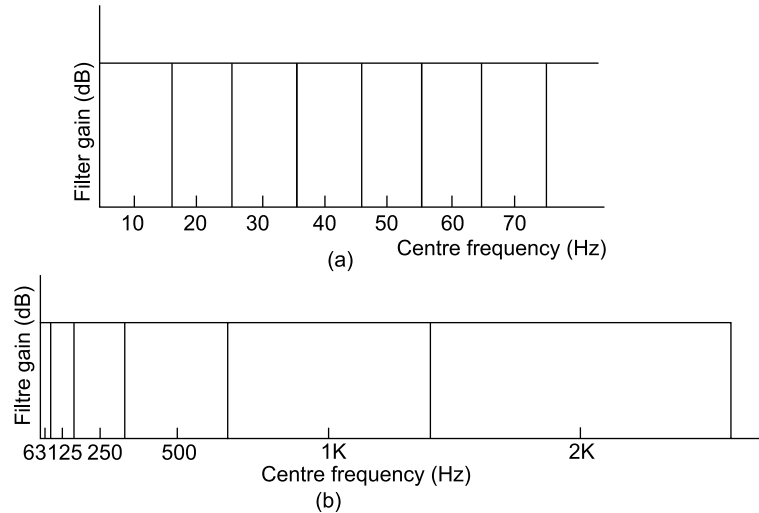
FIGURE **8.10**   Filters used in frequency analysis: (a) constant bandwidth, (b) constant % bandwidth

8.11.1.3   *FFT analysers/Dynamic signal analysers*   Today the trend is to use FFT/signal/real time analysers where the FFT algorithm simulates a set of parallel filters to deliver a spectrum in real time, converting a time domain signal into a combined magnitude and phase representation of the signal in the frequency domain. Though the Fourier transform is reversible, at any given time, only either of these records is available. That is, no frequency information is available in the time-domain signal, and no time information is available in the Fourier transformed signal. Stationary signals, the frequency contents of which are time-invariant, do not require both the time and frequency information at the same time. In addition to signature analysis, the FFT processing of data has resulted in another major application namely modal analysis, which involves the measurement and display of the dynamic characteristics of a structure. Generally, the features available in most modern-day spectrum analysers are: spectrum analysis with superior frequency resolution, filtering ability (including anti-aliasing), weighting/ window functions for the elimination of leakage, averaging capabilities (both in the time and frequency domains), envelope detection (demodulation), transient capture, large memory, order tracking, cascade/waterfall display and zoom features. All the display formats discussed in this chapter also feature in most analysers. Dual and four channel analysers are commercially available. Figure 8.11 shows an FFT analyser. Issues relating to the length of the time record, time domain resolution, FFT analyser setup for waveform collection, etc. are discussed in Section 6.8.

FIGURE **8.11**   FFT analyser (Courtesy of http://www.home.agilent.com)

8.11.2   Typical Machinery Fault Frequencies

Table 8.1 shows some common machinery faults, their predominant frequency components and other salient aspects of their vibration.

TABLE 8.1 Machinery vibration troubleshooting chart (Courtesy of Wavetek Scientific, Inc., U.S.A)

Probable source	Predominant frequency	Dominant plane	Phase angle relationship	Amplitude	Nature of envelope	Comments
UNBALANCE						
(1) Mass imbalance	1 × rotor speed	Radial axial is higher on overhung rotors	1. Force-in-phase (0º) 2. Couple out-of-phase (90º)	Steady and proportional to ω^2	Narrowband	Dangerous at high rotational speeds. Rotor looseness or bow due to thermal stresses may change the amplitude and phase with time and are difficult to correct.
(2) Bent shaft	1 × rpm (2 × rpm if bent at the coupling)	Axial	180° out-of-phase axially	Steady	Narrowband	Run-out at rotor mass appears as unbalance. Run-out at coupling appears misalignment; difficult to correct.
(3) Eccentric motor rotor	1 × rpm, 1 × and 2× line frequency	Radial	N/A	Steady	Narrowband	Will fluctuate (beat) in amplitude and phase if an electrical problem also exists (see electrically induced source).
MISALIGNMENT						
(1) Parallel	1 × and 2 × rpm	Radial	radial- 180° out-of-phase	Steady	Narrowband	Most misalignments are a combination.
(2) Angular	1 × and 2 × rpm	Axial	axial- 180° out-of-phase			Errors are most common in the vertical plane. On long coupling spans, 1× is higher. Harmonics usually up to the third, but sometimes up to the sixth also present.
(3) Both parallel and angular	1× and 2× rpm	Radial and axial	both radial and axial will be 180° out-of phase			

(Contd)

TABLE 8.1 (*Contd*)

ELECTRICALLY INDUCED						
All electrically caused problems can be eliminated by cutting the current to the motor.	2 × slip frequency sidebands around 1 × rpm, 1 × and 2 × line frequency	Radial	N/A	Steady		Will fluctuate in amplitude and phase if mechanical problems (imbalance) also exist.
(1) Eccentric motor						
(2) Loose stator laminations	2 × line frequency and high frequency (1 kHz) sidebands of 2 × line frequency	Radial	N/A	High and steady	Narrowband sidebands of of 2× line frequency	Not usually destructive.
(3) Broken rotor bar	running speed with 2 × slip frequency sidebands	Radial	N/A	Steady	Narrow spike with sidebands	Replace rotor bar.
(4) Unbalanced coil or phase resistance	2 × line frequency	Radial	N/A	Low, steady	Narrowband	
(5) Stator problems (heating, shorts)	2 × line frequency	Radial	N/A	Steady	Narrowband	Vibration will increase as motor heats up.
(6) Loose iron	2× line frequency	Radial	N/A	High and steady	Narrowband	
DEFECTIVE BEARINGS						
(1) Anti-friction	Early stages – 30 Hz –1 kHz depending on size and speed. Late stages - high 1 × and multiple harmonics	Radial, except higher axial on thrust bearing	N/A	Increases as bearing degrades; may disappear just before failure.	Bandwidth broadens as bearing degrades. Baseline may increase across entire spectrum.	Salient vibration frequencies have been discussed in Section 8.2.2.10
(2) Sleeve	Early stages- subharmonics (may be	Radial	Shaft proximity probe orbits will indicate shaft	Increases as bearing degrades.	High baseline energies below 1 ×, 2 × and	Monitoring of rotor position (thrust) via proximity probes

(*Contd*)

TABLE 8.1 (*Contd*)

		Amplitude	Phase	Frequency	Remarks
		noticeable on shafts). Late stages-will appear as mechanical looseness (see below)	position and dynamic changes.	3 × rpm.	can provide reliable protection against thrust bearing failure.
MECHANICAL LOOSENESS					
(1) Bearings, pedestals, etc. (non-rotating)	Radial	1 ×, 2 × and 3 × predominant; may be up to 10 × at lower amplitude	Varies with type of looseness	Steady	May extend up to 10 times shaft speed.	The period of modulation will vary and the time signal will also be truncated (clipped). Also contains wideband noise. In the early stages, a strong 1 × response with some harmonics. As the looseness worsens, the amplitude of the harmonics will increase. Overall RMS value may also decrease. Further deterioration results in fractional harmonics: 1/2, 1/3, $1\frac{1}{2}$, $2\frac{1}{2}$ increasing in amplitude.
(2) Impellers, etc. (rotating)	Radial	1 × predominant, but may have harmonics up to 10 × at low levels.	Will vary from start-up to start-up.	Steady while running; will vary from start-up to start-up.	May extend upto 10 × shaft speed.	Variation of amplitude or phase may be caused by centre of gravity shifts.

TABLE **8.1**   (*Contd*)

OPERATION					
(1) (Process-related) Blade/vane pass	No. of blades/vane × rpm	Radial predominant in the direction of discharge piping.	N/A	Fluctuating	Broadband	More than one discharge volute will produce harmonics of blade passing frequency.
(2) Cavitation or starvation	random-broadband	Radial	N/A	Fluctuating	broadband up to 2000 + Hz	Vane pass frequencies may be superimposed
DRIVE BELTS					
(1) Mis-matched worn or stretched – also applies to adjustable sheave applications.	many multiples of belt frequency but 2 × belt frequency usually dominant	Radial, especially high in line with belts	N/A	May be unsteady and beating if a belt frequency is close to driver or driven speed.	N/A	Belt frequency = 3.14 cpm × pitch diameter/belt length
(2) Eccentric and/or unbalanced sheaves.	1 × shaft speed	Radial	In-phase	Steady	N/A	Balancing is possible with washers applied to taper lock bolts.
(3) Drive belt or sheave face misalignment.	1 × driver shaft	Axial	In-phase	Steady	N/A	Confirm with a straight edge.
(4) Drive belt resonance	Belt resonance with no relationship to rotational speeds	Radial	N/A	May be unsteady	±20 % of resonant frequency depending upon damping	Confirm with strobe light and belt excitation techniques. Change belt tension or belt length to eliminate the problem.
RESONANCE	requires forcing function to excite its natural frequencies	Axial or radial	A centre hung rotor resonance will display 180° out-of-phase bearing	Steady, but baseline energy fluctuations depend on	Appears broad at base; width depends upon damping	Frequency is independent of speed changes.

(*Contd*)

TABLE **8.1** (*Contd*)

			relationships. A component within a structure will display phase relationships dependent upon the bending mode excited.			force and damping.
INSTABILITY						
(1) Oil whirl	42 to 48% of running speed.	Radial	N/A	Steady	Discrete peaks	Considered excessive when amplitudes reach 50% of normal bearing clearances.
(2) Oil whip	Subrotational and equal to shaft resonance.	Radial	N/A	Steady	Discrete peaks	
(3) Rotor rub	50% of running speed and half harmonics	Radial	N/A	Steady	Discrete peaks	Rub increases the resonant frequency to the next highest fraction of running speed.
GEARS						
(1) Transmission error (poorly finished tooth face)	gear mesh frequency (gear rpm × no. of teeth) and harmonics.	Radial for spur gears, axial for helical or herringbone gears.	N/A	Depends on loading, speed and total transmission error.	Usually single peak, but some times low sidebands.	N/A
(2) Pitch line runout, mass unbalance misalignment or faulty tooth	1 × rpm and gear mesh frequency with ± gear rpm sidebands	Radial for spur gears; axial for helical or herringbone gears.	N/A	1 × rpm and gear mesh frequency sidebands depend on fault severity	Discrete peaks	May excite lateral or torsional resonances at various frequencies. Matching errors during hobbing can cause high 2 × or 3 × gear rpm vibration.

Note: Based on casing measurements and amplitude on a logarithmic scale.

8.12 ORDER TRACKING/RPM-BASED MEASUREMENTS

The order-tracking option in a spectrum analyser provides an important capability for the transient analysis of rotating machinery. Irrespective of whether a machine is designed to operate at a fixed speed or over a wide range of speeds, it is important for us to be able to relate the vibration characteristics of a machine to its rotational frequencies (order). The rpm-based analysis option does precisely this by accepting a periodic signal or pulse train from a tachometer or shaft encoder and applying it as a trigger input to the analyser. Thus the sampling of input signals is synchronized to the instantaneous angular position of the machine shaft. Whereas analysis in the frequency domain is done by digitizing the data using a constant number of samples per second, for analysis in the order domain, a constant number of samples per revolution are acquired. Once processed into the order domain, the data may be displayed as revolution histories or order spectra for which the user simply has to specify the machine speed range, the highest order of interest for the analysis and the order resolution required.

8.12.1 Spectrum Cascade/Waterfall Analysis (Amplitude versus Frequency versus Time)

Waterfall diagrams, also known as spectrum cascade plots are nothing but spectra displayed at consecutive instants of time. They are plotted in a three-dimensional manner with time (or machine speed) and frequency along two horizontal axes and vibration amplitude along the vertical axis. Another display format is the spectrogram in which time is plotted on the *X*-axis, frequency on the *Y*-axis and vibration amplitude or energy in colour or tones of grey. Change in spectral information with time or machine speed can be seen easily, making these plots a useful fault detection and trending tool. These displays are useful for transient analysis, identifying at what time instants or rpm, resonant conditions or critical speeds are being excited during a coast up or coast down of the machine. They are also

FIGURE 8.12 Waterfall plot

useful for evaluating the vibration amplitude and frequency characteristics during other transient events such as a transition in load, temperature or other operating variable. Most real-time analysers come with this waterfall option. Also generally available are Campbell diagrams and the option to view any constituent record or 'slice' across all records at a single frequency or instant of time/rpm. Waterfall diagrams are also employed extensively in mapping the time/frequency variations of sounds. Waterfalls can be created from instantaneous or averaged spectra and encompass single and dual-channel measurements. Setting the maximum frequency and the spectrum resolution are important for accurate diagnosis. Figure 8.12 shows a typical waterfall plot of a mass unbalance that excites oil whirl and oil whip, with the oil whip locking on to the critical speed resonance.

8.13   Bodé Plots

Trend monitoring of specific vibration indices or spectrum analysis do not throw light on how a machine behaves during transient conditions. Transient studies may be conducted by observing the response of a machine with respect to (i) machine speed or (ii) time or (iii) a combination of both. Bodé plots constitute a graphical representation of three quantities: signal amplitude, phase angle and machine speed or frequency. They comprise a set of two plots: amplitude versus frequency or rpm and phase versus frequency or rpm. These plots are extremely useful in transient studies involving changes in the response of a machine during a machine's process or load change. They are particularly useful for studying a machine's behaviour during its coast up or coast down when critical speeds or natural resonances are excited. They can also give indications regarding the amount of runout associated with a non-contact transducer, the balance condition of a machine, system damping and the operating phase angle and vibration amplitude at various machine speeds. The three measured variables are usually synchronized to the machine's operating speed using a key phasor signal, but occasionally they may be synchronized to some multiple of machine speed, depending upon the actual

Figure 8.13   Bodé plot

machine fault being investigated. They are very useful for identifying resonant frequencies by their characteristic peak amplitudes and corresponding 180° phase shift as shown in Fig. 8.13, the absence of a phase shift helping to rule our resonance.
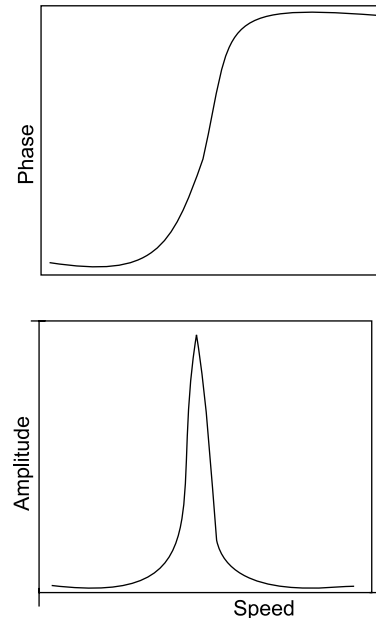
8.14   Polar Plot/Nyquist Plot

The information available in Bodé plots may also be plotted in an alternate manner, i.e. vibration amplitude versus phase, plotted on a polar coordinate graph called the Nyquist plot. This plot is also useful for identifying resonances. The only drawback is that rpm information is not explicitly available in these plots. The variables are plotted on a single circular chart instead of two separate plots. The centre of the plot represents zero speed and zero amplitude with subsequent amplitude and phase angle measurements plotted with their associated machine speeds. Phase angle measurements are plotted around the circumference of the chart, against machine rotation direction. Polar plots are always plotted after filtering in the component corresponding to machine running speed or some multiple of the machine speed, depending upon the fault being investigated. Machine critical speeds and natural resonances are displayed as loops, with the critical speed situated 90° from the start of the loop. This characteristic makes identification of resonances and critical speeds easy. Other information available from the plot are: slow speed runout, system damping, balance condition, operating phase angle and amplitude at various machine speeds. Figure 8.14 shows a typical polar plot of a machine tuned to the 1 × rpm component. Resonance is observed at 1600 rpm.
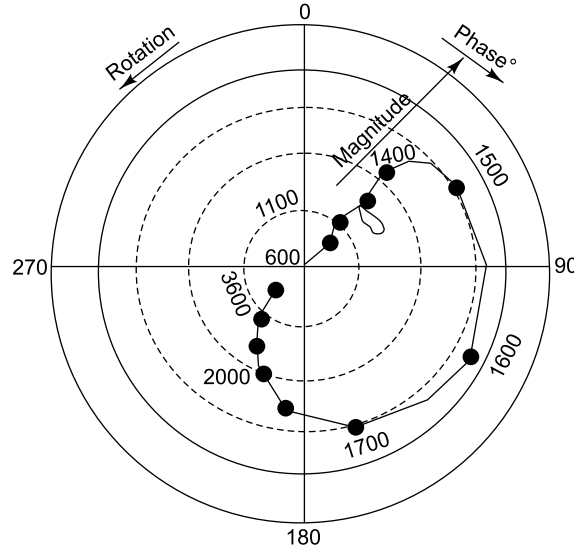
FIGURE 8.14 Polar plot

An alternate way of plotting the frequency response function is with real part on one axis and imaginary part on the other axis. Here again, rpm information is not explicitly available in the plots.

8.15 CEPSTRUM (QUEFRENCY DOMAIN) ANALYSIS

In relatively simple gearboxes, the time waveform or frequency spectra can be used to distinguish impacts due to cracked, chipped, or missing teeth, but when the gearbox contains several sets of mating gears, which is most often the case, the cepstrum is a powerful diagnostic tool. This is a specialized signal processing technique which is used for analysing signals from gearbox or bearing vibrations where modulation components (sidebands) lead to periodicities in the spectra in the form of families of harmonics and uniformly spaced sidebands. Eccentricity of one gearwheel, for instance, would result in AM of the basic vibration generated at the toothmeshing frequency (and its harmonics) by lower frequencies, typically the shaft rotational frequencies. At the same time, the variations in tooth contact pressure, which cause AM, give rise to rotational speed fluctuations, simultaneously producing frequency modulation at the same frequency. Thus, amplitude and frequency modulation often occur together, both giving rise to a family of sidebands with the same spacing (the fundamental modulating frequency); it is to be noted that it is this average sideband spacing which contains the basic diagnostic information regarding the source of modulation. Even for gearboxes in good condition, the spectra normally contain such sidebands, but at constant levels. Changes in the number and strength of sidebands would generally indicate deterioration in condition.

The term 'Cepstrum' is derived by rephrasing the word 'Spectrum'. Other terms involved in cepstrum analysis, like 'quefrency', 'rahmonics', etc. have been derived from the corresponding terms used in spectrum analysis like 'frequency', 'harmonics', etc. The cepstrum was originally defined as the power spectrum of the logarithm of the power spectrum (with logarithmic amplitude, but linear frequency scale), or mathematically as

$$c(\tau) = \left| \Im \left\{ \log \left[F_{xx}(f) \right] \right\} \right|^2 \tag{8.15}$$

where $F_{xx}(f)$ is the power spectrum of the time signal $f_x(t)$.

$$F_{xx}(f) = \left| \Im \left\{ f_x(t) \right\} \right|^2 \tag{8.16}$$

Here $\Im$ represents the forward Fourier transform of the quantity in brackets. The independent variable, τ, i.e. the quefrency has dimensions of time. Just as frequency spectra highlight periodicities in the time waveform, quefrency 'cepstra' highlight periodicities in frequency spectra. A high quefrency represents rapid fluctuations in the spectrum (small frequency spacings) and vice versa. The peaks in the cepstrum result from families of sidebands and the quefrency of the peak represents the periodic time of the modulation, with its reciprocal representing the modulation frequency. The quefrency gives information only about frequency spacings and does not say anything about absolute frequency. The cepstrum can be considered as a data reduction technique, effectively reducing a whole pattern of sidebands into a single line in the cepstrum; this is one of its main advantages. The logarithmic amplitude scale accentuates the periodic nature of the spectrum and reduces the influence of the somewhat random transmission path by which the signal travels from the source to the measurement location.

Over a period of time the cepstrum got redefined as the 'inverse transform of the logarithm of the power spectrum'.

$$c(\tau) = \Im^{-1} \{ \log F_{xx}(f) \} \tag{8.17}$$

This definition is advantageous in that it highlights the connection between the cepstrum and the autocorrelation function, which can be obtained as the inverse transform of the power spectrum i.e.

$$R_{xx}(\tau) = \Im^{-1} \{ F_{xx}(f) \} \tag{8.18}$$

It can be shown that the cepstrum is relatively insensitive to the effect of the signal transmission path. Let the power spectrum of a vibration signal at a measurement location be written as shown below:

$$F_{yy}(f) = F_{xx}(f) \cdot \left| H_{xy}(f) \right|^2 \tag{8.19}$$

where $F_{xx}(f)$ is the power spectrum of the source function and $H_{xy}(f)$ is the frequency response function of the transmission path. Taking logarithms on both sides,

$$\log F_{yy} = \log F_{xx} + 2 \log \left| H_{xy} \right| \tag{8.20}$$

The logarithm operator has transformed the multiplication to an addition which is maintained by the linear Fourier operator.

$$\Im^{-1} \left\{ \log F_{yy} \right\} = \Im^{-1} \left\{ \log F_{xx} \right\} + \Im^{-1} \left\{ 2 \log \left| H_{xy} \right| \right\} \tag{8.21}$$

Thus the source and transmission path effects are additive and since they often have quite different quefrency contents, they are separated in the cepstrum. Therefore, the location of an accelerometer does not change the cepstrum much, though it does the spectrum. Though cepstra usually contain several 'rahmonics' corresponding to each sideband spacing, it is the first or fundamental rahmonic which contains significant information on the average sideband height, while the others represent the distortion.

Figure 8.15(a) shows the cepstrum obtained from a bearing with outer race defect. A1, A2 and A3 are rahmonics of 0.00766 s (the quefrency) which corresponds to the modulating or defect frequency of

129 Hz (1/0.00766 Hz) at 1500 rpm for this bearing. Figure 8.15(b) shows the raw spectrum for the same bearing. From the spectrum it is seen that there are frequency components in the low frequency region (running speed and multiples up to around 120 Hz), structural resonance of the test rig (peak 150 Hz), component due to outer race resonance (3.538 kHz) and side bands corresponding to defect frequency about the resonant frequency. The defect frequency as such is not visible in the 0–500 Hz range making diagnosis from the spectrum difficult.
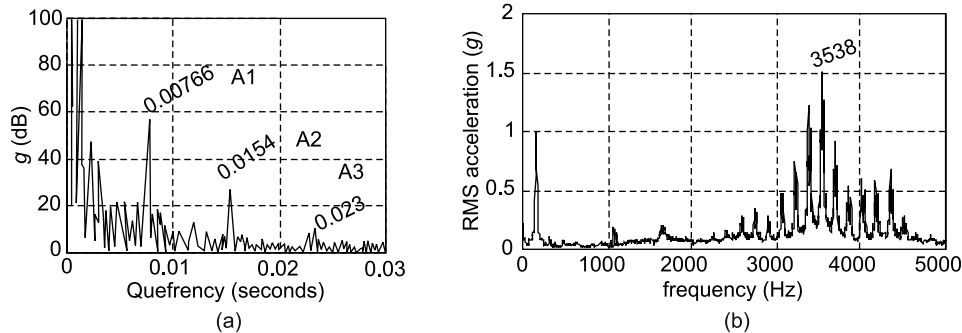
FIGURE 8.15   Vibration of bearing with outer race defect: (a) cepstrum, (b) raw spectrum

8.16   ENVELOPE ANALYSIS/DEMODULATION TECHNIQUE

The envelope analysis/demodulation option is available in some signal analysers and offers amplitude and frequency/phase demodulation. This measurement uses the Hilbert transform to construct the envelope signals, from which the AM and phase modulation (PM) signals are extracted. Demodulation can be combined with zoom to focus the analysis on the frequency range of interest; it can also be combined with synchronous averaging to increase the signal-to-noise ratio. Demodulation is useful when signals of interest are mixed in with other signals in the form of amplitude or frequency modulation and when signals are drowned in noise or harmonics of other signal components.

   This is a technique which is especially suitable for faults which create AM, as in bearings or gears. A vibration signal from a defective bearing comprises low frequency signals from rotational components, defect impulse signals (characterized by the defect frequencies: ball pass frequency outer race (BPFO), ball pass frequency- inner race (BPFI), ball spin frequency (BSF) and fundamental train frequency (FTF)) and machine noise which masks the early stages of bearing faults making spectrum analysis alone a difficult diagnostic tool. The first sign of rolling element bearing deterioration will be an increase in the amplitudes of the spectral components in the 5–20 kHz region. This is because each time a ball passes the fault, the resulting impact excites resonances in the structure. AMs are common in rolling element bearings because the vibration amplitudes vary when the defects on inner race or rolling elements rotate in and out of the bearing load zone. This AM results in side bands in the spectra occurring about the resonant frequencies of the components (inner race, outer race or roller). The latter thus act as the carrier frequencies, ω_r with the bearing defect frequencies and rotational frequencies serving as the modulating frequency.

   In the envelope analysis procedure, a band pass filter is designed taking into account the shaft speed and making an optimal choice between rejection of low frequency rotational signals, wide passband of

harmonic bearing defect components and rejection of unstable higher frequency noise. It is typically designed with a bandwidth of $\pm\ 2f_d$ around f_n (resonance frequency). Using the bandpass filter, all the low frequency high amplitude rotational components such as 1×, 2×, 3× are eliminated from the complex signal and the filter output signal contains only the resonance frequency along with the sidebands of the low amplitude bearing defect harmonics and rotational frequencies.

Removal of these higher amplitude rotational components from the signal results in a significant improvement of the signal-to-noise ratio, allowing the very small-amplitude defect frequency components (sidebands) to be detected. The filtered signal is demodulated to eliminate the carrier frequency. The demodulated signal is then passed through a low pass filter to eliminate the high frequency vibration. The power spectrum of this processed signal is finally obtained. Acceleration enveloping enables the bearing fault to be detected in the incipient stage itself by increasing the signal-to-noise ratio and by enhancing the significant spectral components relating to bearing faults. Figure 8.16 shows a block diagram of the enveloping procedure.
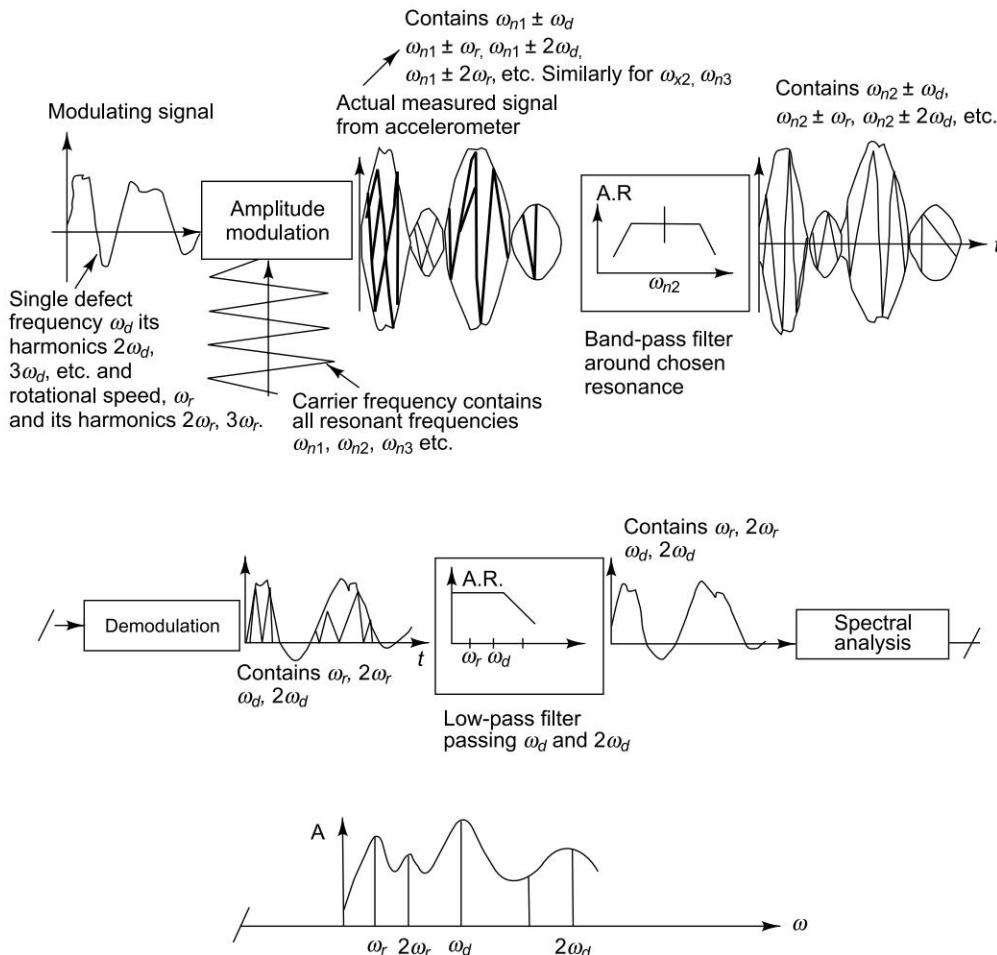
FIGURE 8.16 Block diagram of envelope analysis

Figure 8.17(a)–(d) shows the signals after each step of the enveloping process for the bearing with outer race defect, the results of which have been shown in Fig. 8.15, but now at 1000 rpm. Figure 8.17(a) shows the time domain signal and Fig. 8.17(b) the corresponding raw spectrum. Figure 8.17(c) shows the spectrum after bandpass filtering around the outer race resonance. Figure 8.17(d) shows the spectrum of the demodulated and low-pass filtered signal. A low pass filter with a cut off frequency of $2f_d$ has been used. The defect frequency and its multiples are clearly seen (92.17 Hz, 185.5 Hz) in the enveloped spectrum.
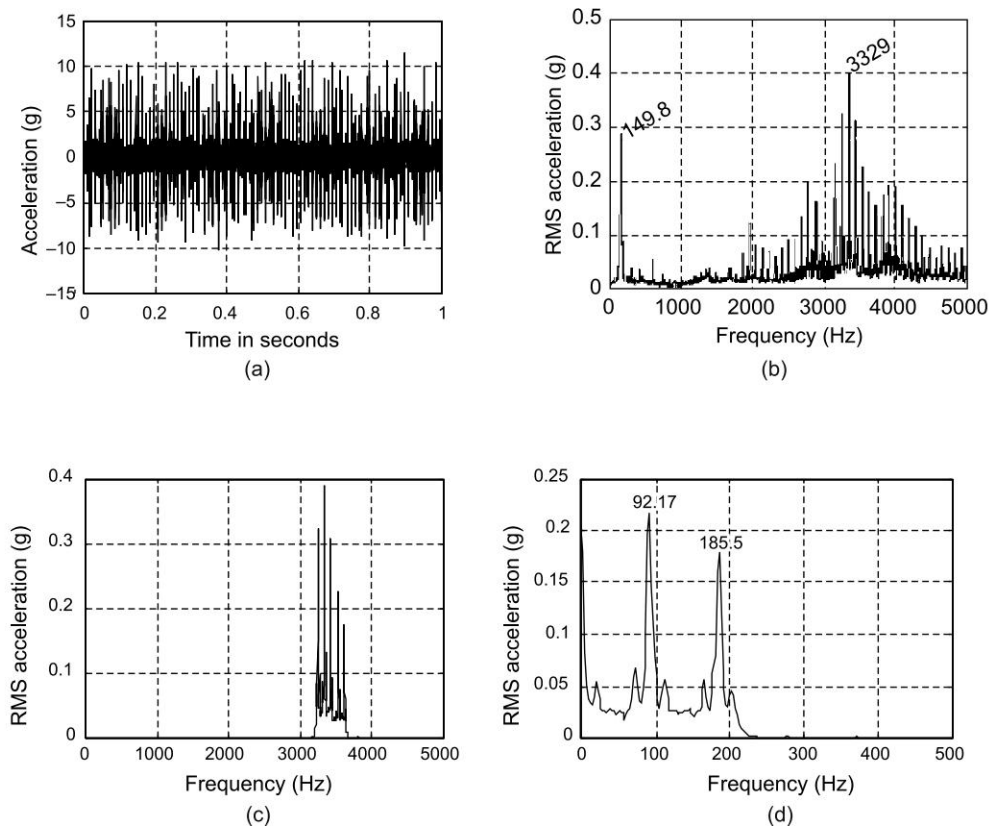
FIGURE 8.17   Results of enveloping process for bearing with outer race defect: (a) time domain signal, (b) raw spectrum, (c) spectrum of bandpass filtered signal, (d) spectrum of demodulated and low-pass filtered signal

8.17   PARAMETRIC SPECTRA USING AUTOREGRASSIVE (AR) AND AUTOREGRESSIVE MOVING AVERAGE (ARMA) MODELS

When only a short length of vibration signal is available, or if the vibration signal is of low energy as generated by bearings of very low speed machines, the FFT often yields poor results and time domain analysis also does not reveal much information. Parametric methods can yield higher resolutions than non-parametric methods for such cases and hence conventional time/frequency domain approaches to condition monitoring are replaced by adaptive methods using parametric models for diagnosis.

Autoregressive (AR) models are often fitted to stationary random time series. In this method, instead of trying to estimate the PSD directly from the data, the data are modeled as the output of a linear system driven by white noise and then the parameters of that linear system are estimated. The output of the filter for white noise input is got as an auto regressive (AR) process. The estimation of parameters for the AR model results in linear equations. An AR model is one where the present value of the signal generated by a process is expressed as a weighted sum of the past values plus a noise term. For the continuous-time case it is represented as

$$x(t) = a_0 x(t-1) + a_1 x(t-2) + \ldots + n(t) \qquad (8.22)$$

where $x(t)$ is the present data value, $x(t-1)$, $x(t-2)$, etc. are the previous data values; a_0, a_1, a_2,..., etc. are the model parameters and $n(t)$ is the noise term. Designing the model involves finding the optimum model order and calculating the model parameters. Along with estimating the model parameters, the modelling error is also calculated and the model order for which this is a minimum is considered the optimum order. A model of too high an order will result in spectral estimates containing spurious peaks, while spectral estimates with insufficient resolution are the result of models with too low an order. Hence the optimal model order may have to be found out by trial and error. Subsequently the power spectral density (PSD) of the signal vector $x(t)$ is determined. This parametric model is the simplest and has a computational advantage over other parametric modelling techniques such as the moving average (MA) and auto regressive moving average (ARMA) techniques, which require highly non-linear equations to be solved.

For many practical cases, the inclusion of both autoregressive and moving average terms results in a model that has fewer parameters than would be necessary for a satisfactory model of pure AR or pure MA form. The autoregressive moving average processes provide a powerful class of models for stationary random processes encountered in practice, because of their flexibility in accounting for a wide range of autocovariance functions. For discrete-time data, the general class of ARMA models consists of a discrete stationary linear transfer function applied to a sequence of white noise, the output being a zero-mean stationary process with frequency characteristics, which are dependent on the parameters of the transfer function. This is a model having a finite number of both autoregressive and moving average terms. Such a mixed process is referred to as autoregressive moving average process. The simplest autoregressive moving average model is the ARMA (1, 1) model. The autoregressive moving average process of order p, q, i.e., ARMA (p, q) is represented by the time series:

$$x_k - \phi_1 x_{k-1} - \ldots - \phi_p x_{k-p} = e_k - \theta_1 e_{k-1} - \ldots - \theta_q e_{k-q} \qquad (8.23)$$

In this equation $\{x_k(k\Delta t), k = 0, 1, 2, \ldots\}$ represents the discrete random process, Δt the sampling time interval, $\{e_k\}$ a zero mean Gaussian white noise of variance σ_e^2, and $(\phi_i, i = 1, 2, \ldots p)$ and $(\theta_i, i = 1, 2, \ldots, q)$ the autoregressive and moving average coefficients, respectively. As in AR modelling, once the autoregressive and moving average coefficients are got, the signal is defined and the PSD may be obtained. Figure 8.18 shows the PSD for the same bearing for which cepstrum and spectrum were shown in Fig. 8.15.
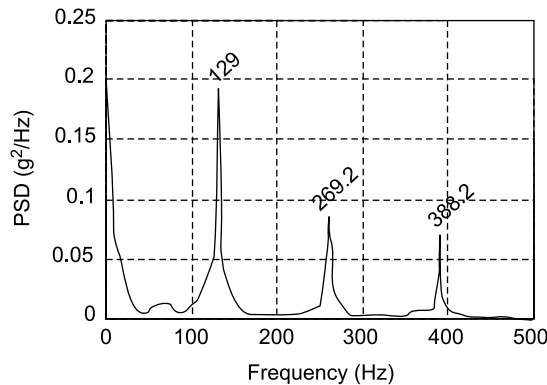
<p align="center">Figure 8.18 PSD from parametric model</p>

8.18 Advanced Fault Diagnostic Techniques

Fault diagnosis using the frequency domain technique is based on the fact that many machine defects generate vibration signals with distinctive frequencies. There are, however, many situations where machine defects cannot be correlated to specific frequencies or when the vibration signals are submerged in noise or vibrations from other machinery. In such cases, advanced diagnostic techniques incorporating new signal processing algorithms may be used. Short-time Fourier transform (STFTs) and wavelet analysis allow both time and frequency domain information to be displayed simultaneously on the same plot and are suitable for signals, the spectral characteristics of which change with respect to time. This enables analysis of short-duration transient events, as well as fault detection in machinery operating in non-steady-state conditions. For machines where the relationships between the faults and the resulting vibration signals are fuzzy, ANNs have been found to provide an excellent basis for detecting and classifying faults.

8.18.1 Time-Frequency Analysis: STFT

Conventional spectral analysis provides averaged spectral coefficients, which are independent of time. Stationary random signals, the frequency contents of which are time-invariant, do not require both the time and frequency information simultaneously. Non-stationarity in signals and systems lead to transient or impulsive events that cause frequency content to change abruptly or vary considerably and irregularly with time and it is important to know, where in time, the spectral components occur. So Fourier transform is not suitable for non-stationary signals and a transform giving the time-frequency representation of the signal is needed. Time-frequency analysis based on the STFT or spectrogram is used to display time-frequency variation and is becoming popular for machinery diagnostics, while it was earlier used essentially for speech analysis. It is accurate for slowly varying changes and it is this transform which is used in signal analysers for waterfall plots. For this application, overlapping short lengths of vibration signals are selected and a Fourier transform is made on each one and displayed as a 3-D or waterfall plot, revealing how the spectrum changes with time.

In the STFT, a window slides over the time record and to define the spectral behaviour of the signal at a particular time, a small section of the signal centred about the window of interest is extracted and Fourier transformed to compute the spectral content of the signal in that window. A local Fourier

transform is thus computed for every portion of the signal within each window which is assumed to be stationary. At each fixed time of interest t, the time domain signal is multiplied by a moving window, which emphasizes the signal centred at time t, and the resultant windowed signal is Fourier transformed to give the STFT. The general definition of STFT is

$$F(\tau,\omega) = \int_{-\infty}^{\infty} f(t)\psi^*(t-\tau)e^{-i\omega t}\,dt \qquad (8.24)$$

where $f(t)$ is the original function in the time domain and $\psi(t)$ is the window function. The spectrogram is the magnitude squared of the STFT and is given as

$$\rho_s(t,f) = |F(\tau,\omega)|^2 \qquad (8.25)$$

When the window $\psi(t)$ is a Gaussian function, the STFT is called a Gabor transform. The fixed time window $\psi(t)$ is the limitation of STFT as it causes a fixed time-frequency resolution i.e. when the length of the window is specified, the time and frequency resolutions are unchanged as illustrated in Figs. 8.19(a) and (b).
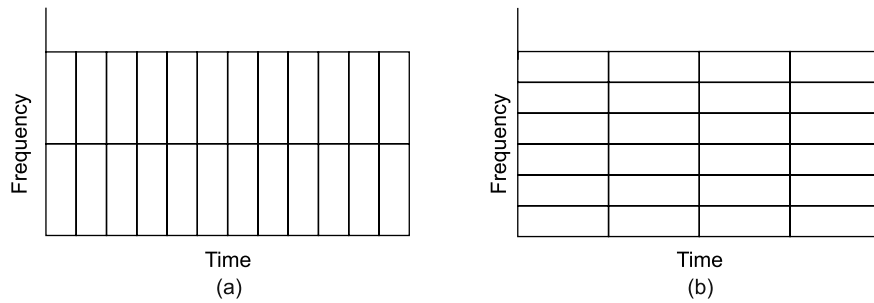
FIGURE 8.19   Short time Fourier transform–comparison of time and frequency resolution: (a) small window, (b) long window

The drawback of STFT may be explained by the Heisenberg uncertainty principle (it states that one cannot assign exact simultaneous values to the position and momentum of a physical system; in the present case, position and momentum are replaced by time and frequency) for the transform pair $\psi(t) \leftrightarrow \Psi(\omega)$

$$\Delta t \Delta \omega \geq \frac{1}{2} \qquad (8.26)$$

where $\Delta\omega$ and Δt are the bandwidth and time spread (i.e. two pulses in time can be discriminated only if they are more than Δt apart) of $\psi(t)$, respectively, and

$$(\Delta t)^2 = \frac{\int t^2 |\psi(t)|^2\,dt}{\int |\psi(t)|^2\,dt} \qquad (8.27)$$

$$(\Delta \omega)^2 = \frac{\int \omega^2 |\psi(\omega)|^2\,d\omega}{\int |\psi(\omega)|^2\,d\omega} \qquad (8.28)$$

As can be seen, the limitation of STFT is that its resolution is not adaptive. In other words the signals are analysed with equal time intervals over the whole range of frequencies. Vibration signals in the presence and growth of machinery faults need to be represented with fine resolution. To solve this problem, there is a need to use narrow windows for higher frequencies and wide windows for lower frequencies.

8.19 Wavelet Analysis

Wavelet analysis is of interest for the analysis of non-stationary signals, because it provides an alternative to the classical STFT. The basic difference between short-time Fourier transform (STFT) and wavelet transform (WT) is that STFT uses a single analysis window, whereas WT uses short time windows at higher frequencies and long windows at low frequencies or in other words, the WT zooms in on short lived, high frequency phenomena and zooms out of long lived, low frequency phenomena. A contraction in one domain is accompanied by an expansion in the other, but in a non-uniform way over the time-frequency plane. Thus WT has the property of being localized in time (space) as well as scale (frequency).

The WT was first proposed as a tool for signal analysis by geophysicist Morlet. WT involves representing general functions in terms of simple, fixed building blocks with a fixed shape called 'wavelets' at different positions (translated in time) and scales (dilated in time to give different frequency components) and then studying each component with a resolution matched to its frequency. This is called 'constant-Q' or constant relative bandwidth frequency analysis. It is akin to analysis using $1/n$ octave filters. It provides a time-scale map of a signal, enabling the extraction of time-varying features, and making wavelets an ideal tool for analysing signals of a transient or non-stationary nature such as signals containing discontinuities and sharp spikes. Wavelets give a better time localization at high frequencies than STFT and for that reason can be useful for detecting local events in a signal. It can conveniently be used to analyse gear vibration which manifests as amplitude and PM of tooth meshing vibration.

The 'analysing' wavelet is so called because: (i) wavelet means a small wave, of finite length and (ii) wave refers to the oscillatory nature of this function. The basis function $\psi(a, \tau)$ or the mother wavelet used in the transformation process, is so called because other functions are derived from it; it consists of a family of short-duration, high frequency and long-duration, low frequency functions. Wavelet analysis measures the time-frequency content in a signal indexed by the translation (τ) and scale parameters (a). The translation operation involves shifting of the mother wavelet along the time axis; each wavelet is local in the sense that it decays to zero when sufficiently away from its centre. Then the translated version is used to capture the time information of the function to be analysed. Depending on the dilation parameter a, the wavelet function dilates or contracts in time, causing the corresponding contraction or dilation in the frequency domain. At the finest scale, wavelets may be very short and at a coarse scale they may be very long. Any particular local feature of a signal can be identified from the scale and position of the wavelet into which it is decomposed.

8.19.1 Continuous Wavelet Transform

The basis function $\psi(a, \tau)$ may be defined as

$$\psi(a,\tau) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-\tau}{a}\right), \quad (a > 0) \tag{8.29}$$

The dilation factor a represents log frequency as for constant percentage bandwidth filters. When a is large ($a > 1$), the basis function becomes a stretched version of the mother wavelet ($a = 1$) and shows low

frequency characteristics. When *a* is small ($a < 1$), the basis function is a contracted version of the mother wavelet function and displays high frequency characteristics. Thus, we see that the CWT analyses the low frequency content of a signal with a wide duration function and, conversely, analyses high frequency phenomena with a short duration function overcoming the deficiencies of Fourier analysis. Figure 8.20 shows translated and scaled versions of a mother wavelet.
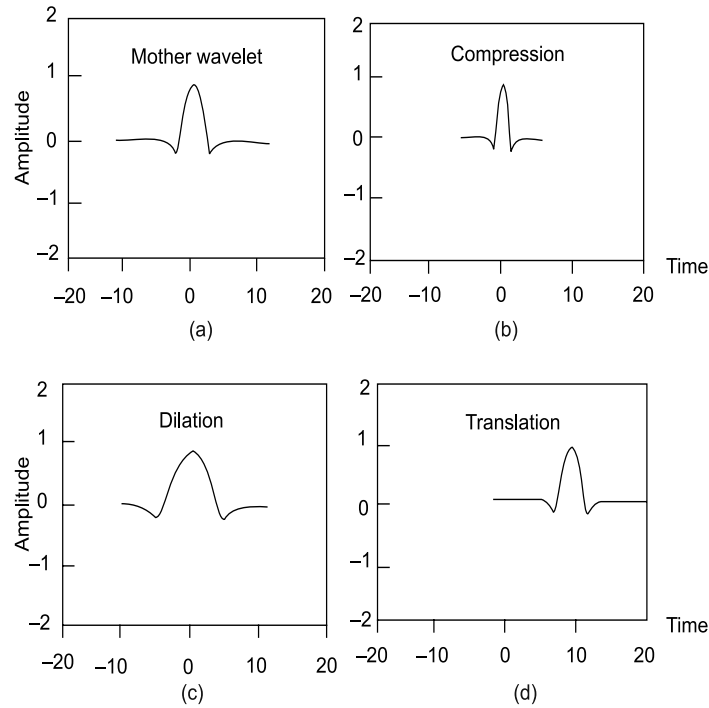
FIGURE 8.20   Examples of scaling, translation and dilation of wavelets: (a) $a = 1$, $b = 0$ (b) $a = 0.25$, $b = 0$, (c) $a = 5$, $b = 0$ (d) $a = 1$, $b = 10$

The formula for the WT is

$$W(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \psi^* \left(\frac{t - \tau}{a} \right) dt \qquad (8.30)$$

where $\psi^*(t)$ is the complex conjugate of the analysing wavelet $\psi(t)$. Since this formula represents a convolution, the wavelets can be considered as a set of impulse responses of filters, which because of the dilation factor have constant percentage bandwidth properties. The real or complex continuous-time function $\psi(a, t)$ is a mother wavelet if it satisfies the two properties listed below:

(i) The function integrates to zero

$$\int_{-\infty}^{\infty} \psi(t) dt = 0 \qquad (8.31)$$

(ii) It is square integrable or, equivalently, has finite energy.

$$\int\limits_{-\infty}^{\infty} |\psi(t)|^2 \, dt < \infty \tag{8.32}$$

There are many types of wavelets which may be chosen depending on the input signal. To name a few, they are Daubechies, Morlet, Symlet, Biorthogonal, Coiflets, Mexican Hat, Haar, etc. The Morlet wavelet is one of the most interesting ones used for machinery diagnostics. A remarkable feature of this wavelet is that its Fourier spectrum is Gaussian. The names of the Daubechies family wavelets are written as dbN, where N is the order, and db the 'surname' of the wavelet family. The wavelet functions ψ of the members of the Daubechies family from order 2–6 are shown in Fig. 8.21.
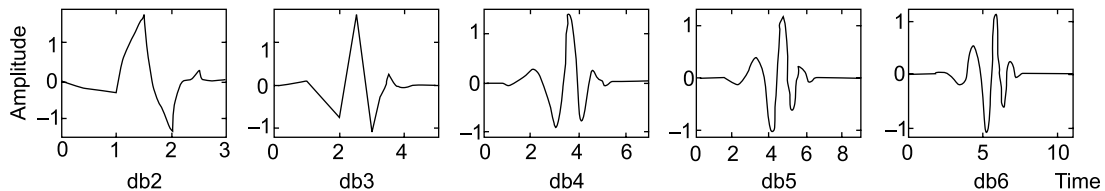
FIGURE 8.21   Wavelet functions ψ of Daubechies wavelet

Magnitude plots of WT with Morlet wavelets are shown in Fig. 8.23 for a healthy gear as well as one with part of a tooth removed. The energy of the vibration signal is uniformly distributed around the fundamental meshing frequency at 320 Hz as shown in Fig. 8.22(a) for the healthy case. In the presence of a fault the signal energy changes around the gear mesh frequency at the fault location (180°) as seen in Fig. 8.22(b), spreading to the second harmonic of gear meshing frequency, i.e. 640 Hz.
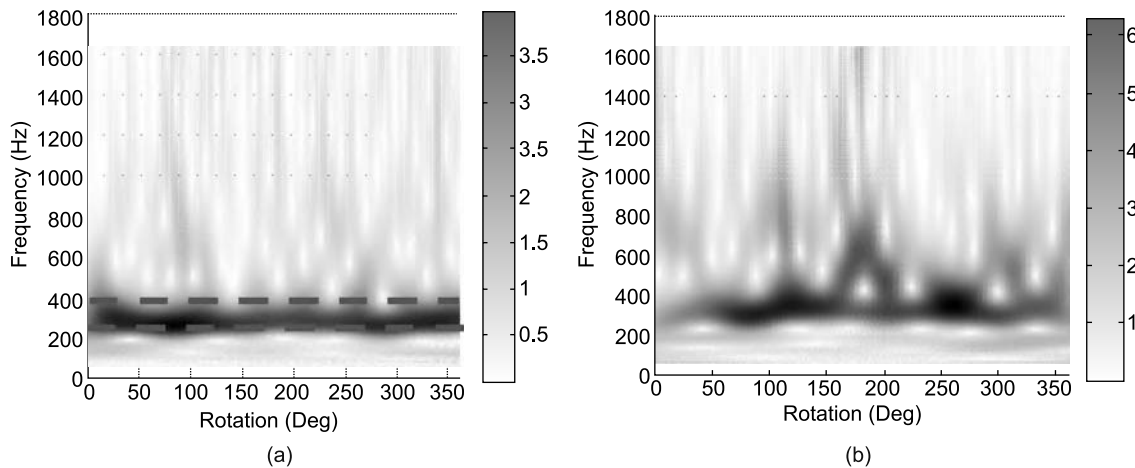
FIGURE 8.22   Wavelet magnitude plots for a gearbox: (a) healthy, (b) defective

8.19.2   Discrete Wavelet Transform

As mentioned before, the CWT maps a one-dimensional signal to a two-dimensional time-scale joint representation. However, it has certain drawbacks: (i) Calculating wavelet coefficients at all possible

scales amounts to a lot of work and generates a lot of redundant data. (ii) Even without the redundancy, there are still an infinite number of wavelets in the WT; it is desirable to have this number reduced to a more manageable count and (iii) For most functions the WTs have no analytical solutions and they can be calculated only numerically. Hence fast algorithms are needed to be able make WTs computable and these compute using discrete wavelets. Discrete wavelets do the scaling and translation operations in discrete steps, limiting the choice of wavelet scales and translations a and τ to discrete numbers, but the analysis is still sufficiently accurate. In other words, this transform decomposes the signal into a mutually orthogonal set of wavelets. This is achieved by modifying the wavelet representation (8.29) to create (8.33).

$$\psi_{j,k}(t) = \frac{1}{\sqrt{s_0^j}} \psi\left(\frac{t - k\tau_0 s_o^j}{s_o^j}\right) \tag{8.33}$$

In Eq. (8.33), j and k are integers. $s_0 > 1$ implies a fixed dilation step and it is generally chosen as 2 so that the sampling of the frequency axis amounts to dyadic sampling. This is a natural choice for computers and the human ear. The translation factor τ_0 depends on the dilation step and is typically chosen as 1 to result in dyadic sampling of the time axis also. The time-scale space is sampled at discrete intervals as a result of discretizing the wavelet. It results in a series of wavelet coefficients, and the process is called wavelet series decomposition. It is important that the signal be reconstructable from these coefficients and this is possible by making the discrete wavelets orthogonal to their own dilations and translations by special choices of the mother wavelet, which means

$$\int \psi_{j,k}(t)\psi_{m,n}^*(t)dt = 1, \ j = m \ \text{ and } k = n$$
$$= 0, \text{ otherwise} \tag{8.34}$$

An arbitrary signal can be reconstructed by summing the orthogonal wavelet basis functions, weighted by the WT coefficients.

8.20 Fault Classification Using ANN

ANNs are non-parametric classification techniques. They are based on the behaviour of the brain as a network of units called neurons and have been found to be an effective tool for pattern recognition in situations where data are fuzzy or incomplete and in tasks that previously relied on human judgement to take decisions. ANNs are highly suitable for condition monitoring in machinery where the number of components and processes are too many or complex to be mathematically modelled properly. Mimicking the human brain, ANNs learn from experience, generalized from previous examples to new ones and extract essential characteristics from the inputs containing randomness and irrelevant data. The neuron is the information processing unit that is fundamental to the operation of a neural network and is shown in Fig. 8.23. Each input is multiplied by a corresponding weight, analogous to a synaptic strength, and all of the weighted inputs are summed to determine the activation level of the neuron.
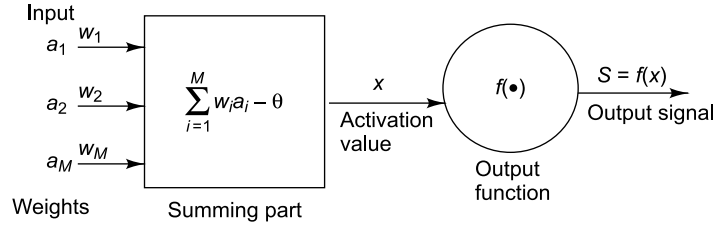
FIGURE 8.23   Typical neuron model

In the model, the activation value (x) is obtained as the weighted sum of its M input values (a_i) and a bias term (θ). The output signal (S) is typically a non-linear function $f(x)$ of the activation value. The following equations describe the typical neuron model.

Activation:

$$x = \sum_{i=1}^{M} w_i a_i - \theta \tag{8.35}$$

Output signal:

$$S = f(x) \tag{8.36}$$

A common differentiable output function used in the learning process is one which possesses a sigmoid non-linearity. Two examples of non-linear functions are the logistic function and hyperbolic tangent function described in the equations below and shown in Figs. 8.24(a) and (b).

Logistic function:

$$f(x) = \frac{1}{1 + e^{-x}}, \quad -\infty < x < \infty \tag{8.37}$$

Hyperbolic tangent function:

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}, \quad -\infty < x < \infty \tag{8.38}$$
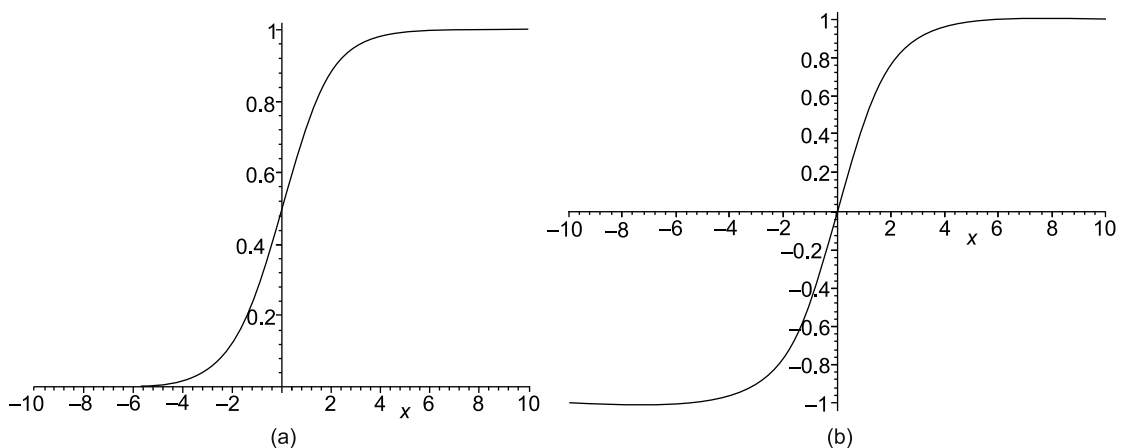
FIGURE 8.24   Typical output functions $f(x)$: (a) logistic function, (b) hyperbolic tangent function

For the logistic function the limit is $0 \leq f(x) \leq 1$ and for the hyperbolic tangent function, the limit is $-1 \leq f(x) \leq +1$. A typical feedforward neural network is discussed in the next section.

The error back propagation (EBP) training algorithm is a very popular scheme which performs an input to output mapping by minimizing a cost function using a gradient search technique. The cost function is equal to the mean squared difference between the desired and actual network outputs. There are two stages in the EBP sequence, namely a forward pass and a backward pass. During the forward pass, all the weights of the network are initialized randomly and the network outputs and the differences between the actual and target outputs, i.e. errors are calculated for the initialized weights. During the backward pass, the initialized weights are adjusted to minimize the error by propagating the error backwards. The network outputs and errors are calculated again with the updated weights and the process is repeated till the error is acceptably small.

8.20.1   Multilayer Feedforward Neural Network Model

A three layered feedforward neural network is as shown in Fig. 8.25. The network consists of three layers:

(i)   The input layers that receive preprocessed data from machines/structures
(ii) The hidden layer (or layers) which processes the data
(iii) The output layer that provides the result of the analysis, i.e. whether the machine or structure is healthy or damaged
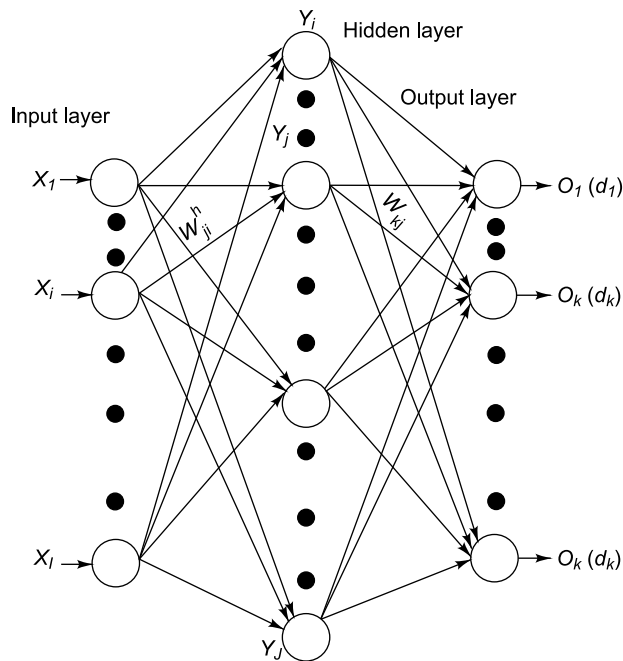
FIGURE 8.25   Three layered feed forward neural network

The input layer has *I* linear input units indexed by **i,** the hidden layer has *J* non-linear units indexed by *j* and the output layer has *K* non-linear units indexed by *k.* There may be a number of hidden layers. Since the input vector {*X*} is given at the input layer and the desired output {*d* } is available only at the output layer, the error between the desired output vector {*d*} and actual output vector {*O*} is available only at the output layer. Using this error, it is necessary to adjust the weights (W_{ji}^h) from the input units to the hidden units, and the weights (W_{kj}) from the hidden units to the output units.

It is left to the vibration engineer to decide what input and output parameters are to be used. In the case of bearing vibration diagnostics, the input vector might comprise the peak, RMS value, skewness, kurtosis or average spectral heights of selective bands. The output vector may comprise the following conditions: good bearing, defective inner race, defective outer race or ball defect.

8.21 CLASSIFICATION USING SUPPORT VECTOR MACHINES (SVM)

As seen earlier, classifying is a requirement in learning the condition of a machine. The SVM is a powerful pattern classifier pioneered by Vapnik. It comprises a set of supervised learning methods used for classification and regression, which can learn even from a small training data set. Support vector machine models are a close cousin to classical multilayer perceptron neural networks. In fact, an SVM model using a sigmoid kernel function is equivalent to a two-layer, perceptron neural network. Given two classes of data points, one belonging to a data set, say, from a healthy machine and the other from a faulty machine, the aim of a fault classifier would be to decide which class a new data point would belong to and this is what the SVM does. In this technique, each (raw) data point (attribute) in a class (input vector) is mapped into an *n*-dimensional feature space, i.e. it is transformed to an *n*-dimensional vector (feature). The task of choosing the most suitable representation is known as feature selection. Once the features are selected, the aim of the SVM is to see whether such transformed points (features) can be separated with an *n* – 1 dimensional hyperplane or decision surface. Typically there are many hyperplanes that might classify the data. However, the objective is to achieve maximum separation between the two classes or in other words, a hyperplane is to be constructed in such a way that the separation between the two classes of data is maximized. Such a hyperplane is known as the maximum-margin hyperplane and such a classifier is known as a maximum margin classifier. The separation between the hyperplane and the closest data point is called the margin of separation and is denoted by ρ. The goal of a support vector machine is to find a particular hyperplane for which the margin of separation ρ is maximized. Under this condition, the decision surface is referred to as optimal hyperplane. The vectors on the margins near the hyperplane are called support vectors. Figure 8.26(a) shows the working of an SVM algorithm and Fig. 8.26(b) shows the classification obtained from an SVM of a set of points from two different classes of data, class A (asterisk), say from a good machine and class B (circles), say, from a faulty machine. It also shows the support vectors and optimal hyperplane.

A multiclass SVM aims to classify attributes by using support vector machines, where the labels of different classes are drawn from a finite set. The main way of doing this is by reducing the single multiclass problem into multiple binary problems. Thus each problem yields a binary classifier, which is assumed to produce an output function that gives relatively large values for attributes from one class and relatively small values for attributes from the other.
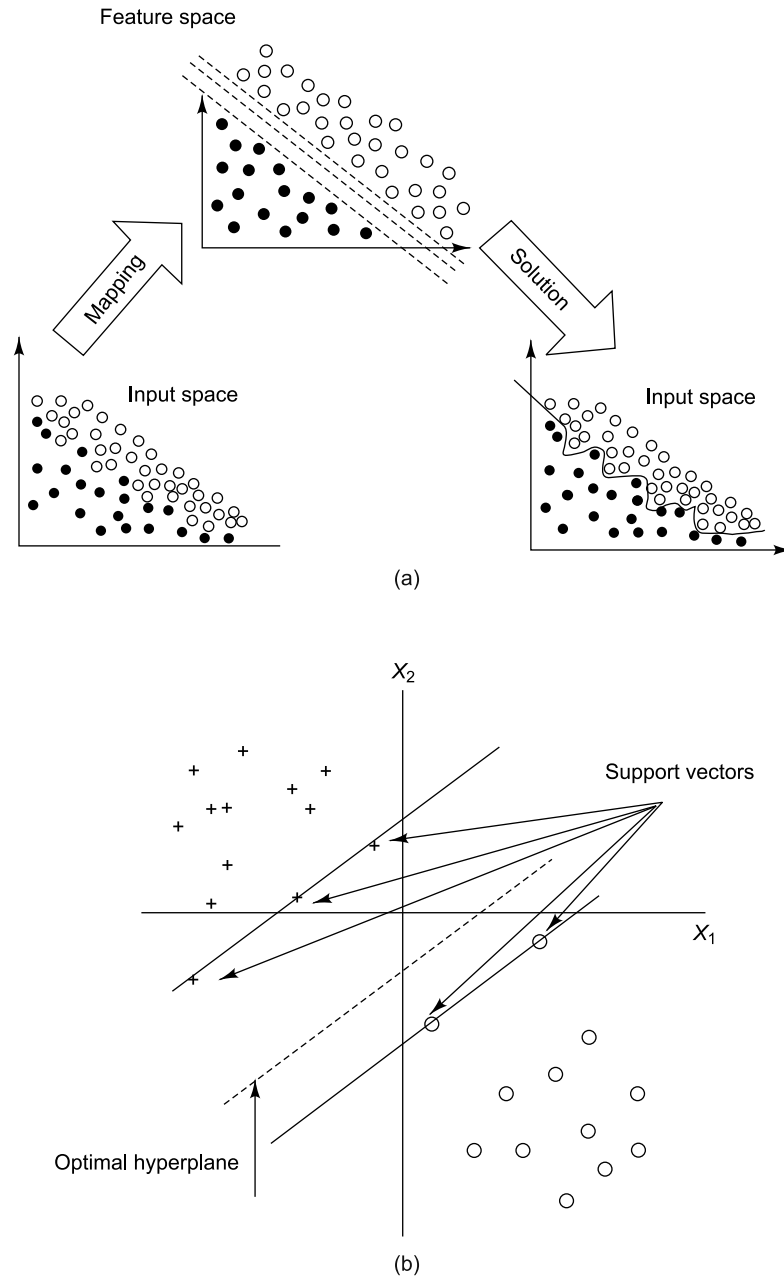
FIGURE 8.26   Support vector machine: (a) SVM procedure, (b) support vectors and optimal hyperplane for linearly separable patterns

Non-linear mapping of an input pattern vector into a higher dimensional feature space is performed in accordance with Cover's theorem on separability patterns. For an input pattern space made up of non-linearly separable patterns, Cover's theorem states that such a multi-dimensional space may be transformed into a new feature space where the patterns are linearly separable with a high probability,

provided the transformation is non-linear, and the dimension of the feature space is high enough. The separating hyperplane is defined as a linear function of the vectors drawn from the feature space. The optimal hyperplane is defined as

$$\sum_{i=1}^{N_L} \alpha_i d_i K(X, X_i) = 0 \tag{8.39}$$

where X_i is a support vector, X is an m-dimensional vector drawn from the input space, $\{\alpha_i\}_{i=1}^{N_L}$ are the Langrange multipliers which extend the first-order condition (derivative or gradient equal to zero) for optimization to the case of equality constraints. $\{d_i\}_{i=1}^{N_L}$ are the desired responses (target outputs), $d_i \in \{1, -1\}$ and N_L is the number of support vectors for which the Lagrange multipliers are non-zero. $K(X, X_i)$ is the inner product kernel defined by

$$K(X, X_i) = \phi^T(X)\phi(X_i)$$
$$= \sum_{j=1}^{m} \phi_j(X)\phi_j(X_i), \, i = 1, 2, \ldots, N_L \tag{8.40}$$

Here $\{\phi_j(X)\}_{j=1}^{m}$ denotes a set of non-linear transformations from the input space to the m-dimensional feature space. From Eq. (8.39) it is seen that the construction of the optimal hyperplane is based on the evaluation of an inner-product kernel $K(X, X_i)$ which is used to construct the optimal hyperplane in the feature space without having to consider the feature space itself explicitly. Since the design of an SVM involves finding an optimum hyperplane, it is necessary to find the optimal Lagrange multipliers which are obtained from the given training samples.

There is a modified maximum margin idea that allows for mislabeled examples. If there exists no hyperplane that can split the two classes, a soft margin method can be used which will choose a hyperplane that classifies the examples as correctly as possible, while still maximizing the distance to the nearest correctly classified examples and the optimization becomes a trade-off between a large margin and a small error penalty. The parameters of the maximum-margin hyperplane are derived by solving an optimization problem which simultaneously minimizes the empirical classification error and maximizes the geometric margin.

FURTHER READINGS

1. Barkova, A.V and Barkora, N.A., The artificial intelligence systems for machine condition monitoring and diagnostics by vibration, In *Proceedings of the Saint Petersburg Post-graduate Institute of the Russian Federation Power Industry and Vibration Institute, USA*, Vol. 9, Saint Petersburg, 1999.
2. Braun, S. *Mechanical Signature Analysis: Theory and Applications*, Academic Press, London, 1986.
3. Burrus, C.S., Gopinath, R.A. and Guo, H., *Introduction to Wavelets and Wavelet Transforms: A Primer*, Prentice Hall, Upper Saddle River, NJ (USA), 1998.
4. Collacott, R.A, *Vibration Monitoring and Diagnosis*, George Godwin Ltd., London, 1979.
5. Cortes, C. and Vapnik, V., *Support-Vector Networks*, Machine Learning, 20, 273-298, 1995.
6. Daubechies, I., Ten lectures on wavelets, In *Proceedings of the CBMS-NSF Regional Conference Series in Applied Mathematics*, Philadelphia, SIAM, Vol. 61, 1992.

7. Eisenmann, R.C. Sr. and Eisenmann, R.C. Jr., *Machinery Malfunction: Diagnosis and Correction*, Prentice Hall, New York, 1998.
8. Goldman, S., *Vibration Spectrum Analysis*, Industrial Press, New York, 1999.
9. Haykin, S., *Neural Networks and Learning Machines*, Prentice Hall, Upper Saddle River, New Jersey, 2008.
10. Joachims, T., *Learning to Classify Text Using Support Vector Machines: Methods, Theory, and Algorithms*, Kluwer Academic Publishers, Norwell, MA, 2002.
11. Kaiser, G., *A Friendly Guide to Wavelets*, Birkhäuser, Boston, 1994.
12. Kecman, V., *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*, MIT Press, Cambridge, 2001.
13. Körner, T. W., *Fourier Analysis*, Cambridge University Press, Cambridge, UK, 1996.
14. Lalanne, C., *Mechanical Vibration & Shock*, Taylor and Francis, North America, 2002.
15. Lyon, R.H., *Machinery Noise and Diagnostics*, Butterworth, Boston, 1987.
16. Mathew, J. and Alfredson, R.J, The condition monitoring of rolling element bearings using vibration analysis, *Trans. ASME J. Vib. Acoust. Stress Reliab. Des.*, 106, 447–457, 1984.
17. McFadden, P.D. and Smith, J.D., A signal processing technique for detecting local defects in a gear from the average of the vibration, *Proc. IMechE*, 199, 99–112, 1985.
18. Mechefske, C. K. and Mathew. J., Parametric spectral estimation to detect and diagnose faults in low speed rolling element bearings: Preliminary investigations, *Mech. Syst. Signal Proc.*, 7(1), 1–12, 1993.
19. Mechefske, C.K. and Mathew, J., Fault detection in low speed rolling element bearings, Part I: the use of parametric spectra, *Mech. Syst. Signal Proc.*, 6, 297–308, 1992.
20. Mitchell, J.S., *An Introduction to Machinery Analysis and Monitoring*, Penwell, Los Angeles, 1981.
21. Mitchell, L.D, Signal processing with the Fast Fourier Transform (FFT) analyzer, In *Proceedings of Fall Meeting—Recent Advances in Experimental Characterization of Composites*, Salt Lake City, Utah, Nov. 6–10, pp. 22–32, 1983.
22. Piersol. A.G., *Harris' Shock and Vibration Handbook*, McGraw-Hill Professional, New York, 2002.
23. Randall, R.B., *Frequency Analysis*, Brüel & Kjær, Copenhagen, 1987.
24. Rao, B.K.N., *The Handbook of Condition Monitoring*, Elsevier, Oxford, 1996.
25. Reeves, C.W., *The Vibration Monitoring Handbook*, Coxmoor, Oxford, 1999.
26. Smith, J.D., *Gears and their Vibrations, Marcel Dekker*, New York, 1983.
27. South T., *Managing Noise and Vibration at Work: A Practical Guide to Assessment, Measurement and Control*, Elsevier Butterworth-Heinemann, Oxford, 2004.
28. Taylor, J.I., *The Vibration Analysis Handbook*, Vibration Consultants, Tampa, 1994.
29. Taylor, J.S. and Cristianini, N., *Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, New York, 2000.
30. Vapnik, V. and Kotz. S., *Estimation of Dependences Based on Empirical Data*, Springer-Verlag, New York, 2006.
31. Williams, J. H., Davies, A. and Drake P. R., *Condition-based Maintenance and Machine Diagnostics*, Springer, Technology & Engineering, Berlin, 1994.
32. Wowk, V., *Machinery Vibration: Measurements and Analysis*, McGraw-Hill, New York, 1991.

<div style="text-align: right">

Chapter

9

</div>

Vibration Severity and Standards

9.1 INTRODUCTION

In this chapter various vibration standards which serve as guidelines for prevention of vibration severity are discussed. These may be classified from the point of view of: (i) machinery vibration and (ii) vibration of the human body. Both whole body vibration (WBV) and hand-arm vibration are discussed in connection with the latter. Assessment of ride comfort in road and rail vehicles is described. Finally the effect of vibration in buildings on human beings is spelt out.

9.2 CLASSIFICATION OF SEVERITY OF MACHINERY VIBRATION

In the classification of severity of machinery vibration, the allowable displacement, velocity or acceleration depends on the type of standard, the frequency range, the direction in which measurement is made, the size, horse power and support condition of the machine and other factors.

9.2.1 ISO Standard 2372 (10816)

The International Standards Organization has a special measure, 'vibration severity' which is defined as the highest value of the broadband, root mean square (RMS) value of the velocity amplitude (not spectral levels) in the frequency range 10–1000 Hz, as evaluated on a machine at prescribed points. ISO standard 2372 is applicable to rotating machinery having rigid rotors and to those machines having flexible rotors and in which the bearing housing vibration is a measure of the shaft vibration. Severity is considered at frequencies between 30% of the rotational frequency to at least three times the rotational frequency evaluated on the structure. This standard is applicable to machines running at speeds in the range 600–12,000 rpm. It includes the most common causes of vibration in rotating machinery resulting from (i) non-synchronous excitation such as rubs and rotor whirl, (ii) rotor unbalance, (iii) electric field excitation and harmonics and (iv) harmonics of synchronous rotor excitation. Because of the limited high-frequency range, rolling element bearing problems can be easily missed. Machinery can be subdivided into four basic categories for purposes of vibration measurement, choice of transducer location and evaluation.

(i) Reciprocating machinery having both rotating and reciprocating components, such as diesel engines and certain types of compressors and pumps; vibration is measured on the main structure of the machine at low frequencies.

(ii) Rotating machinery having rigid rotors, such as certain types of electric motors, single stage pumps and slow-speed pumps; vibration is measured on the main structure (bearing housing or pedestals).

(iii) Rotating machinery having flexible rotors, such as large steam turbine generators, multistage pumps and compressors; vibration is measured on the shaft directly.

(iv) Rotating machinery having quasi-rigid rotors, such as low-pressure steam turbines, axial flow compressors and fans; vibration amplitude is measured on the bearing housing.

The allowable vibration severity for various classes of machines can be seen in Table 9.1. The values given are RMS values. To get peak values, they should be multiplied by $\sqrt{2}$. In the table, the various classes of rotating machinery considered are as described below:

Class I: Individual components integrally connected with the complete machine or small machines in their normal operating conditions (i.e. electric motors up to 15 kW).

Class II: Medium sized machines (i.e. 15–75 kW electric motors and 300 kW engines on special foundations).

Class III: Large prime movers or machines mounted on heavy rigid foundations.

Class IV: Large prime movers mounted on relatively soft, lightweight structures/foundations.

TABLE 9.1   Vibration severity as per ISO 2372 (10816)

RMS velocity for vibration severity	Vibration severity for separate classes of machines			
mm/s (RMS)	Class I	Class II	Class III	Class IV
0.28				
0.45	A	A		
0.71				
1.12	B			
1.8		B	A	A
2.8				
4.5	C		B	
7.1	D	C		B
11.2		D	C	
18			D	C
28				D
45				
71				

Vibration severity is divided into four ranges: the letters A, B, C and D representing machine vibration quality grades, ranging from A: good (smoothest) to D: unacceptable (roughest). The particular range selected by the user is based on a number of considerations such as type and size of machine, service expected and effect of machinery vibration on the surrounding environment. The severity rating of a machine also depends on the classification of the supports in the mounting system for the machine. The

supports are said to be soft if the fundamental frequency of the machine on its support is lower than its main excitation frequency and hard if the fundamental frequency of the machine on its support is higher than its main excitation frequency. As a thumb rule, if the persistent value of vibration velocity exceeds 15 mm/s (RMS), it is not advisable to run the machine further. If the predominant frequency of excitation is known, the permissible values of displacements and acceleration may be computed as outlined in Section 3.2.1.

9.2.2   VDI 2056 Standard

VDI 2056 is another standard which was proposed by Verein Deutscher Ingenieure (VDI or Association of German Engineers) and has been serving as a guideline for selection of the measuring points and limits of vibration severity. This has now been replaced by ISO 10816. In accordance with VDI 2056, the standard vibration parameter is the RMS value of the vibration velocity measured in the frequency range 10–1000 Hz. This guideline specifies that the measurements should preferably be taken at the bearings of the machine at the following six locations: drive and non-drive ends in the horizontal, vertical and axial directions. To evaluate the vibration level, the largest amplitude of the six measured points is selected and this is defined as the vibration severity. VDI 2056 classifies machinery into six groups, of which we are primarily concerned with three:

Group K: Small machines, individual parts of engines and machines and complete machines, especially production electric motors up to 20 HP.

Group M: Medium-sized machines, especially electric motors between 20 and 100 HP, without special foundations; rigidly mounted machines on special foundations, with rotating parts only, up to 400 HP.

Group G: Large-sized machines mounted on rigid and heavy foundations; large engines and machines with rotating parts only.

In accordance with VDI 2056, the limiting values for vibration severity for Groups K, M and G are as spelt out in Table 9.2.

TABLE **9.2**   Group grade vibration velocity (mm/s RMS)

	Group K	*Group M*	*Group G*
Good up to	0.69	1.09	1.78
Definitely permissible	0.69–1.8	1.09–2.8	1.78–4.5
Still permissible	1.8–4.5	2.8–6.99	4.5–11
Not permissible above	4.5	6.99	11

9.3   SENSITIVITY OF HUMAN BEINGS TO VIBRATION

Vibrations acting on the human body are called human vibration and may be of the following kinds: (i) motion sickness, typically found on ships or in vehicles with very soft suspension; these occur in the frequency range 0.1–0.63 Hz or (ii) WBV in vehicles and buildings occurring in the frequency range 1–80 Hz. The main purpose of measuring human vibration is for prevention of health risks and evaluation of comfort, for example in vehicles. Prolonged exposure to vibration affects the human body adversely. Two categories of human vibration are as follows:

Whole body vibrations (WBV): These act via the buttocks, the back and the feet of a sitting person, the feet of a standing person or the back and the head of a recumbent person. Such vibrations may cause backache or damage to the spinal column. Drivers of vehicles such as automobiles, trucks, buses, tractors, taxis and locomotives, as well as pilots of helicopters experience WBV. There are two common motion-induced vibratory forces that an operator encounters during daily work; they are: (i) an instantaneous shock with a high-peak level (enough to jar a person out of seat) and (ii) repeated exposures to vibration with low-peak amplitudes caused by regular motion of a vehicle on rough terrain. The effects of such forces vary with the duration of exposure and they are more difficult to quantify than instantaneous damage caused by high-shock loads.

Hand-arm vibrations: These are induced through the hands into the body and are caused by machine tools or workplaces. When powered hand-held tools like chain saws and jack hammers were introduced in the early 1900s, operators started experiencing vascular disorders in fingers and hands and blanching of fingers. This was also noticed among stone masons when they started using pneumatic drills and air hammers instead of mallets and hammers. This vibration disease was called Raynaud's phenomenon, or white finger (WF), or most commonly vibration induced white finger (VWF) and results in reduction of the hand's ability to feel or to regulate its temperature. It may also result in numbness and excessive sensitivity to low temperatures, damage of tissue in hands and arms and also circulatory disorders, as well as bone, joint or muscle diseases.

9.3.1   Vibratory Model of Human Body

Some of the effects of WBV are frequency related, since different parts of the human body have their own natural frequencies and do not vibrate as a single mass, causing resonances of individual parts with the vibration and leading to amplification or attenuation of input vibrations in certain parts. One of the most important parts of the body which is highly sensitive to vibration and shocks is the abdomen with resonances in the vertical direction occurring in the 4–8 Hz range with amplification of up to 200%. Vibrations between 2.5 and 5 Hz generate strong resonances in the vertebrae of the neck and lumbar region with amplification of up to 240%. Vibrations in the range 20–30 Hz setup the strongest resonance between head and shoulders with amplification of up to 350%. Also a vibration in the frequency region 20–90 Hz correlates with the eyeball resonance and affects vision. A vibratory model of the human body is shown in Fig. 9.1. This model is not very useful above 100 Hz, and other more complex models have to be used.

9.3.2   Whole Body Vibration

WBV is typically due to forces in the frequency range 0.5–80 Hz to which the human body responds significantly. Vibrations get transferred from the seat, steering wheels and pedals of a running vehicle to the driver or passenger. The significant effects on health, task performance and comfort due to WBV are described below.

Effects on health: The probability and extent of an injury or health damage is related to WBV. From epidemiological studies, subjective data, biodynamic models and knowledge of the physical properties of the human body, some health effects have been established. Acute effects due to short term vibration exposure are headache, chest pain, abdominal pain, nausea and loss of balance. Chronic effects due to long-term exposure are explained in the following paragraphs.
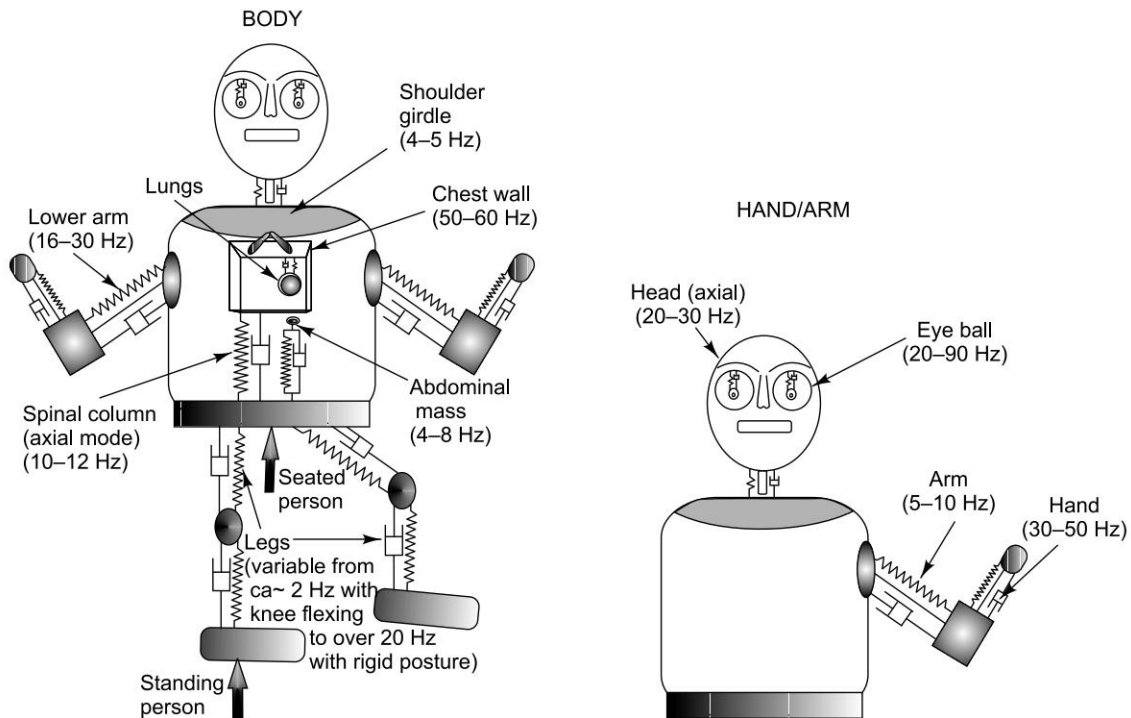
FIGURE 9.1   Vibratory model of the human body

Spinal column disease and complaints: These are the most common diseases associated with long term exposure to WBV, as the spine is especially sensitive to the frequency range 4–12 Hz.

Digestive system diseases: These are observed in persons exposed to WBV over a long period of time and are associated with the resonant movement of the stomach at frequencies between 4 and 5 Hz.

Cardiovascular system effects: Prolonged exposure to WBV at frequencies below 20 Hz could result in hyperventilation, increase in heart rate, pulmonary ventilation and respiratory rate.

Effects on task performance: Some of the WBV effects may not result in injury or health damage, but can cause problems performing different tasks. They may affect the following:

- Senses: problems are created with collecting information
- Processing information
- Level of arousal, motivation or fatigue
- Intentional actions

In general, WBV causes imbalance, disorientation and lack of coordination, which then lead to stress, fatigue, interference with reading of instruments, operation of tools, etc. They can also result in impaired reflex action, distraction and annoyance. Some adverse effects are due to the combination of vibration with other physical factors like noise, temperature, non-ergonomic design of equipment, protective clothing, etc. that are common at workplaces. This combination can also cause stress, fatigue and problems with employee task performance.

Effects on comfort: The impact of WBV on comfort was studied mainly by the transport industry in order to improve ride comfort in vehicles such as buses and trains. It was established that the change in

discomfort is proportional to a change in magnitude of vibration; that is a doubling in vibration magnitude causes a doubling of discomfort. Discomfort is also closely related to vibration frequency; at low frequencies (1–2 Hz), vibration is transmitted without magnification to the human body; at slightly higher frequencies various body resonances tend to amplify the motion, causing overall discomfort. If the frequency is increased further, the human body provides an increasing attenuation of vibration, decreasing the discomfort. Discomfort tends to increase with increasing duration of vibration. It is reflected by the 'fourth power' formula: root mean quad (RMQ) or vibration dose value (VDV). There are also many other factors affecting comfort like body posture, age, gender and noise. Random vibration and multiple axis vibration are known to produce more discomfort than harmonic vibration at a single frequency or along a single axis.

9.3.3 Evaluation of Human Exposure to WBV

Evaluation of WBV involves assessment of human sensitivity to vibration, which depends not only on the physiological and biomechanical response of the human body, but also on a number of psychological and environmental factors. As discussed earlier, human reaction to vibration is a function of amplitude and frequency of acceleration applied to the body, direction (vertical and horizontal) and character of the motion (linear or rotation). The sensitivity of the human body to vibrations also varies with body position (sitting, standing and lying). In the design of vehicle systems, comfort criteria based on power spectral density are regarded as most convenient. In this method, the acceleration spectrum is divided into one third-octave frequency bands (please see Section 13.2.2) and a correlation of RMS acceleration in each band, with human response to vibration is used to specify the criteria.

Ride quality is concerned with the sensation or feel of the passenger in the environment of a vehicle. The most widely followed standards worldwide for evaluating human exposure to WBV are NASA algorithm, ISVR practice, British Standard 6841 (1987) and ISO 2631(1997). The NASA algorithm is based on the measurement of floor vibration in five directions viz. vertical, lateral, longitudinal, pitch and roll. The latter three recommend 12 axes measurements: three translational and three rotational each at the seat, three translational at the seat back and three translational at the feet. The most popular standards are:

(i) ISO Standard 2631-1 (1997) 'Mechanical vibration and shock-evaluation of human exposure to WBV'.
(ii) British Standard 6841 (1987) 'Measurement and evaluation of human exposure to WBV'.

Posture is another frequency dependent parameter; people subjected to vibration in the frequency range from 1 to 30 Hz experience difficulties in maintaining a correct posture and experience an increased postural swing. In a human body, such situations may create chronic stresses and sometimes even permanent damage to the affected organs or body parts. However, the effect of posture is not accounted for in ISO 2631 and British Standard 6841.

9.3.3.1 ISO 2631 guidelines International Standards Organization (ISO 2631, 1985(E)) has specified numerical values for limits of exposure to WBV transmitted from a solid surface to the human body in the frequency range 1–80 Hz. These limits are intended to be used for occupants of vehicles and cover human sensitivity to periodic vibration exposure in the vertical and lateral directions with time ranging from 1 min to 24 hours. Vibration limits are presented as a function of frequency and exposure time for the transverse (front-to-back) and longitudinal (foot-to-head) directions. The International

Standard ISO 2631-1 (1997) emerged as the long-awaited result of an extensive revision process of ISO 2631-1 (1985). The various sub sections of ISO 2631 deal with the following:

Part 1: General requirements

Part 2: Continuous and shock-induced vibration in buildings (1–80 Hz)

Part 3: Whole-body *z*-axis vertical vibration (0.1–0.63 Hz)

Part 4: Vibration on board sea-going ships (1–80 Hz)

Axes of vibration: ISO 2631 has proposed a model (Fig. 9.2) considering various locations in the human body from where the vibrations are transmitted and weighting functions for analysing the effect of vibration on comfort (discomfort model) and for evaluation of ride quality of vehicles. Vibration is measured in terms of its acceleration level along three orthogonal axes, *x*, *y* and *z*. The *x*- and *y*-axes refer to the directions back to chest and right- to left-hand side of a human being as shown in the figure, while *z* refers to direction from feet to head in the case of a standing or recumbent human being and from buttocks to head in the case of a seated human being. In principle, all the translational axes on the surface which supports the subject are included.

FIGURE 9.2   Discomfort model

The model assumes that the main vibration inputs influencing vibration discomfort in a seated position occur at three locations: supporting seat surface, seat back and feet. At each of the input positions, vibration may occur in three orthogonal translational axes defined relative to the orientation of the body, X_s, Y_s and Z_s on the seat, X_b, Y_b and Z_b at the back rest and X_f, Y_f and Z_f on the floor. In addition, vibration in three rotational axes at the seat may be assessed in the R_x (roll), R_y (pitch) and R_z (yaw) axes. Figure 9.3 shows a typical triaxial accelerometer used for measurement of WBV.

Frequency of vibration and weighting curves: Third octave band analysis of acceleration spectra may be done for evaluating influence of vibration on the human body. According to this, the frequency range 0–80 Hz is divided into sub-ranges of one-third octave bands. Effective acceleration is determined by weighting the acceleration in the one-third octave bands in accordance with the weighting curves of ISO 2631-1 and ISO 5349 which

FIGURE 9.3   Triaxial accelerometer for human body vibration measurement (Courtesy of Brüel & Kjær, Denmark)

give different frequency weightings for different axes. There are six frequency curves W_b, W_c, W_d, W_e, W_j, and W_k, which cover the most significant combinations of WBV axes

and effects (health, comfort, perception and motion sickness). Three of these weighting curves are used in the evaluation of vibration from the point of view of health effects and two in the procedure of assessment of vibration severity. There are also weighting curves W_h for hand-arm vibration, W_j for vibration of the head in the axial direction and W_m for WBV in buildings. The various weighting curves are:

W_b : Weighting for vertical WBV, z-axis, seated, standing or recumbent person, based on ISO 2631-4.

W_c : Weighting for horizontal WBV, x-axis, seat back, seated person, based on ISO 2631-1.

W_d : Weighting for horizontal WBV, x- or y-axis, seated, standing or recumbent person, based on ISO 2631-1.

W_e : Weighting for rotational WBV, all directions, seated person, based on ISO 2631-1.

W_f : Weighting for vertical WBV, z-axis motion sickness, seated or standing person, based on ISO 2631-1.

W_h : Weighting for hand-arm vibration, all directions, based on ISO 5349-1.

W_j : Weighting for vertical head vibration (under the head), x-axis recumbent person, based on ISO 2631-1.

W_k : Weighting for vertical WBV, z-axis seated, standing or recumbent person, based on ISO 2631-1.

W_m : Weighting for WBV in buildings, all directions, based on ISO 2631-2.

Magnitude of vibration: The evaluation of vibration severity makes use of the weighted RMS acceleration a_w. The overall weighted vibration level is

$$a_w = \left[\sum_{i=1}^{n} (w_i \cdot a_{w,i})^2 \right]^{1/2} \qquad (9.1)$$

Duration of vibration: Vibration assessment depends strongly on the duration of exposure. ISO curves have been defined for different durations ranging from 1 min to 24 hr; the longer the duration, smaller is the allowable vibration value.

ISO 2631 defines three criteria and they are:
 (i) Fatigue decreased proficiency boundary (preservation of working efficiency).
 (ii) Exposure limits (preservation of health and safety).
(iii) Reduced comfort boundary (preservation of comfort).

The extent to which the ability to work is actually impaired depends on many factors like amplitude, frequency and duration of exposure to vibration. Amplitude of acceleration in the vertical direction for fatigue decreased proficiency is shown in Fig. 9.4 as a function of centre frequency of one third octave bands, fatigue time being the parameter of the curves. Figure 9.5 shows the corresponding weighting factors for one-third octave frequencies. For exposure limits, the acceleration values in Fig. 9.4 are to be multiplied by 2 (i.e. 6 dB higher) and for reduced comfort boundary, the values are to be divided by 3.15 (i.e. 10 dB lower).

The fatigue decreased proficiency curves and weighting curve for lateral vibration are shown in Figs 9.6 and 9.7, respectively.
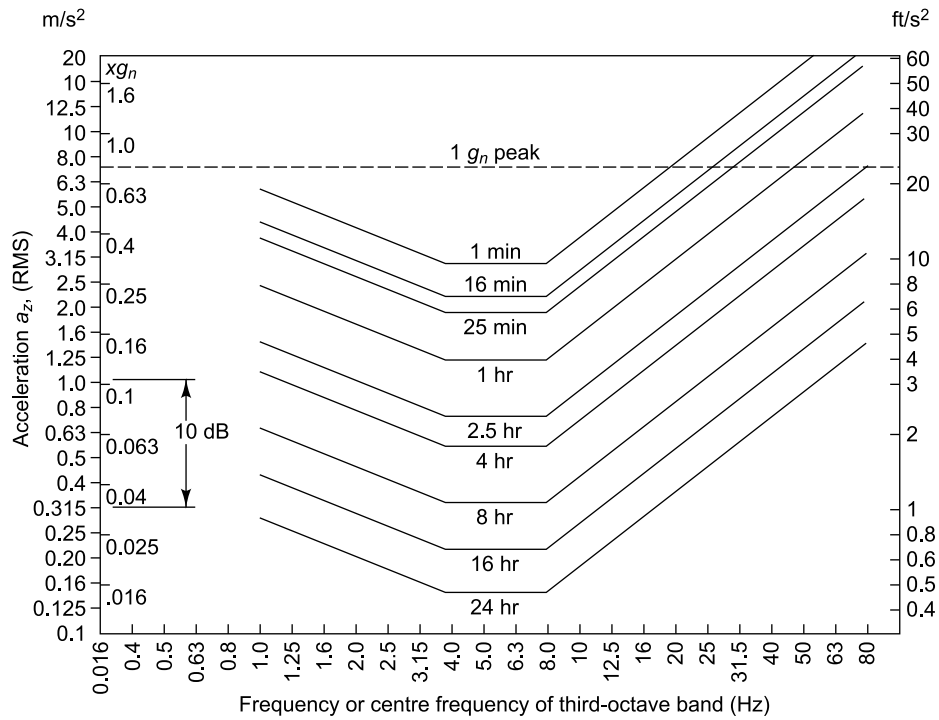
FIGURE 9.4   ISO 2631 Fatigue decreased proficiency boundary–vertical direction (Courtesy of ISO 2631)
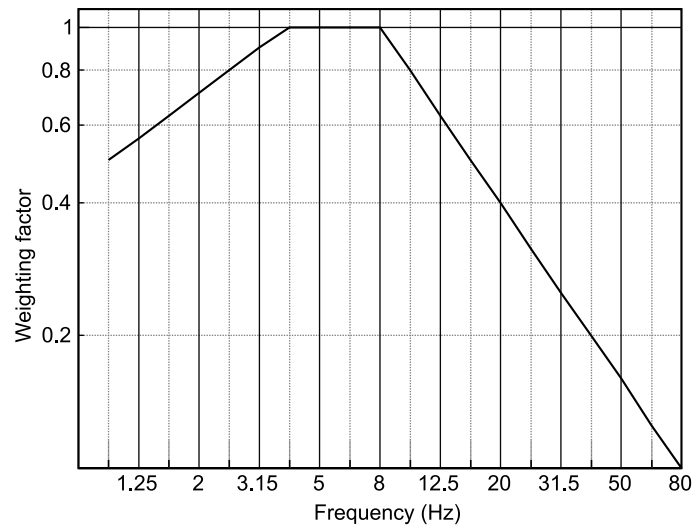
FIGURE 9.5   ISO 2631 Weighting factor W_k—vertical direction
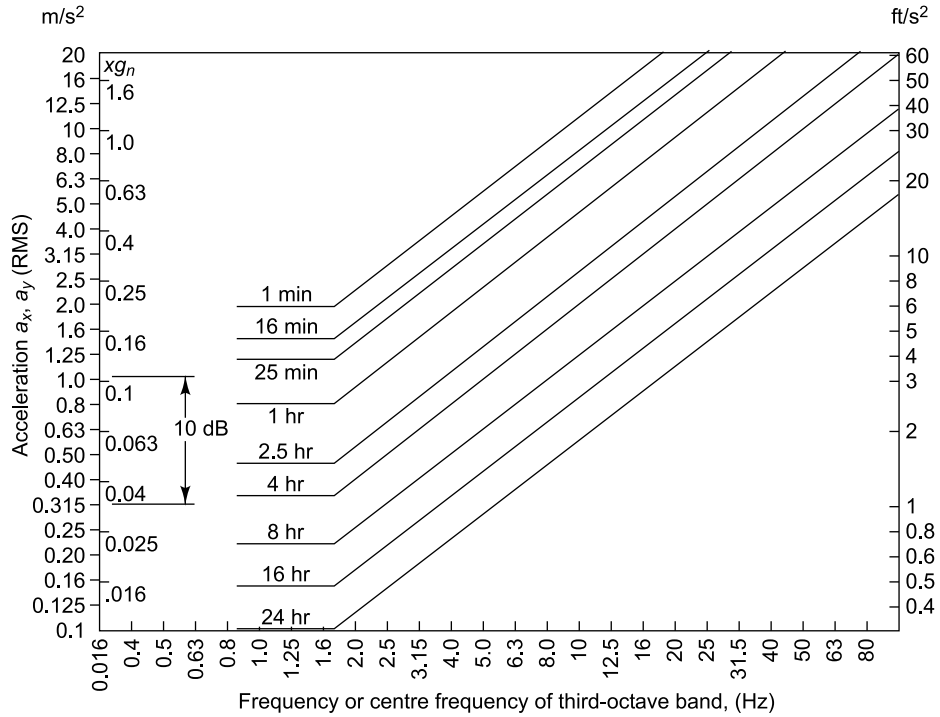
FIGURE 9.6 ISO 2631 fatigue decreased proficiency boundary—horizontal direction (Courtesy of ISO 2631)
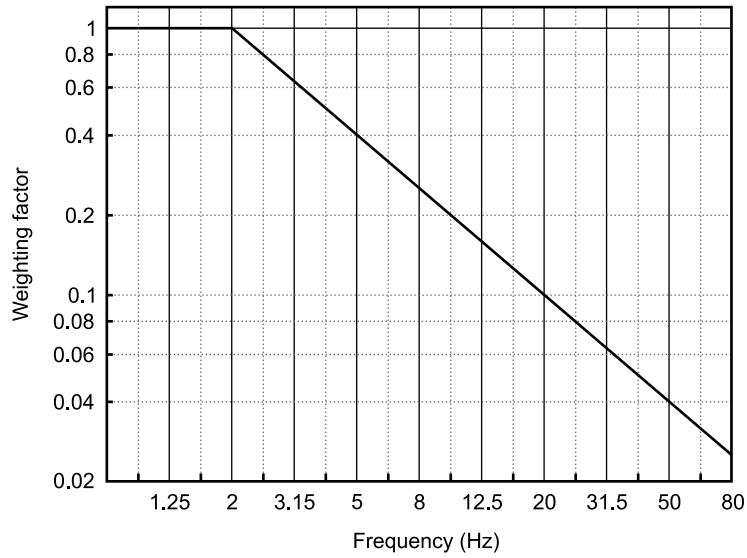
FIGURE 9.7 ISO 2631 weighting factor W_d—horizontal direction

Exposure time based on response: For determination of allowable exposure time to vibration response based on ISO 2631, one-third octave band analysis of acceleration spectra is done and effective (overall) acceleration is determined after weighting the acceleration in the one-third octave bands in accordance with the ISO weighting curve, i.e. Fig. 9.5 for acceleration in the vertical direction and Fig. 9.7 for acceleration in the horizontal direction. The overall weighted vibration levels are then compared with the permissible values in the 4–8 Hz frequency range for WBV in the vertical direction, or in the 1–2 Hz range for WBV in the horizontal direction in order to evaluate exposure time. Figure 9.8 is used for acceleration in the vertical direction and Fig. 9.9 for acceleration in the horizontal direction.
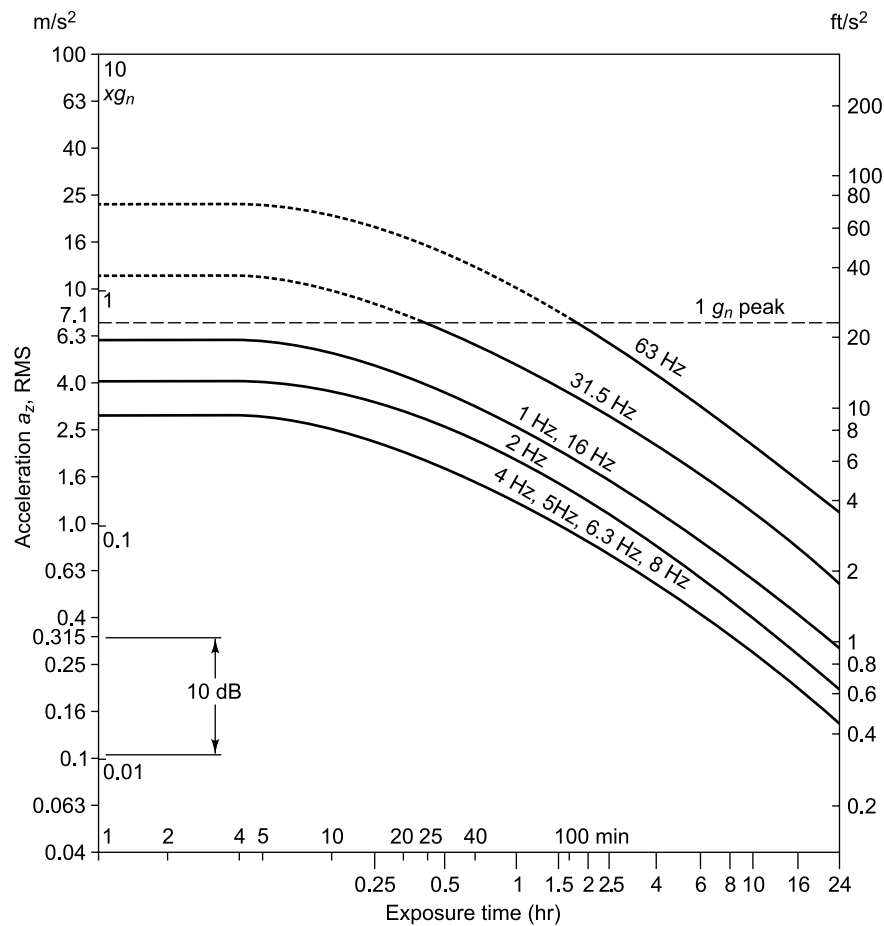
FIGURE 9.8 Vertical acceleration limits as a function of frequency and exposure time—fatigue decreased proficiency boundary (Courtesy of ISO 2631)
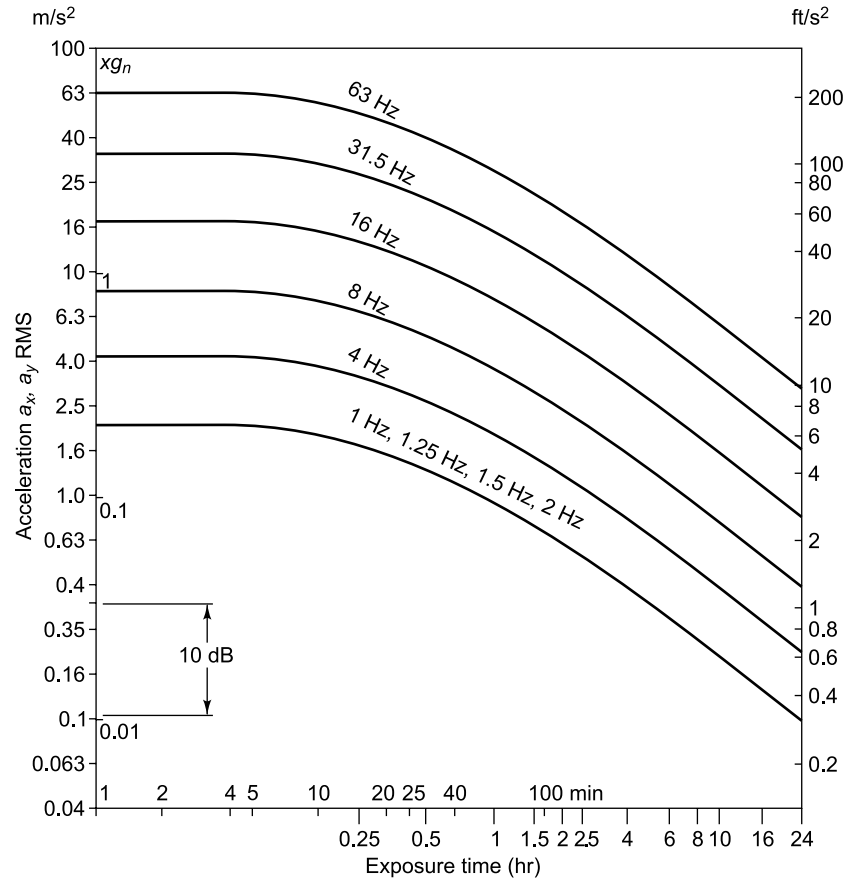
FIGURE 9.9 Horizontal acceleration limits as a function of frequency and exposure time—fatigue decreased proficiency boundary (Courtesy of ISO 2631)

9.3.3.2 BS 6841 standards The British Standard BS 6841 was established in 1987 in response to the perception in Britain as a failure by ISO standard 2631 (1985) to properly address some of the issues of human exposure to WBV and it appears to have served its purpose by appeasing users.

Axes of vibration: BS 6841 uses the same biodynamic coordinate systems as in the case of ISO 2631 for the x-, y- and z-axes for a person standing, sitting and lying down. It also defines three rotational axes r_x, r_y and r_z for rotation about the x-, y- and z-axes, respectively. The standard recommends that measurements be performed on three translational and three rotational axes at the seat, three translational axes between the back and the seat back rest and three translational axes beneath the feet as in the case of ISO 2631.

Frequency of vibration and weighting curves: Six frequency weighting curves are provided covering most combinations of vibration axes and effects (health, hand control, vision, comfort, perception and motion sickness) and are illustrated in Fig. 9.10.
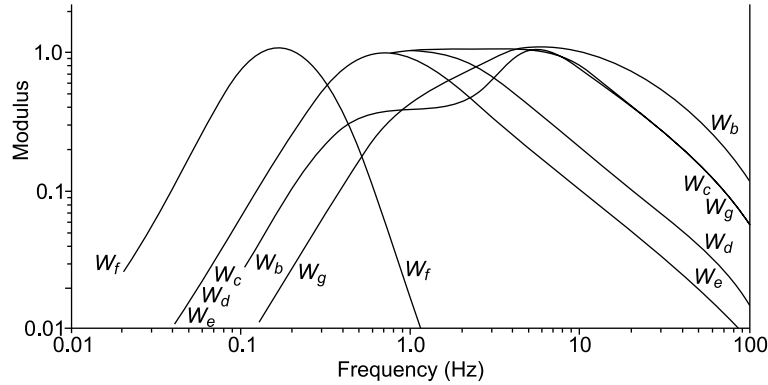
FIGURE 9.10 Frequency weighting curves of BS 6841 (Courtesy of BS 6841)

Only three curves W_b, W_c and W_d are used for the evaluation of the whole-body effects of vibration on health. Weighting W_b is applicable for the z-axis on a supporting surface and weighting W_c for the x-axis on a seat back rest. Weighting W_d is applicable for the x- and y-axes on a supporting surface. The most widely used of these frequency weightings W_b mimics human sensitivity to vertical motions and is based mostly on subjective responses with some input provided by transmissibility data.

Magnitude of vibration: The standard recommends the use of triaxial RMS acceleration. However, when evaluating and assessing vibration for health effects or when the crest factor, which is the ratio of the peak acceleration to the RMS acceleration, is above 6, as in situations of impulsive vibration, VDV is the only option. The VDV was developed as a measure that gives a better indication of the risks from vibrations including shocks (than RMS value) and requires specialized measuring equipment. The units for VDV are m/s$^{1.75}$, and unlike the RMS vibration magnitude, the measured VDV is a cumulative value, based on the fourth root of the time integrated fourth power of the triaxial acceleration signal. VDV can be expressed as

$$\text{VDV} = \left[\int_0^T a^4(t)dt\right]^{0.25} \tag{9.2}$$

In Eq. (9.2) $a(t)$ is the frequency-weighted acceleration in m/s^2 and T is the total period of the day (in seconds) during which vibration may occur.

BS 6841 also states that when the crest factor is below 6 or when only measurements of weighted accelerations a_w are available, the estimated vibration dose value (eVDV) may be used. This is defined as shown below:

$$\text{eVDV} = [(1.4a)^4 \, t]^{1/4} = 1.4 \, at^{1/4} \tag{9.3}$$

eVDV is in m/s$^{1.75}$, the same as for VDV.

Duration of vibration: A unique threshold limiting curve is provided, which applies to all durations from less than a minute to 24 hours.

9.3.3.3   *Absorbed power method*   A different approach to analysing the ride quality and comfort called absorbed power method was first computed by a group of U.S. Army engineers in the mid-1960s

to evaluate and quantify seat-transmitted vibration. It has been proved through tests that the rate of flow of energy becomes the parameter that characterizes the interaction of the vibrating human body with the environment. The coupling between the vibrating structure and the body is inherent in the method, making the result sensitive to both posture of the subject and seat contact dynamic forces. The power absorbed by the human body subjected to vibration is calculated from the simultaneous measurement of force, acceleration and their relative phase. Thus, this method might be a better measure of the physical stress imposed on the body than the ISO or BS standards.

Axes of vibration: This method has been suggested for the measurement of the vibration of seated subjects, but can be extended to the analysis of standing or recumbent subjects and can be used to measure the power absorbed along any given axis. Power being a scalar quantity, it is easy to calculate the cumulative effect of exposure to all three components of vibration. The measurement locations are left to the discretion of the vibration engineer.

Frequency of vibration: The effect of different frequencies is accounted for in the computed value of the absorbed power. Hence, no separate frequency weighting is necessary to link absorbed power to health effects. A time-integrated value of absorbed energy may be obtained by combining the absorbed power with the duration of exposure to vibration to give an assessment of vibration severity.

Magnitude of vibration: This is quantified by the absorbed power. Absorbed power varies linearly as a^2, except for vertical vibration at frequencies around 5 Hz, where the relationship becomes non-linear, with power increasing more rapidly and requiring higher order terms proportional to a^4. This makes a direct conversion of acceleration spectra to power spectra complicated. Frequency-integrated power however is, to a very good approximation linear with a^2. The energy flow that takes place as a result of the complex damped elastic properties of the human body has been designated average absorbed power P_{av}.

$$P_{av} = \frac{1}{T} \int_0^T F(t) V(t)\, dt \tag{9.4}$$

Here $F(t)$ is the input force and $V(t)$ is the input velocity. Absorbed power can also be described in the frequency domain as shown below:

$$P_{av} = \sum_{i=0}^{N} K_i (A_{i,\text{rms}}^2) \tag{9.5}$$

Here $A_{i,\text{rms}}^2$ is the mean square acceleration and K_i is a frequency dependent weighting parameter. Values of K_i for vertical response as a function of frequency are available in literature and have been shown in the frequency range 0–10 Hz in Table 9.3.

9.3.3.4 *Hand-arm vibration testing according to ISO 5349*

Hand-arm vibration as mentioned earlier is vibration transmitted to the hands and arms of a person by the use of vibrating hand-held power tools. The tools known to cause hand-arm vibration are chainsaws, concrete breakers, cut-off saws, hammer drills, hand-held grinders, impact wrenches, pedestal grinders, polishers, power hammers and chisels, power lawn mowers and trimmers. Regular exposure to hand-arm vibration may cause pain in the arms and shoulders, vibration-induced WF syndrome, reduced grip strength and carpal tunnel syndrome. These effects can further lead to sleep disturbances and reduction of efficiency at work. The duration of operation and the frequency of vibration of the tools increase the risk of health effects.

The industries which are to be concerned with hand-arm vibration are those involved in construction and maintenance of buildings, roads and railways, heavy manufacturing, shipbuilding, forestry, mining and quarrying, foundries, etc.

TABLE 9.3 Frequency dependent weighting parameters for calculation of absorbed power (Lee and Pradko, 1960)

Sl. No.	f (Hz)	K_i
1	0.625	0.0040265
2	1.25	0.015101
3	1.875	0.037814
4	2.5	0.077843
5	3.125	0.127606
6	3.75	0.174830
7	4.375	0.2010225
8	5.0	0.198689
9	5.65	0.17282
10	6.25	0.134302
11	6.875	0.10045
12	7.5	0.075581
13	8.125	0.06107
14	8.75	0.0537
15	9.375	0.0505
16	10.0	0.048340

ISO 5349-1 relates to hand-arm vibration only; it defines both a basicentric coordinate system (in respect of the hand-grip feature on the tool) and a biodynamic coordinate system (in respect of the anatomy of the hand) as shown in Fig. 9.11. Measurements have to be made along three axes since one cannot assume that the tool has a dominant axis.
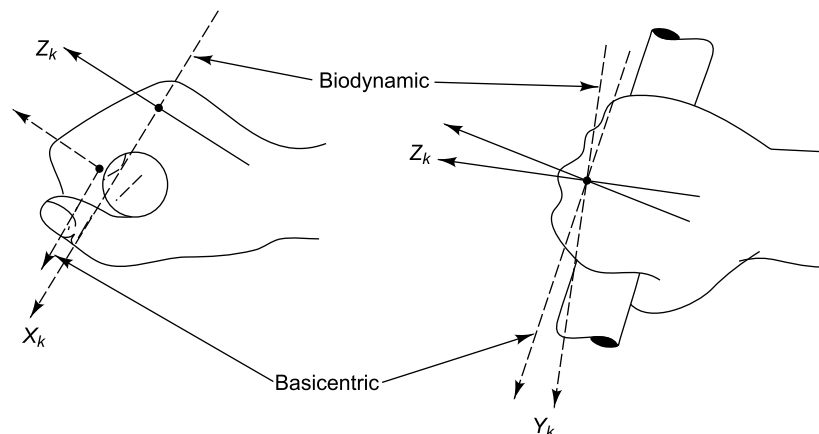
FIGURE 9.11 Biodynamic and basicentric coordinate systems for hand-arm vibration

Hand-arm vibration measurement involves measuring the accelerations at the contact surface between the hand and the tool, i.e. at the handle position of a hand-held power tool. The measurements may be

done on one or two handles simultaneously using a transducer with a small mass as compared to the object under test in order to avoid mass loading. The vibration magnitude, a_{hv}, is computed as a superposition of the vibrations along all three axes and is expressed as:

$$a_{hv} = \sqrt{\left(a_{hux}^2 + a_{hwy}^2 + a_{hwz}^2\right)} \tag{9.6}$$

where a_{hux}, a_{hwy} and a_{hwz} are the RMS acceleration magnitudes in m/s^2, measured in the three orthogonal directions, x, y and z at the vibrating surface in contact with the hand, and frequency-weighted using the weighting W_h (discussed in Section 9.3.3.1) as shown in Fig. 9.12.
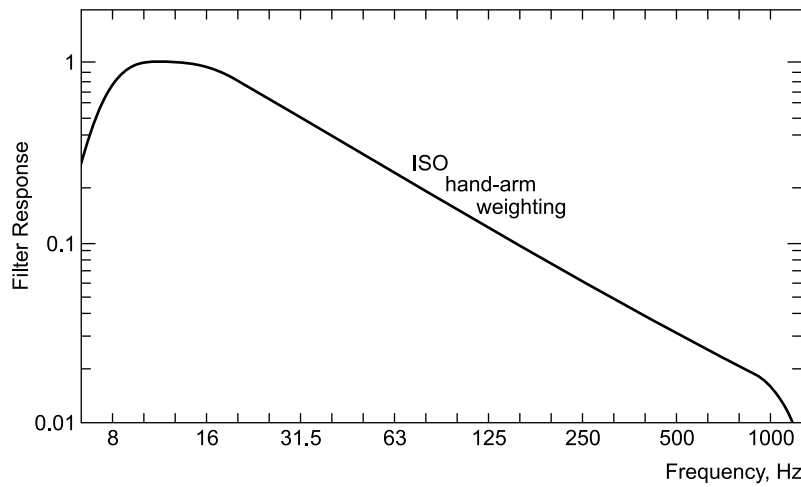
FIGURE 9.12   Frequency weighting W_h for hand-arm vibration

The total daily vibration exposure $A(8)$ is defined as the equivalent continuous RMS acceleration that would have to be maintained over a time duration T_0 = 8 hr to give the worker the same vibration dose as was actually experienced in the working day (irrespective of how many hours). To avoid confusion between vibration magnitude and daily exposure to vibration, it is customary to express daily exposure to vibration $A(8)$ in m/s^2 which is calculated using the formula:

$$A(8) = a_{hv}\sqrt{\frac{T}{T_0}} \tag{9.7}$$

where a_{hv} is the vibration magnitude, in m/s^2, T is the duration of exposure to the vibration magnitude a_{hv} and T_0 is the reference duration of 8 hours (28,800 s). In situations where the resultant acceleration fluctuates over time, a_{hv} is taken to be the equivalent continuous RMS acceleration. If a machine tool operator performs several tasks during a work day, being subjected to continuous accelerations for each task, the total daily vibration exposure $A(8)$ can be calculated as

$$A(8) = \sqrt{\frac{1}{T_0}\sum_{i=1}^{n}a_{hwi}^2 T_i} \tag{9.8}$$

where n is the number of individual operations within the working day, a_{hwi} is the vibration magnitude for operation i and T_i is the duration of operation i. Where both hands are exposed to vibration, the greater of the two magnitudes a_{hv} is used to ascertain the daily exposure.

9.3.3.5 *Directive 2002/44/EC of the European Parliament* Directive 2002/44/EC of the European Parliament and of the Council places requirements on employers to ensure that risk arising to workers from exposure to vibrations are eliminated or reduced to a minimum. According to the directive, human exposure to WBV should be evaluated using the method defined in International Standard ISO 2631-1:1997. The RMS vibration magnitude is expressed in terms of the frequency-weighted acceleration at the seat of a seated person or the feet of a standing person and is expressed in units of m/s². The RMS vibration magnitude used for risk assessment is the highest of the three values along the three orthogonal axes. The EU Physical Agents Directive (Vibration) or PAD(V) quantifies continuous, non-impulsive vibration exposure in terms of A (8) which has been defined earlier.

The directive also defines daily exposure limit values and daily action limit values above which workers should not be exposed. Once the exposure action value is exceeded, the employer should establish measures to arrive at a minimum exposure to mechanical vibration. In any event, workers shall not be exposed to values above the exposure limit. The assessment of the level of exposure to vibration is based on the calculation of daily exposure A (8) including all hand-arm vibration exposures during the day. For the determination of A (8) it is sufficient to make short-term measurements during representative work steps and to extrapolate the results to 8 hours. Daily exposure is calculated as follows:

$$A_x \ (8) = \ 1.4 a_{wx} \ \sqrt{\frac{T_{\exp}}{T_0}}$$

$$A_y \ (8) = \ 1.4 a_{wy} \ \sqrt{\frac{T_{\exp}}{T_0}} \tag{9.9}$$

$$A_z \ (8) = \ a_{wz} \ \sqrt{\frac{T_{\exp}}{T_0}}$$

Here a_{wx}, a_{wy} and a_{wz} are the RMS acceleration magnitudes in m/s², during the daily exposure measured in three orthogonal directions x, y and z at the vibrating surface and frequency-weighted using the weighting W_h. $T_{\exp}$ is the total duration of exposure during one work day and T_0 is the reference duration of 8 hours.

The VDV provides an alternative measure of vibration exposure. VDV increases with measurement time and it is therefore important that in any measurement of VDV, the period over which the vibration is measured is noted. $\text{VDV}_{\exp, \ x}$, $\text{VDV}_{\exp, \ y}$ and $\text{VDV}_{\exp, \ z}$ give the VDV exposure values in the three orthogonal directions. The highest of the three orthogonal axes values $1.4 \times \text{VDV}_{\exp.x}$, $1.4 \times \text{VDV}_{\exp.y}$ and $\text{VDV}_{\exp.z}$ is used for risk assessment.

$$\text{VDV}_{\exp.x} = 1.4 \times \text{VDV}_x \left(\sqrt{\frac{T_{\exp}}{T_0}} \right)^{0.25}$$

$$\text{VDV}_{\text{exp}.y} = 1.4 \times \text{VDV}_y \left(\sqrt{\frac{T_{\text{exp}}}{T_0}} \right)^{0.25} \tag{9.10}$$

$$\text{VDV}_{\text{exp}.z} = \text{VDV}_z \left(\sqrt{\frac{T_{\text{exp}}}{T_0}} \right)^{0.25}$$

The EU Physical Agents Directive has set the following limiting values for exposure to vibration in the workplace as shown in Table 9.4 for hand-arm and WB vibrations. Corresponding VDV values are also shown.

TABLE 9.4   Daily exposure $A(8)$ for hand-arm and whole body vibrations

	Hand-arm (m/s²)	Whole body (m/s²)
Daily exposure limit value $A(8)$	5	1.15 (VDV = 21 m/s¹·⁷⁵)
Daily exposure action value $A(8)$	2.5	0.5 (VDV = 9.1 m/s¹·⁷⁵)

9.4   RIDE COMFORT ANALYSIS IN RAIL VEHICLES

Sperling's Ride Index (Wz) was introduced by Sperling, in order to evaluate ride quality and ride comfort of a railroad vehicle. The vehicle itself is judged by the ride quality and ride comfort implies that the vehicle is assessed according to the effect of mechanical vibration on the occupant.

$$Wz = (a^2 B^2)^{1/6.67} \tag{9.11}$$

where a is the amplitude of acceleration in cm/s² and B is the acceleration weighting factor, which is given by

$$B = 1.14 \left[\frac{[(1 - 0.056 f^2)^2] + (0.0645)^2 (3.35 f^2)}{[(1 - 0.252 f^2)^2] + (1.547 f - 0.00444 f^3)^2 (1 + 3.35 f^2)} \right]^{1/2} \tag{9.12}$$

where f is the frequency in Hz. The weighting factor B_s for ride comfort in the vertical direction is given by

$$B_s = 0.588 \left[\frac{1.911 f^2 + (0.25 f^2)^2}{(1 - 0.277 f^2)^2 + (1.563 f - 0.0368 f^3)^2} \right]^{1/2} \tag{9.13}$$

Since the vibration of the vehicle body encompasses a whole spectrum of frequencies in which the natural frequencies of the vehicle are pronounced, the ride index calculation has to be done for the entire spectrum. The Wz ride factor is determined for each individual frequency from the equations mentioned above and the total Wz factor is calculated as

$$Wz_{\text{total}} = \left(Wz_1^{10} + Wz_2^{10} + Wz_3^{10} + \ldots + Wz_n^{10} \right)^{1/10} \tag{9.14}$$

Evaluation standards for ride quality and ride comfort indices are shown in Table 9.5.

TABLE 9.5   Ride evaluation scales—ride quality and ride comfort

Ride index Wz	Ride quality
1	Very good
2	Good
3	Satisfactory
4	Acceptable for running
4.5	Not acceptable for running
5	Dangerous
Ride Index Wz	*Ride comfort*
1	Just noticeable
2	Clearly noticeable
2.5	More pronounced but not unpleasant
3	Strong, irregular, but still tolerable
3.25	Very irregular
3.5	Extremely irregular, unpleasant, annoying; prolonged exposure intolerable
4	Extremely unpleasant ; prolonged exposure harmful

9.5   VIBRATIONS AFFECTING PEOPLE IN BUILDINGS

Building vibrations fall into two categories: those that arise from internal sources and those that arise from external sources. Most vibrations generated inside buildings originate from machines (cranes, trolleys, elevators, fans, pumps, drop hammers and punching presses) and from the activities of people (walking, jumping, dancing and running). Externally generated vibrations commonly arise from road and rail traffic, subways, construction activities (pile driving, blasting, excavation and compacting of soil), sonic booms, strong winds and earthquakes. The resulting vibrations in buildings may cause annoyance to the occupants, impaired function of instruments, or structural damage.

Most vibration problems can be described in terms of the source, the transmission path and the receiver. This provides a convenient subdivision of the overall problem and a simpler approach to many remedial measures. When vibrations are transmitted from the source through the soil into the building foundations, they propagate sideways and upwards throughout the building with the soil having a strong influence on the intensity of the vibrations received in the building. The type of foundation selected for the building is also important. Certain research organizations have established building vibration limits to lie between 12.5 and 50 mm/s before expected damage.

There are basically two kinds of vibration that affect people in buildings, namely:

 (i) Vibration transmitted to the human body as a whole through the supporting surface: through the feet when standing, the buttocks when seated, or the supporting area when reclining. The general case of human response to vibration as described above is covered by ANSI S3.18-1979, which recommends vibration magnitudes corresponding to task proficiency, human comfort, health and safety.

(ii) Vibrations of the building and the resulting reactions of the occupants: this second kind of exposure results from the vibration of the whole structure, floor vibration (primarily vertical motion) and wall vibrations (primarily horizontal motions). These vibrations evoke fear of damage to the structure or its contents, startle, and interference with sleep or conversation.

There are various standards pertaining to vibrations in buildings. Human discomfort values for continuous vibration have been extensively researched and published in BS 6472 (1984 and 1992), ISO 2631.2-1989 and AS 2670.2-1990. The American National Standards Institute (ANSI) in conjunction with the Acoustical Society of America have developed and published ANSI S3.29-1983.

9.5.1   BS 6472:1992

This standard deals with human exposure to vibration in buildings in the frequency range 1-80 Hz. It uses the same coordinate systems as BS 6841 for the *x*-, *y*- and *z*-axes for a person standing, sitting and lying down. This standard uses frequency-weighted one-third octave band vibration values to evaluate and assess the effects of continuous and impulsive vibration on building occupants. The weighted RMS acceleration curve used is shown in Fig. 9.13 for *z*-axis vibrations. The minimum value, i.e. 0.005 m/s² (RMS) roughly corresponds to the threshold of perception of continuous *z*-axis vibration for most people. A similar curve for *x*- and *y*-axes vibrations is shown in Fig. 9.14.
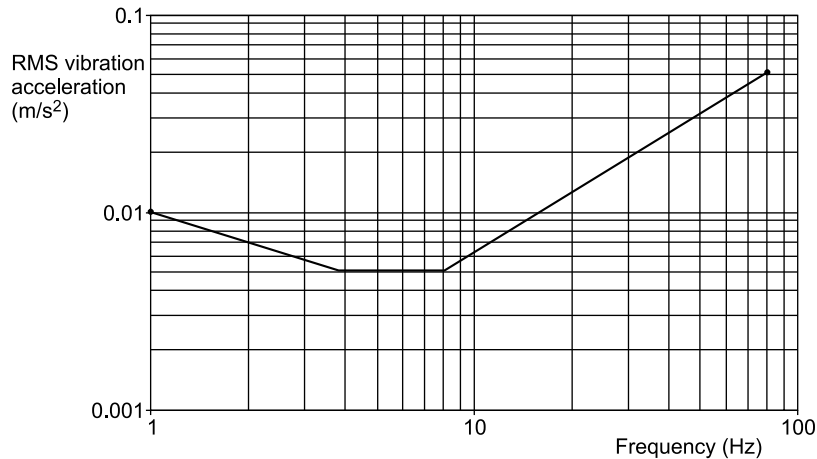
FIGURE 9.13   Weighted *z*-axis vibration RMS acceleration (Courtesy of BS 6472-1992)
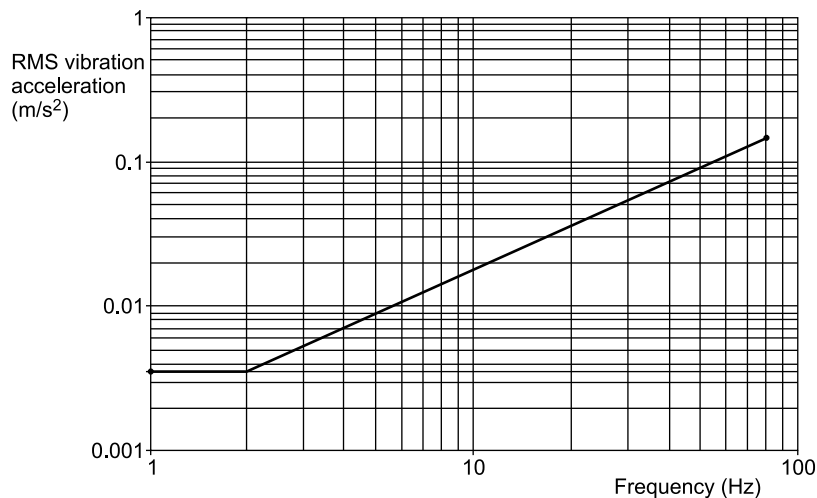
FIGURE 9.14   Weighted *x*- and *y*-axes RMS acceleration (Courtesy of BS 6472-1992)

The 'threshold of perception' curves shown in Figs 9.13 and 9.14 can also be expressed as equivalent weighted peak vibration velocity curves as depicted in Figs 9.15 and 9.16, respectively. The minimum value for z-axis vibration corresponds approximately to the threshold of perception, i.e. 0.14 mm/s.
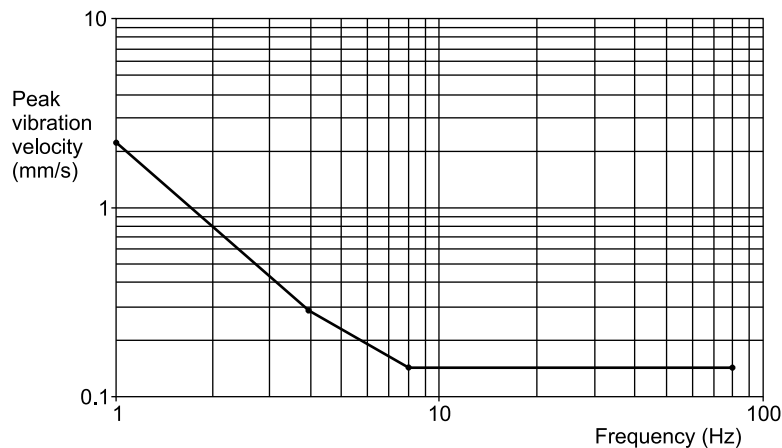
FIGURE 9.15   Weighted z-axis peak velocity (Courtesy of BS 6472–1992)
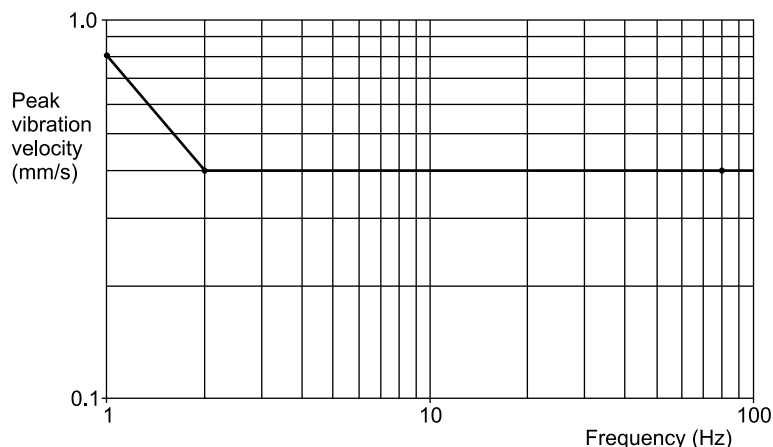
FIGURE 9.16   Weighted x- and y-axes peak velocity (Courtesy of BS 6472–1992)

It is seen from Fig. 9.15 that z-axis velocity requires no frequency weighting in the range 8–80 Hz in order to determine annoyance. Likewise x- and y-axes vibration velocities do not require any weighting in the range 2–80 Hz as seen from Fig. 9.16.

Based on these threshold of perception curves, the standards relate the degree of adverse comments from building occupants to multiples of the base-weighted z-axis vibration values (i.e. the values specified in the curve in Fig. 9.13). Allowable magnitudes of building vibration are specified in accordance with the type of occupancy, period of the day and type of vibration (i.e. continuous or impulsive) by applying the multiplication factors shown in Table 9.6 to the threshold values in Fig. 9.13 to get Fig. 9.17. Fig. 9.18 shows the x- and y-axes curves corresponding to different multiplication factors for building vibration.

TABLE 9.6   Multiplying factors for satisfactory magnitudes of building vibration (BS 6472–1992)

Place	Time (hr) Day = 16 Night = 8	Multiplying factors (curve numbers)	
		Exposure to continuous vibration	Impulsive vibration excitation with up to three occurrences
Critical working areas (e.g. operating theatres, precision laboratories)	Day	1	1
	Night	1	1
Residential	Day	2–4	60–90
	Night	1.4	20
Office	Day	4	128
	Night	4	128
Workshops	Day	8	128
	Night	8	128
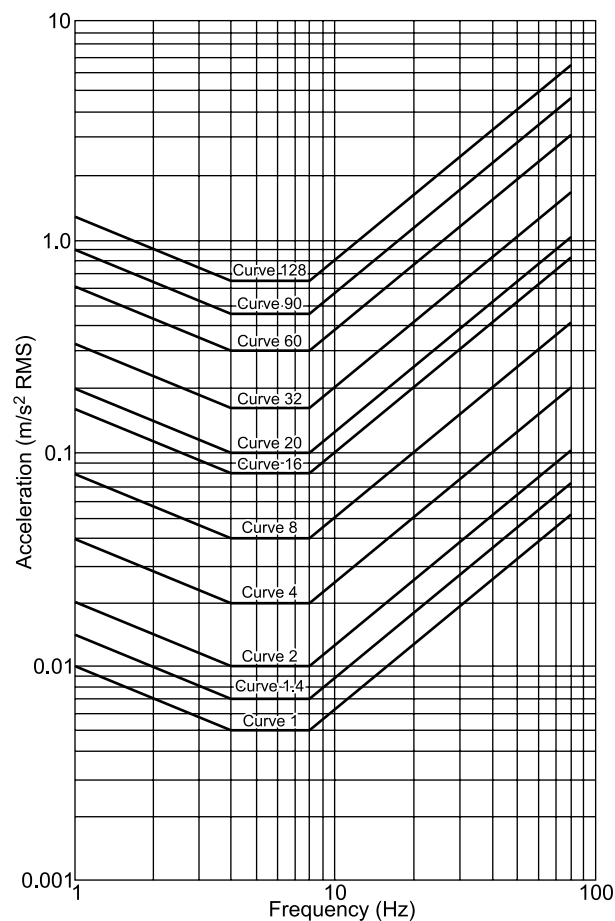
FIGURE **9.17**   Building vibration *z*-axis curves for RMS acceleration (Courtesy of BS 6472–1992)
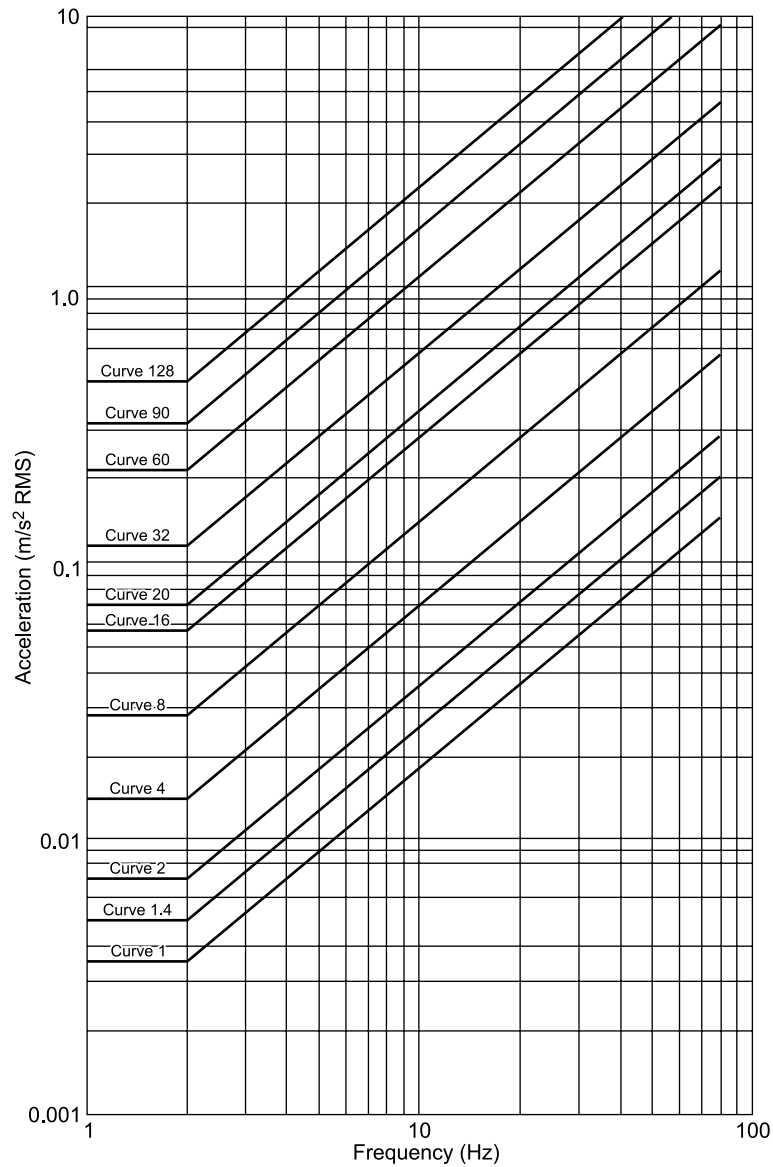
FIGURE 9.18   Building vibration *x*- and *y*-axes curves for RMS acceleration (Courtesy of BS 6472–1992)

9.5.2   Short-Term Vibration—DIN 4150:1986 Standard

This standard specifies limits for short term vibration from any activity measured on the foundation as well as the uppermost storey of a building. The values given are combined-direction peak velocity values.

TABLE 9.7   Vibration criteria for short term vibration in structures (DIN 4150-3:1986)

Type of structure	Peak component of velocity at foundation (mm/s)			On floor of uppermost storey (mm/s)
	Vibration frequency			With many frequency
	<10 Hz	10–50 Hz	50–100 Hz	components
Industrial buildings	20	20–40	40–50	40
Dwellings	5	5–15	15–20	15
Buildings under a preservation order	3	3–8	8–10	8

FURTHER READINGS

1. Allen, D.E. and Rainer, J.H., Floor Vibrations, *National Research Council Canada, Division of Building Research, Canadian Building Digest 173*, source: http://irc.nrc-cnrc.gc.ca/cbd/cbd232e.html, September 1975.

2. ANSI S3.18-1979 (reaffirmation of ANSI S3.18-1979 (R1993)): (R1999) Whole-Body Vibration, *Guide for the Evaluation of Human Exposure*, 1979.

3. ANSI S3.29-1983 (ASA 48-1983), *Guide to the Evaluation of Human Exposure to Vibration in Buildings*, American National Standard, Standards Secretariat, Acoustical Society of America, NY 10017, 1983.

4. British Standards Institution BS 6841: 1987 Guide to Measurement and Evaluation of Human Exposure to Whole Body Mechanical Vibration and Repeated Shock, London, 1987.

5. Brüel & Kjær (B&K), *Human Body Vibration: Technical Review*, Denmark, 1982.

6. BS 6472: 1984, *Guide to Evaluation of Human Exposure to Vibration in Buildings* (1 Hz to 80 Hz), British Standards Institution, London, 1984.

7. BS 6472:1992 *Guide to Evaluation of Human Exposure to Vibration in Buildings* (1 Hz to 80 Hz), 1992.

8. DIN 4150:1986 Part 3 Structural Vibration in Buildings, 1986.

9. Drogicina, E.A. and Razumov, I.K., *Vibration, Encyclopedia of Occupational Health and Safety*, Vol. 2, McGraw-Hill Book Co., New York, 1972.

10. Engelhardt, R.E., Mills, K.D. and Schneider, K., *Shock and vibration in road and rail vehicles, Shock and Vibration Handbook*, Vol. 3, McGraw-Hill Co., New York, pp. 45–1–45–19, 1961.

11. Goldman, D.E. and Henning E. Von Gierke, *Effect of shock and vibration on man, Shock and Vibration Handbook*, Vol. 3, McGraw-Hill Co., New York, pp. 44–1 to 44–49, 1961.

12. Griffin, M.J., A comparison of standardized methods for predicting the hazards of whole body vibration and repeated shock, *Journal of Sound and Vibration*, 215, 883-914, 1998.

13. *International Organisation for Standardisation* ISO 2631–1: 1985 (E), Mechanical Vibration and Shock-Evaluation of Human Exposure to Whole Body Vibration, Part 1: General Requirements, ISO, Geneva, Switzerland, 1997.

14. *International Organisation for Standardisation* ISO 2631–1: 1985 (E), Mechanical Vibration and Shock-Evaluation of Human Exposure to Whole Body Vibration, Part 2: Continuous and Shock-induced Vibration in Buildings (1 to 80 Hz), ISO, Geneva, Switzerland, 1997.

15. *International Organisation for Standardisation* ISO 2631-3-1985(E), Mechanical Vibration and Shock-Evaluation of Human Exposure to Whole Body Vibration, Part 3: Whole-body *z*-axis Vertical Vibration in the Frequency Range 0.1 to 0.63 Hz, ISO, Geneva, Switzerland, 1997.
16. *International Organisation for Standardisation* ISO 2631-1: 1985 (E), Mechanical Vibration and Shock-Evaluation of Human Exposure to Whole Body Vibration, Part 4: Vibration on Board Sea-going Ships (1 to 80 Hz), ISO, Geneva, Switzerland, 1997.
17. *International Organisation for Standardisation*, ISO 5349-1: 2001, Mechanical Vibration and Shock-Measurement and Evaluation of Human exposure to Hand-transmitted Vibration, Part 1: General Requirements, ISO, Switzerland, 2001.
18. ISO 6897:1984, *Guidelines for the Evaluation of the Response of Occupants of Fixed Structures*, Especially Buildings and Off-shore Structures, to Low-frequency Horizontal Motion (0.063 to 1 Hz), 1984.
19. *International Organisation for Standardisation*, ISO 2372: 1974, ISO 10816 -1:1995, Mechanical Vibration of Machines with Operating Speeds from 10 to 200 rev/s—Basis for Specifying Evaluation Standards, 1995.
20. Lee, R.A. and Pradko, F., *Analytical Analysis of Human Vibration*, SAE Paper 680091, pp. 346–370, 1968.
21. Lewis, C.H. and Griffin, M.J., A comparison of evaluations and assessments obtained using alternative standards for predicting the hazards of whole-body vibration and repeated shock, *Journal of Sound and Vibration*, 215, 915–926, 1998.
22. Lundström, R. and Holmlund, P., Absorption of energy during whole-body vibration exposure, *Journal of Sound and Vibration*, 215, 801–812, 1998.
23. Mansfield, N.J. and Griffin, M.J., Effect of magnitude of vertical whole-body vibration on absorbed power for the seated human body, *Journal of Sound and Vibration*, 215, 813–825, 1998.
24. ORE Report, C116/8, Interaction between Vehicle and Track, Utrecht, 1978.
25. Reiher, H. and Meister, F.J., The sensitivity of people against vibration, *Forschung ang dern Gebiet des Ingeniee erweigens*, 2(11), 381–386, 1931.
26. Seidel, H., Blüthner, R., Hinz, B. and Schust, M., On the health risk of the lumbar spine due to whole-body vibration-theoretical approach, experimental data and evaluation of whole-body vibration, *Journal of Sound and Vibration*, 215, 723–741, 1998.
27. Stikeleather, L.F., Review of ride vibration standards and tolerance criteria, *SAE Trans.* No. 760413, 1460–1467, 1976.
28. VDI 2056 Evaluation of Mechanical Vibrations of Rotating Machinery (withdrawn and replaced by ISO 10816), 1964.

Basics of Experimental Modal Analysis

10.1 INTRODUCTION

Experimental modal analysis deals with the determination of modal parameters, such as natural frequencies, damping properties and mode shapes of a structure through experiments. This was once considered an area requiring expert knowledge, but with the advent of personal computers and the development of inexpensive, user-friendly software packages for modal parameter extraction, as well as signal processing software and hardware, it has evolved into a common tool readily available in vibration toolboxes and accessible to most test engineers. In any experimental modal analysis procedure, modal parameters are to be estimated from the measured frequency response functions (FRFs), i.e. from the output response and input force data and the quality of an experimental modal model is only as good as the quality of the FRFs. Hence, several important experimental aspects are to be considered while conducting modal tests in order to obtain valid modal data. Choice of excitation of the structure, in terms of excitation location and mechanism, excitation signal, frequency range and amplitude of the excitation force, fixing/mounting of test structure so as to minimize exciter/test structure interaction are to be looked into. Besides, selection of appropriate transducers and their positioning and mounting, as well as aspects related to signal processing, have to be considered and are described in this chapter.

Though the entire process of obtaining the required data is experimental in nature, many mathematical techniques are involved in the computation of FRFs and modal parameter extraction. Experimental modal analysis is therefore a difficult application, especially where high accuracy is required; besides an in-depth knowledge of structural dynamics is also required. Therefore mathematical theory, as found necessary and drawn heavily from Ewins (2003), has been presented in this chapter to obtain a logical development of these techniques. Starting with the basics of the representation and properties of various forms of FRFs of a single-degree-of-freedom (SDOF) linear dynamic system, we progress to FRF representation of multiple-degree-of-freedom (MDOF) systems, since this is how most physical structures are modelled. The discussion of undamped cases is followed by those with viscous and structural damping. Finally, parameter estimation techniques are presented to gain insight, first into the single mode concepts, followed by concepts related to MDOF systems. Modal testing, with its structural dynamics and modal parameter estimation theories, is based on an ideal physical model, i.e. the system is linear, time-invariant, causal and observable. However, most structures deviate from this ideal behaviour and the resulting problems have to be anticipated and dealt with in any modal test setup. It is hoped that the material presented here will give the reader a sound understanding of modal analysis concepts so as to enable him/her to confidently embark on a modal analysis exercise.

10.2   IMPORTANT EXPERIMENTAL ASPECTS OF MODAL TESTING

Some of the important practical aspects to be considered in modal analysis are the boundary conditions of the test structure, minimisation of exciter/test structure interaction, choice of exciters/shakers, problems in the measurement of excitation force, difficulties encountered in impact testing, sensing techniques and fixtures as well as selection of excitation signals for modal testing. These have been discussed in the sections to follow.

10.2.1   Support Conditions of Test Structure

For structures such as buildings, towers, machinery, vehicles, etc., modal testing is done to evaluate system dynamics, or from the point of view of redesign or modification of actual fixing conditions in order to improve their vibration isolation characteristics or fatigue lives. The fixture location and the point of application of excitation force (either through hammer or shaker) require careful consideration. The support and fixing conditions used for a test structure largely dictate the success of a modal test. Often the fixture can distort the test input so badly that the specimen is overtested at some frequencies and undertested at others, causing the results to be meaningless. Hence, any basic test setup used for modal analysis should be carefully planned, taking into account the type of the structure to be tested and the level of accuracy of the desired results. It is important to find out all the fixity conditions in which the structure may be used under actual operating conditions and choose the appropriate fixtures required to obtain the desired support constraints or boundary conditions. If the structure is not properly prepared, then there is degradation of the overall structural characteristics obtained. This affects subsequent analyses such as structural modification, finite element correlation and substructure coupling.

The main decision to be taken is whether to conduct the test in the (i) free condition or (ii) fixed/grounded condition or (iii) actual fixity conditions at site. Analytically, the first two boundary conditions can be achieved very easily by either allowing all degrees of freedom for the first case, or arresting the DOF at the support for the second case. However, it is not so easy to simulate these conditions while testing. Free condition implies that the structure is not attached to the ground anywhere and is freely suspended or floating in space. With such a condition, the structure exhibits rigid body behaviour ideally at low frequencies close to zero frequency and does not undergo any bending or flexing. The airplane or rocket in flight is an example of this free condition. In practice however, the rigid body frequencies (three translational and three rotational) will not be zero. Testing with this boundary condition therefore allows determination of the rigid body modes and hence the mass and inertia properties of the structure. Physically, achieving a truly free support is not realizable, so the structure must be held in some manner. This may be achieved by suspending the structure using very soft, highly elastic bands or suspension cords or by placing the test structure on very soft sponge or cushion or springs. Due to such an arrangement, the structure will be constrained to some extent and the rigid body modes will therefore have low frequencies. If a sufficiently resilient support is used, the frequencies corresponding to the rigid body modes will be quite low as compared to the frequencies of the bending modes and will thus have negligible influence. As a guideline in testing with free supports, the frequency of the highest rigid body mode should be less than one tenth that of the first flexible mode. The excitation frequency band during modal testing should be carefully chosen to avoid exciting the rigid body modes which can cause failure of the suspension cords.

A fixed boundary condition implies that all motion: displacements as well as rotations, at the clamped edge, or at the support, are forced to zero. This condition is much more difficult to achieve than a free boundary condition, since most structures have some amount of flexibility at the clamps. The base to which the structure is attached will typically have some motion of its own due to flexibility of the bolted, riveted or welded connections. To ascertain whether this motion is negligible or not, FRFs may be measured at the base over the frequency range of interest and compared with the corresponding response of the structure. If the former is negligibly small, the structural data may be taken to be accurate. Figure 10.1 shows free and fixed boundary conditions for a gear wheel and beam respectively.

(a)   (b)

FIGURE 10.1   Boundary conditions in modal testing: (a) free, (b) fixed

From a practical point of view, the actual boundary conditions chosen for a test may depend on the weight and size of the structure. For instance, it is not possible to support a large structure weighing a few tonnes in a free test state. Also it may not be convenient to fix certain structures in the clamped condition. There are also situations requiring both test conditions as in the case of a space vehicle. It is desirable to test this structure in the free condition to simulate its operating environment in space. However, it is also required to do the test in the grounded condition to study dynamic behaviour in launch environment. Sometimes the boundary conditions can lead to uncertainties and non-linear behaviour of the test structures and therefore fixing conditions should be carefully considered. Besides, it is to be remembered that the clamping mechanism also adds damping to the structure due to movement in the fixture joints.

10.2.1.1   *Minimizing exciter/test structure interaction*
The shaker physically interacts with the structure through the force transducer, altering the system dynamics. In the case of lightweight structures, the load cell adapter may lead to considerable mass loading, causing the measured force to be greater than that actually applied. Since this extra mass is between the structure and the load cell, the latter senses it as a part of the structure. The shaker should transmit forces in line with the main axis of the load cell. However, it is possible that there is rotation about the other two axes. To minimize this problem, the exciter is connected to the load cell through a slender rod called the stinger, which allows the load to move freely in other directions. This stinger is essentially a rod with a strong axial stiffness like a truss, but weak bending and shear stiffness. Figure 4.2 of Chapter 4 shows the positioning of the stinger in modal testing. It is also desirable to isolate the main body of the exciter from the structure to prevent reaction forces from being transmitted back to the structure. There are a few ways in which this could be done:

(i) by supporting the shaker on a mechanically isolated foundation.
(ii) by mounting the shaker on a solid floor with the structure suspended from above.
(iii) by suspending the shaker and using an inertial mass attached to the shaker to generate a measurable force.

The concept of impedance is useful in understanding the interaction between the shaker and the test specimen. Mechanical impedance, in simple terms, is defined as the force required to produce a desired motion (velocity to be precise). A force applied to a test structure by means of an exciter drops close to zero at the natural frequency when the vibration amplitude becomes high. Such reduction in force is due to the fact that, at resonance, the test structure becomes highly compliant and vibrates with large amplitudes, even with very low input forces. The exciter then uses all the available energy to accelerate its own mechanical parts, leaving very little force for driving the test structure. The signal level of the force may, in fact, drop to the noise floor in the instrumentation. Complex systems, with several natural frequencies, load the vibration exciter each time the excitation frequency coincides with a natural frequency. The lighter the mass of the moving elements in the exciter, the less this problem will be. An ideal exciter is one with infinite mechanical impedance; however, this is not the case in practice. This problem may also be alleviated by effective mechanical de-coupling of the structure under test and the modal exciter, through appropriate test fixtures as discussed above, thus minimizing the change in mechanical impedance caused by the exciter. This is required so that when modal parameters are later extracted from the vibration information around resonance, force drop-offs are reduced, thereby increasing the accuracy of the FRFs.

If the input is not a force applied to the mass being tested, but instead is a motion (displacement) applied to the base, the result is that instead of the zero force required to maintain motion, an oscillating force of specific magnitude is needed to maintain the motion of the test mass. Thus base excitation of a system exhibits high impedance. In the limit, if the test mass approaches infinity, the impedance also approaches infinity.

It is clear that the concept of infinite impedance is inherent in any modal test program, so that the dynamic input to the specimen is not altered by the specimen itself. The design of the fixture is directly tied to the need for a shaker with infinite impedance. The fixture must be as stiff as possible, so that it is not deflected by the force and transfers motion from the exciter to the specimen with zero distortion at all amplitudes and frequencies. Generally, however, a fixture is never stiff enough to result in the natural frequency of the test system being above the range of test frequencies. Instead, the natural frequency generally lies within the test frequency range. Test specifications typically require that the dynamic input to a test specimen be kept at a constant level, regardless of fixture loading. Electromagnetic and electrohydraulic vibration exciters handle this problem with servo controllers. They sense any deviation from the specified vibration level and vary the gain of the amplifiers to minimize the deviation. With other exciters, the modal analyst has to take all precautions. To reduce impedance mismatch between the shaker and the structure, the following methods may be adopted:

(i) using shakers with different coil sizes depending on the structure.
(ii) using a shaker with a constant current amplifier and positioning the shaker at a point with large effective mass.

10.2.2 Choice of Exciters/Shakers

There are many exciters which may be used for experimental modal analysis. They include hydraulic shakers, electrodynamics shakers, unbalance mass exciters, inertial shakers, etc.; with such exciters the specimen is fixed onto the shaker table and tested for identifying resonance behaviour. In the case of unbalance mass exciters, the force produced is proportional to the square of the rotational speed of the unbalance mass. Since these exciters are capable of producing large forces, they are used to test structures like bridges, masts, towers, dams, foundations, etc., the shaker itself offering negligible mass loading in these cases. For imparting forces to lightweight structures, or for modal analysis of machines at site, the impact hammer is a very useful device. The only difficulties involved in a hand-controlled hammer are occurrences of multiple impacts and poor repeatability of impact conditions, namely the force amplitude and direction. These problems can be got rid of to a large extent by using electric hammers. An interesting development in the field of exciters is that of inertial shakers (reaction mass shakers) and non-contact electromagnetic exciters for rotating machines. The latter are particularly convenient for exciting lightweight and rotating structures. They do not have the problems associated with impact testing using a hammer. They can be used to transmit forces to rotating disks, say, for vibration control or for estimation of the stiffness and damping of bearings. These exciters are also useful for evaluating the dynamic response of impellers, circular saws, etc. Inertial shakers are useful for obtaining the response of a structure in field conditions within the shortest test time possible, without the use of any additional test fixtures. They can very conveniently be used for modal testing of a ship hull or automobile body which requires fixing the exciter with arbitrary orientations. Chapter 4 deals exhaustively with excitation mechanisms with Tables 4.1 and 4.3 summarizing salient features of intrusive and non-intrusive excitation techniques, respectively.

10.2.2.1 *Problems in the measurement of excitation force*

Measurement of the input force is one of the most crucial aspects of a successful modal test. Errors originating here will not only affect each and every FRF obtained, but will also typically be 'hidden', and unlike those in the measured response, be difficult to detect visually. Obtaining accurate and reliable force measurements is therefore very important in order to ensure that the extracted modal parameters are correct. With a conventional force transducer used in a modal test setup, the measured force will not be directly proportional to and in phase with the excitation drive signal from the signal generator due to the masses of the moving parts of the exciter, as well as the mass of the force transducer, which are not negligible. If, however, the moving parts of the excitation drive-train may be considered to be part of the structure being tested, then the force can be indirectly computed by sensing the current to the electrodynamic shaker as discussed in Section 4.2.1.1. The forces measured may also be erroneous due to the asymmetric construction of the force transducer and the influence of lateral degrees of freedom of motion at the contact point of the transducer and test structure. Mass cancellation circuits may also be used for this.

10.2.3 Impact Testing and Difficulties

Impact testing is probably the most popular excitation technique used in experimental modal analysis. Its popularity is due largely to its ease of use, ability to adapt to difficult testing environments, and self windowing potential for minimal signal processing errors. All these aspects have been described in Section 4.3.4. All modal parameter extraction methods used for obtaining FRFs with shaker excitation have been adapted to impact testing also. Impacts may be imparted using an ordinary instrumented

hammer or a modal punch or an electric hammer. The challenge in impact testing of large structures is in supplying an impact with sufficient energy to uniformly excite the test structure. Since most of the dynamic characteristics of interest in large structures are at low frequencies, a means of restricting the input force spectrum is an important consideration. Electrodynamic shakers can also be used to deliver impulsive loads to test objects by feeding them with voltage pulses. This method has conventionally been used for shock qualification of components; however, there is no reason why such shocks cannot be used in modal testing.

One major problem with impact testing is generation of multiple impacts while hitting. This can be eliminated by the use of a force window which is nothing but a rectangular window. It suppresses the second and subsequent pulses, if any, in the measured force that is used in the FRF computation. However, the response to the suppressed pulses exists in the output spectrum used in the FRF calculation, as a result of which the computed FRF does not represent the system accurately. Therefore, multiple impacts should be avoided. In practice it is usually easier to simply discard a record with a double hit from the ensemble of records and this is done by most modal software packages. If a small number of averages are taken for an impact measurement, even one unsatisfactory record can adversely affect the averaged set of measurements.

Another problem in impact testing involves use on non-linear structures. If impact forces are high, they can overdrive the system and aggravate its non-linear response. Every effort should be made to linearise the system when making FRF measurements using impact testing for estimation of modal parameters, since all the theory behind FRF computation and modal parameter extraction is based on the assumption of a linear time-invariant system. Non-linearities should be eliminated from the test system as far as possible. Static preloads applied with soft springs can tighten clearances and constrain the system into a comparatively linear state. The low stiffness of preload in series with the stiffness of the structure causes a negligible effect on the system dynamics.

Yet another problem is impacting at locations which are not accessible with a conventional impact hammer. A punch impactor may be used under such circumstances to precisely locate impacts, to minimize the variation of the impact location and to impact in a skewed direction (non-orthogonal to the coordinates). The punch impactor should be fitted with a hard tip, typically a metal or plastic tip. It should be calibrated in the configuration in which it is to be tested since the sensitivity depends on the tip-structure combination.

10.2.4   Sensing Techniques

In Chapter 3, we discussed various types of transducers which may be used in vibration testing. Displacement, velocity or acceleration transducers should be used, depending on whether we require receptance, mobility or acceleration FRFs (described in Section 10.3.1). Mass loading due to the sensor could be a severe problem for highly flexible structures, especially if the sensor mass is considerable compared to the effective mass of a particular mode of the structure. Mass loading can also be reduced by choosing lightweight transducers or even better, non-intrusive sensors which cause minimal mass loading on the structure. Details of non-intrusive and intrusive sensors are given in Section 3.10. There are many transducers which are suitable for non-contact vibration response measurement and many of these have the desirable small target footprints. In modal analysis the use of miniature accelerometers and strain gauges for sensing the response is popular; they are not truly non-intrusive, but they hardly alter the system dynamics. Many of the non-contact response measurement techniques allowing direct

interpretation of experimental data are optical in nature and have very good resolution, dynamic range and flexibility necessary for modal analysis applications. Some of these transducers are fibre-optic sensors, laser holography, laser Doppler velocimetry (LDV), laser triangulation techniques, and line scan cameras. There are other non-optical techniques using proximity probes of both inductive and capacitive type and acoustic sensors. A lot of information is given in Chapter 3 regarding a whole range of transducers of the contact type, which may also be used for modal analysis after taking into account their salient features.

Considerations regarding the choice, location and orientation of transducer and transducer mounting have been discussed in detail in Section 8.3. Local conditions of a structure may drastically vary, influencing the local stiffness at the transducer location and hence choice of location is very important. The transducer should not be mounted at or close to a node of any structural mode of interest. For point FRF measurements, the force and the response are required to be measured at the same point. This may be done using an impedance head, which has both force gauge and the accelerometer combined in a single housing. It is important to note the direction of the force and response, and when at a location both translational and rotational DOF are present, the errors in the mobility ratio can be considerable. Most of the present modal analysis techniques use translational DOF. Bi- and triaxial transducers are also available to perform detailed modal analysis.

Due to mass loading effects of a contact type of transducer, the vibration measured turns out to be different from what would be produced in the absence of loading. Fortunately, technology exists for performing effective non-contact modal analysis of such structures, but careful consideration must be given to the entire test system compatibility. For example, consider the read/write head and assembly on a typical computer hard disk, the dynamic response of which is critical to the performance of the system. The mass of the entire read/write mechanism being quite small (approximately 30–40 gm), the mounting of either an excitation or response transducer can alter the system dynamics. The mass loading of even the lightest accelerometer (less than a gram) can significantly alter the system response.

10.2.5 Selection of Excitation Signals for Modal Testing

The choice of excitation system, as well as excitation function, can make the difference between a good measurement and a bad one. Once the excitation system, i.e. the shaker or excitation mechanism that best suits the application has been decided on, the next step in the measurement process involves selecting an excitation function such as sine sweep, random, transient, etc. The excitation signal generally depends on the type of test to be conducted. Though the level of non-linearities can be measured and characterized effectively with sine sweeps, a random function is most suitable to estimate the best linear model of a non-linear system. The amount of damping and the density of the modes within the structure can also dictate the use of specific excitation functions. The exciter control electronics must therefore have a signal generator and a gain control, dependent on the feedback from the test object. Computer-aided vibration control systems which are simple to operate and interactive are commercially available for shaker control. They utilize highly developed and powerful digital signal processing hardware and/or software. The excitation signal can be one of the following:

  (i) stepped sinusoidal
 (ii) slow sinusoidal sweep
(iii) periodic

   (iv) true random
- sine on random
- random on random
   (v) pseudo random
   (vi) transient
- chirp random
- sine chirp
- burst chirp
- external pulse
- shock

Most of the present day analysers come built in with many/all the options mentioned above. Stepped and swept sinusoidal, as well as periodic and random excitation signals, can also be got from an oscillator or signal/function generator. The latter signal can also be got from a random noise generator.

10.2.5.1   *Stepped sine excitation*   In this method, the signal given to the exciter is a discrete sinusoidal voltage or current with a fixed amplitude and frequency. In order to encompass the entire frequency range of interest, the command signal frequency is stepped from one discrete value to another in such a way as to provide the necessary frequency resolution in the FRF plot. Care should be taken to ensure that the excitation frequency is not changed abruptly and that sufficient time is given for the system to settle down. Especially if one is in the vicinity of resonance, or if one is measuring lightly damped or closely spaced modes, care should be taken to avoid transient effects by delaying each measurement and ensuring that steady state conditions have been attained before the readings are taken.

   The major advantage with stepped sine excitation is that it gives the vibration engineer flexibility in choosing the frequency resolution. Near resonance, tests can be conducted at a larger number of frequencies and away from resonance, excitations can be given at a few frequencies only. It is especially important to have closely spaced points around resonances if one is interested in measuring modal damping. Considering that excitation is always imparted at a single well-known frequency, it is possible to use simple measuring techniques for obtaining FRFs and doing away with the need for sophisticated equipment like FFT analysers. It is possible to optimize the measurement process, irrespective of whether one is doing the whole process manually or is using computer control. Though sine testing gives the best signal-to-noise ratio (SNR), it is extremely slow for broadband measurements.

   Stepped sine excitation is ideally suited for the evaluation of non-linearities since it is very easy to control amplitude, frequency and phase of the excitations. The presence of non-linearities in the system may be found out by analysing the dependence of the response level on the excitation level and the manifestation of higher order harmonic components in the response to sinusoidal excitation of a single frequency. This testing technique is typically used for large, complex structures and for cases where there is a high level of background noise present. For large structures requiring a multiple excitation/response measurement, one has to use a controller which defines a phase/amplitude relationship between the different shakers. This relationship has to be maintained throughout the entire sweep.

10.2.5.2   *Slow sine sweep*   This involves the use of a sweep oscillator to provide a sinusoidal command signal, the frequency of which is varied continuously and slowly throughout the frequency range of interest. Swept sine capability allows conventional sine testing of non-linear structures where high energy at a single frequency is necessary. Figure 10.2 shows a typical sine sweep signal.
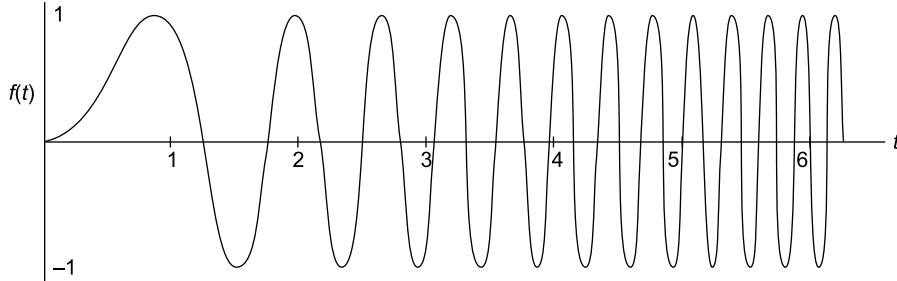
FIGURE 10.2   Sine sweep

As in the case of stepped sine excitation, it is imperative in this case also to ensure that the frequency sweep is done sufficiently slowly to enable the system to settle down to steady-state conditions before measurements are made. If the sweep rate is too high, then the FRF gets distorted with the true resonance peak getting shifted to the right (higher frequency) with increasing sweep rate. A good way to check whether a sweep rate is alright, is to measure twice, once sweeping upwards, and the second time downwards through the frequency range. It can be said that the sweep rate is correct, if both the sweeps give rise to the same FRF. It is possible to prescribe an optimum sweep rate for a given structure, taking into account its damping levels. A good practical approach to guarantee that we can very closely approach the desired condition is by using a logarithmic sweep rate. The ISO standards on 'methods for the experimental determination of mechanical mobility' prescribe the maximum allowable linear and logarithmic sweep rates through a resonance. These are given by the formulae below.

Linear sweep

$$S_{max} < 54(f_r)^2(\eta_r)^2 \text{ Hz/min} \tag{10.1a}$$

or
$$S_{max} < 216(f_r)^2(\zeta_r)^2 \text{ Hz/min} \tag{10.1b}$$

Logarithmic sweep

$$S_{max} < 78(f_r)^2(\eta_r)^2 \text{ octaves/min} \tag{10.2a}$$

or
$$S_{max} < 310(f_r)^2(\zeta_r)^2 \text{ octaves/min} \tag{10.2b}$$

Here S_{max} denotes the maximum allowable sweep rate, f_r the resonant frequency of the mode under study, ζ_r its damping ratio and η_r the corresponding loss factor.

10.2.5.3   Periodic excitation   A complex periodic input signal is sometimes used for obtaining FRFs. Such a signal contains not one, but a superposition of several sinusoidal signals simultaneously. With such an excitation signal, a spectrum analyser can extract the response to each of these sinusoidal components simultaneously. Two types of periodic signals are typically used; both types are generated within most analysers in order to ensure perfect synchronization with the analysis. One is a synthesized signal with various sinusoidal components having ordered amplitude, frequency and phase relationships, as in a square wave. The other is of the pseudo random type, which is discussed later.

10.2.5.4   True random excitation   True random signal is the most popular signal for shaker excitation since measurement time is minimal. The choice of this excitation signal can well be imposed by the

dynamics of the test structure. This signal may be obtained from signal generators or random noise generators, which are commercially available. Several types of noise signals are typically used, namely uniform white noise, Gaussian white noise, pink noise, and periodic random noise. Ideal white noise has equal power per unit bandwidth, resulting in a flat power spectral density across the frequency range of interest. Pink noise, also called weighted white noise, on the other hand contains equal energy per measurement band and has an octave or one-third octave band energy level which is constant with frequency band. Thus the lower frequencies have a higher energy level. In practical measurements involving white noise, an infinite number of samples are required for achieving a flat power spectral density. Hence, when making measurements with white noise excitation, the power spectra are usually averaged; the larger the number of averages, the flatter is the resulting power spectrum.

The terms uniform and Gaussian refer to the nature of the probability density function (PDF) of the noise. For uniform white noise, the PDF curve is flat, while for Gaussian white noise, the PDF is bell-shaped. Figure 10.3 shows the time record of white noise. Figures 7.5 and 7.6(a) show the PDFs of uniform white noise and Gaussian white noise, respectively. Gaussian white noise may be obtained by passing uniform white noise through a linear filter.
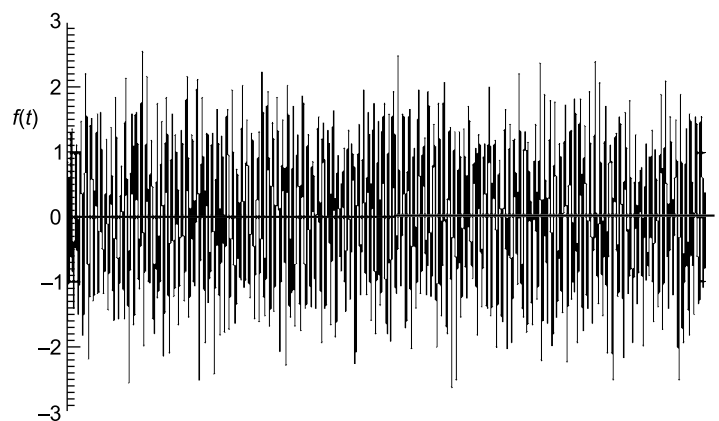
FIGURE 10.3   Time record of white noise

This signal has the advantage that both its peak to RMS ratio (crest factor) and RMS value can be controlled. It also has a reasonably good SNR. A random function is most suitable for obtaining the best linear model of a non-linear system. Typically in testing, the true random excitation signal is sent out during a time period T to the shaker by the controller and the structure's response is measured synchronously. During the next time period T, a different random excitation record (with the same RMS level and crest factor however) is sent out. Thus the excitation level at a given frequency is different during each period T, allowing the test engineer to get the best linear approximation of a non-linear system by using an averaging process. Though averaging helps in determining the best linear model and in reducing uncorrelated noise, it increases the total analysis time. This technique also has the disadvantage that there is leakage in the analysis window due to the non-periodic nature of signal in this window. This problem is overcome by using pseudo-random excitation which is periodic in the analyser window considered. Band-limited random noise allows exciter energy to be concentrated in a small frequency range for improved SNR. We have already seen the time history of true random noise (Fig. 10.3). Figure 10.4 shows the PSDs of true random (white noise) and band-limited white noise signals.
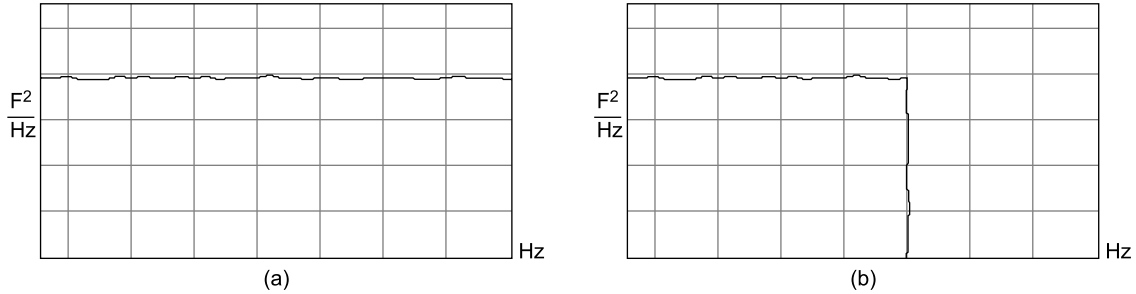
FIGURE 10.4   PSD of random noise: (a) white noise, (b) band-limited white noise

10.2.5.5   *Pseudo-random excitation*   Pseudo- or periodic-random noise (PRN) consists of a summation of all possible sinusoids with frequencies that can be represented as a fundamental and higher harmonics in the number of samples being processed. PRN is a superposition of sinusoidal signals with equal energy at all frequencies and with random phases. The major advantage of this type of excitation signal is its exact periodicity in the analyser bandwidth, resulting in zero leakage errors. PRN is thus self-windowing and there is no need to do windowing before spectral analysis. The main difference between PRN and white noise is that PRN does not have a continuous PSD curve, as does white noise, but has energy only at discrete frequencies that correspond to harmonics of a fundamental frequency. This fundamental frequency is equal to the sampling frequency divided by the number of samples. The most favourable feature of PRN is that the FRF of a linear system may be computed with just one time record, instead of averaging a number of FRFs over several time records, as is done for non-periodic random noise sources. Figure 10.5 shows the PSD of PRN.
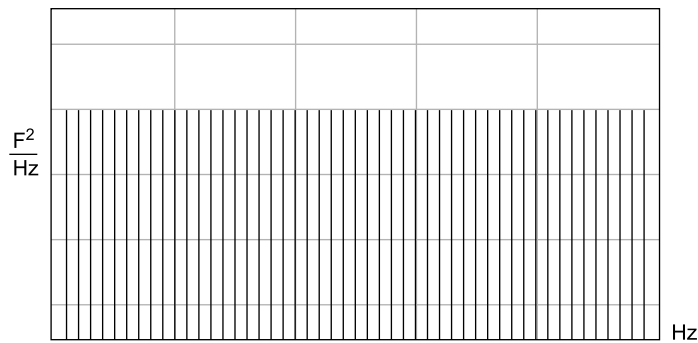
FIGURE 10.5   PSD of periodic random noise

10.2.5.6   *Transient excitation*   Transient excitation is different from steady state excitation in that it poses the constraint that the response signal should have died by the end of the sampling window. This type of excitation is essentially one of short duration and requires that lightly damped structures have a rather long sampling window, limiting the frequency range that can be studied. With transient excitation, as in the case of random excitation, it is the practice to make a large number of repeated measurements under nominally identical conditions, and then to do averaging, resulting in averaged FRF estimates. Though a large number of averages tremendously improves the SNR, it may diminish any advantage of the time taken for analysis, which is a potential attraction of the transient excitation technique. Most

analysers provide the user with a transient excitation signal called the chirp signal which is nothing but a rapid swept sine excitation from the lowest to the highest frequency in the spectrum over the relatively short sampling time window. This signal is so called because it sounds like the chirping of a bird. Chirp signals encompass transient signals such as chirp random, sine chirp and burst chirp. A burst chirp signal is similar to the chirp, except that it is turned off prior to the end of the sampling window. Chirp signals, like all the previously mentioned excitation signals, can be fed to an attached shaker. We have already seen in Sections 4.3.4 and 10.2.3 that a transient input can be imparted to a system using an instrumented hammer also. Though all the transient signals mentioned above are very similar from an analysis point of view, the sine chirp signal has the advantage that it offers greater controllability in terms of the amplitude and frequency content of the input and allows larger vibration energy, while burst noise excitation provides decreased test time and reduced leakage errors. Figure 10.6 shows some typical transient signals.
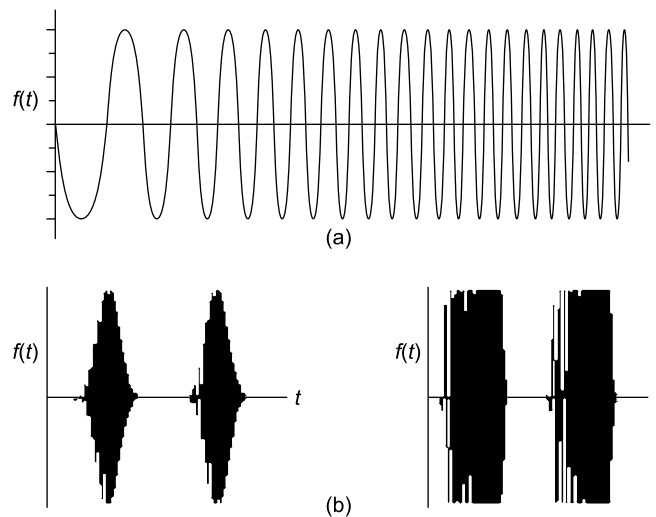
FIGURE 10.6   Transient signals for excitation: (a) sine chirp, (b) burst signals

10.2.5.7   *Burst random excitation*   A true random signal is used in burst random testing, but it is turned off well before the end of the finite length sampling window. This ensures that both the excitation and response signals are completely contained within the sampling window T and that the measured signals are periodic in the window, making them self windowing in nature and doing away with the need for windows. The advantage of the burst signal is that the structure has returned to rest before the next burst is given. The final FRF is obtained as the average with multiple burst excitations. This is the most appropriate excitation function for systems with closely coupled and or lightly damped modes, since it allows measurements in a leakage free manner. This excitation technique offers a very good compromise between measurement accuracy and speed. It is quite fast and like the true random, offers a very good linear approximation of a non-linear system. However, one cannot reduce the duration of the random signal too much, since a minimum amount of energy (or RMS level) is needed for effective excitation of the test structure. Figure 10.7 shows the zoomed waveform of a typical burst random signal. Electrodynamic shakers can also be used to deliver burst loads to test structures, though this method is used more for shock qualification of components. Table 10.1 gives a comparison of various excitation signals.
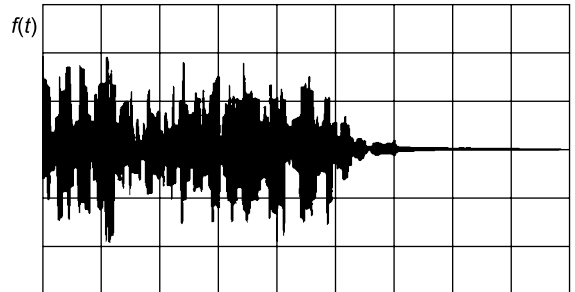
FIGURE 10.7 Zoomed version of burst random signal

TABLE 10.1 Comparison of various excitation signals

Excitation signal		Test time	SNR	Crest factor	Controlled frequency content	Controlled amplitude content	Removes distortion	Reduces leakage	Characterizes non-linearity
Sine steady state		Very long	Very high	High	Yes	Yes	No	No	Yes
True random		Good	Fair	Fair	Yes	No	Yes	No	No
Periodic in analyser window	Pseudo random	Very good	Fair	Fair	Yes	Yes	No	Yes	No
	Random	Fair	Fair	Fair	Yes	No	Yes	Yes	No
	Fast sine	Fair	High	High	Yes	Yes	No	Yes	Yes
Transient in analyser window	Impact	Very good	Low	Low	No	No	No	Yes	No
	Burst sine	Very good	High	High	Yes	Yes	No	Yes	Yes
	Burst random	Very good	Fair	Fair	Yes	No	Yes	Yes	No

10.2.6 Applications and Features of a Signal Generator

The signal generator or function generator is a device which generates the input to the power amplifier which drives the electrodynamic shaker/shakers and hence controls the nature of vibration energy input to the test structure, in terms of excitation waveform and signal level. Basically the signal generator should be able to deliver different kinds of signals, the choice of which have been seen in the preceding sections. Most generators also produce user definable waveforms with any signal shape such as triangle or square or arbitrary shape outputs. User defined waveforms can also be swept up and down in frequency similar to sine.

Typical technical specifications of a signal generator

Output

Impedance   :   50 Ω
Amplitude   :   0.1–10 V peak

Typical outputs

White noise	:	With flatness of ±1 dB up to 40 kHz and multiple uncorrelated channels
Band-limited noise	:	With filter roll off characteristics >20 dB/octave
Burst	:	All waveform types with continuous or single shot operation
Fixed frequency	:	Sine, square and triangular waveforms
Frequency range	:	1 Hz – 40 kHz
Frequency accuracy	:	0.2%
Frequency stability	:	0.2%
Total harmonic distortion	:	<2%

Swept frequency

Sweep types	:	Linear or logarithmic
Sweep rate	:	Programmable with pause, sweep up and sweep down

User defined waveform (arbitrary waveform)

Buffer size	:	8192 samples of 12-bit data

10.3 REPRESENTATION AND PROPERTIES OF FRF DATA OF SDOF AND MDOF SYSTEMS

Having seen the experimental aspects related to modal testing, we now move on to the theory behind modal analysis. In Section 2.4.5, the expressions for the FRF characterising the dynamic properties of a linear time invariant system have been derived. In the present section are discussed various mathematical forms of the FRF and ways of presenting or graphically displaying them. The plotting of FRF data is an involved process for the reason that there are three quantities, frequency along with two parts of the complex FRF, amplitude ratio and phase angle between the force and the response, and these cannot be fully displayed on a standard *x–y* graph. Thus, a full representation of an FRF in a single plot can only be done using a three-dimensional display as illustrated in Fig. 10.8. This is obviously not a convenient way of graphically representing the FRF. Any simple *x–y* graph can only show two of the three quantities and so there are different possibilities available, all of which are used from time to time. Some discussion on this has already been presented in Sections 8.13 and 8.14.
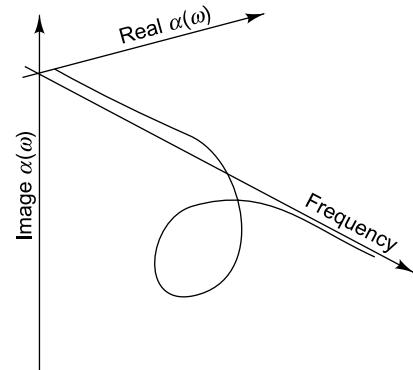
FIGURE 10.8 Three-dimensional plot of the receptance of an SDOF system

10.3.1 Graphical Display of FRF Data for SDOF Systems

The FRF of an SDOF system may be defined in terms of receptance, mobility and accelerance or inertance. These quantities are described for an undamped system in the following discussion. Let $x(t)$,

$v(t)$ and $a(t)$ represent the displacement, velocity and acceleration responses to a harmonic exciting force $f(t)$. Receptance is defined as the output displacement to unit input harmonic force and may be described by the expression below for an undamped system.

$$\alpha(\omega) = \frac{Xe^{i\omega t}}{Fe^{i\omega t}} = \frac{1}{k - \omega^2 m} \tag{10.3}$$

For sinusoidal vibration, we have the following simple harmonic relationships for the force and response:

$$f(t) = Fe^{i\omega t} \tag{10.4}$$

$$x(t) = Xe^{i\omega t} \tag{10.5}$$

$$v(t) = \dot{x}(t) = i\omega Xe^{i\omega t} \tag{10.6}$$

and $$a(t) = \ddot{x}(t) = -\omega^2 Xe^{i\omega t} \tag{10.7}$$

Mobility is defined as the output velocity to unit input force and has been obtained based on the above harmonic relationships.

$$Y(\omega) = \frac{Ve^{i\omega t}}{Fe^{i\omega t}} = i\omega \frac{x}{f} = i\omega\alpha(\omega) \tag{10.8a}$$

or $$|Y(\omega)| = \omega|\alpha(\omega)| \tag{10.8b}$$

and $$\theta_Y = \theta_\alpha - 90° \tag{10.8c}$$

Inertance or accelerance is the ratio of acceleration response to unit input harmonic force as shown below.

$$A(\omega) = \frac{Ae^{i\omega t}}{Fe^{i\omega t}} = -\omega^2\alpha(\omega) \tag{10.9a}$$

or $$|A(\omega)| = \omega^2|\alpha(\omega)| \tag{10.9b}$$

and $$\theta_A = \theta_\alpha - 180° \tag{10.9c}$$

Thus receptance, mobility and inertance are the main formats in which an FRF may be displayed, depending on whether displacement, velocity or acceleration measurements have been made. There do exist other possible ways of displaying FRFs in terms of functions which are inverses of the above as shown below

$$\text{Dynamic stiffness} = \frac{\text{Force}}{\text{Displacement}} \tag{10.10a}$$

$$\text{Mechanical impedance} = \frac{\text{Force}}{\text{Velocity}} \tag{10.10b}$$

$$\text{Apparent mass} = \frac{\text{Force}}{\text{Acceleration}} \tag{10.10c}$$

However, the inverse formats are used only in special cases, since they can give rise to considerable confusion and error, especially for MDOF systems. Table 10.2 summarizes all the FRF parameters, both standard and inverse, along with the various notations utilized.

TABLE **10.2** Definitions of frequency response functions

Response parameter	Standard FRF (response/force)	Inverse FRF (force/response)
Displacement	Receptance, admittance dynamic compliance, dynamic flexibility	Dynamic stiffness
Velocity	Mobility	Mechanical impedance
Acceleration	Inertance, accelerance	Apparent mass

Receptance for a damped SDOF system is given by the expression

$$\alpha(\omega) = \frac{Xe^{i\omega t}}{Fe^{i\omega t}} = \frac{1}{k - \omega^2 m + i\omega c} \tag{10.11a}$$

The angle of lag ϕ is given by

$$\tan\phi = \frac{\omega c}{k - m\omega^2} \tag{10.11b}$$

Figure 10.9 shows a typical receptance plot.
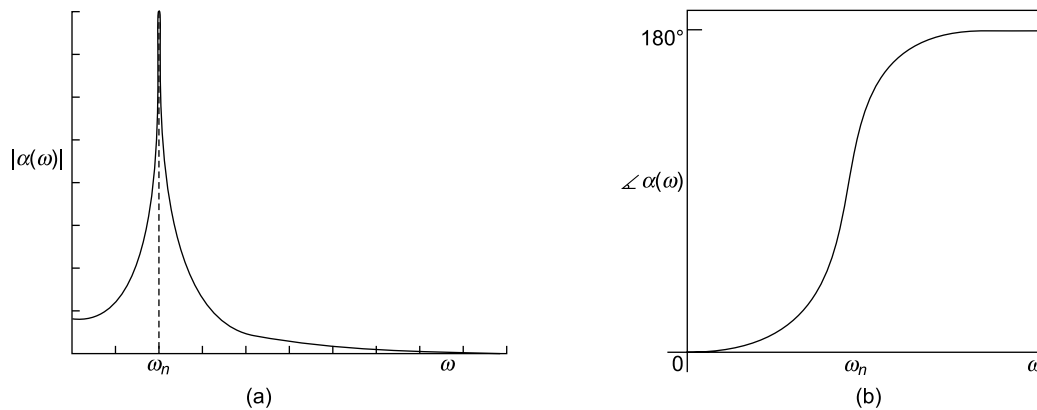
FIGURE **10.9**   Receptance plot: (a) magnitude, (b) phase

Many real structures have a frequency dependent damping called structural or hysteretic damping which has been discussed in Section 2.2.5.2. The receptance in this case is described by

$$\alpha(\omega) = \frac{1}{k - \omega^2 m + ih} \tag{10.12a}$$

$$\tan\phi = \frac{h}{k - m\omega^2} \tag{10.12b}$$

The three most commonly used graphical presentations are as follows:
 (i) Bodé plots, consisting of two graphs: Modulus of FRF versus frequency and phase of FRF versus frequency.
 (ii) Real part of FRF versus frequency and imaginary part versus frequency (two plots) and
 (iii) Nyquist plot: a single plot consisting of real part versus imaginary part; this representation does not explicitly contain frequency information.

10.3.1.1 Bodé plots This style of displaying FRF data applies to both damped and undamped systems, while the other forms are applicable only to damped systems and are sensitive to the type of damping. Irrespective of the type of FRF used, a wide range of values result, typically of the order of 1 to 10000, both in the FRF magnitude, as well as frequency, and these may be compressed using a logarithmic scale, otherwise the smaller values would get submerged if linear scales are used. Plots for receptance, mobility and inertance of an undamped SDOF system are shown in Figs. 10.10(a)–(c), respectively with logarithmic scales for both the *x*- and *y*-axes. Data which would get displayed as curves with linear scales, become asymptotic to the straight mass and stiffness lines on a plot depicting log modulus versus log frequency. This is advantageous as it provides a simple means of checking the validity of a plot and also allows for easy identification of the mass and stiffness characteristics of the system under study.

Let us consider a rigid mass *m*, free in space upon which a harmonic force *f* is applied. The corresponding equation of motion is given by Newton's law

$$m\ddot{x} = f \tag{10.13}$$

Then, the receptance of this system would be

$$\alpha(\omega) = -\frac{1}{\omega^2 m} \tag{10.14a}$$

and the magnitude of $\alpha(\omega)$ in logarithmic terms would be

$$\log|\alpha(\omega)| = -\log(m) - 2\log(\omega) \tag{10.14b}$$

A log–log plot of the magnitude of $\alpha(\omega)$ as a function of frequency is therefore a straight line with a slope of –2. Similarly, the receptance of a simple isolated massless spring is given by

$$\alpha(\omega) = \frac{1}{k} \tag{10.15a}$$

Therefore,

$$\log|\alpha(\omega)| = -\log(k) \tag{10.15b}$$

Similarly the magnitudes and logarithms of magnitudes of the FRF function represented in terms of mobility and accelerance may be derived and are shown in Table 10.3.

TABLE 10.3 Frequency responses of mass and stiffness elements

FRF parameter	Mass	Stiffness
Receptance $\alpha(\omega)$	$-1/\omega^2 m$	$1/k$
$\log|\alpha(\omega)|$	$-\log(m) - 2\log(\omega)$	$-\log(k)$
Mobility $Y(\omega)$	$-i/\omega m$	$i\omega/k$
$\log|Y(\omega)|$	$-\log(m) - \log(\omega)$	$\log(\omega) - \log(k)$
Inertance $A(\omega)$	$1/m$	$-\omega^2/k$
$\log|A(\omega)|$	$-\log(m)$	$2\log(\omega) - \log(k)$

From Table 10.3 it is seen that mass and stiffness properties will always appear as straight lines on log modulus versus log frequency plots. Thus any typical FRF plot can be divided into the following three regions:

 (i) a low-frequency straight-line region characterized by the stiffness
 (ii) a high-frequency straight-line characterized by the mass

(iii) the resonant region with its abrupt magnitude and phase variations characterised by the damping in the system

Figure 10.10 shows the log–log FRF representations of an undamped SDOF system. From this figure and Table 10.3, it is clear that when $\omega \ll \omega_r$, the frequency response is approximately equal to an asymptote called the stiffness line and has a slope of 0, 1 or 2 for displacement, velocity and acceleration responses, respectively. When $\omega \gg \omega_r$, the frequency response is approximately equal to an asymptote called the mass line and has a slope of –2, –1 or 0 for displacement, velocity, or acceleration responses, respectively. The relevant FRF characteristics of simple mass and spring elements may be superimposed on these log–log plots as a grid of lines to enable finding out mass and stiffness of the system under consideration.
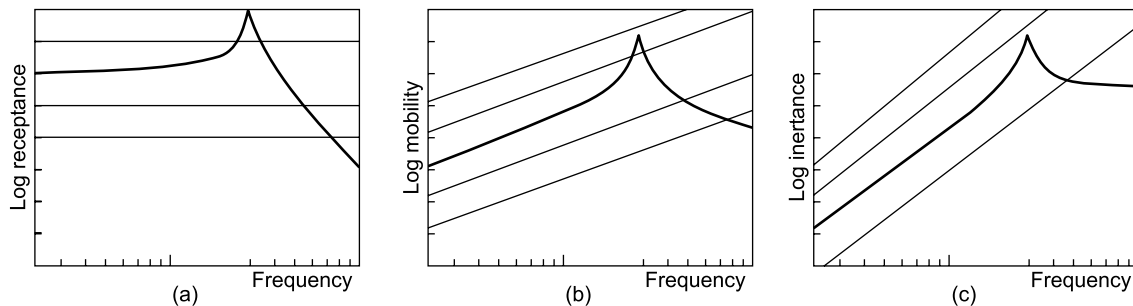
FIGURE 10.10   Log–log FRF representations of an undamped SDOF system: (a) log–log receptance, (b) log–log mobility, (c) log–log inertance

10.3.1.2   *Real part and imaginary part versus frequency*   Receptance FRF of a lightly damped SDOF system has been shown in Fig. 10.11 in the form of real and imaginary parts versus frequency and from these it can be seen that the phase change in the vicinity of the resonance frequency causes a sign change in one part, and a peak (maximum or minimum) value in the other part. It is not possible to use logarithmic scales in this case, as in the modulus plots, because both positive and negative values have to be shown. Due to this reason, this display format is not as widely used as the others, even though it is available as a standard feature in most analysers.
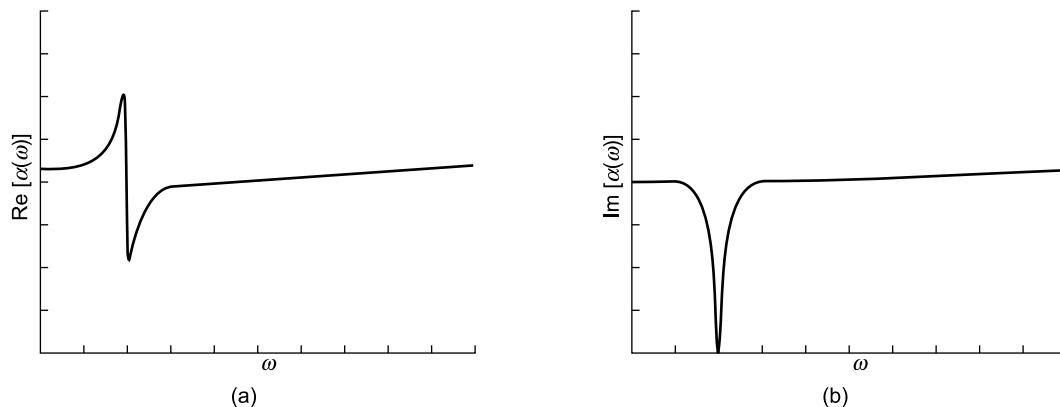
FIGURE 10.11   Receptance plot of real and imaginary parts of FRFs for a damped SDOF system

10.3.1.3   Nyquist plot   The Nyquist or Argand plane plot or vector response plot consists of a plot of real part of FRF versus the imaginary part and is widely used as it helps display the resonance region in detail. This display format consists of only a single plot and hence, the missing frequency information should be added at appropriate points on the curve, usually at regular frequency intervals. Figures 10.12(a) and (b) show the Nyquist plots of an SDOF system with viscous and structural damping, respectively. A striking feature of these plots is that only points very close to resonance are clearly identifiable while those away are very crowded. Indeed, it is this feature of favourably distorting the plot so as to focus on the resonance region that makes the Nyquist plot very attractive for modal analysis.
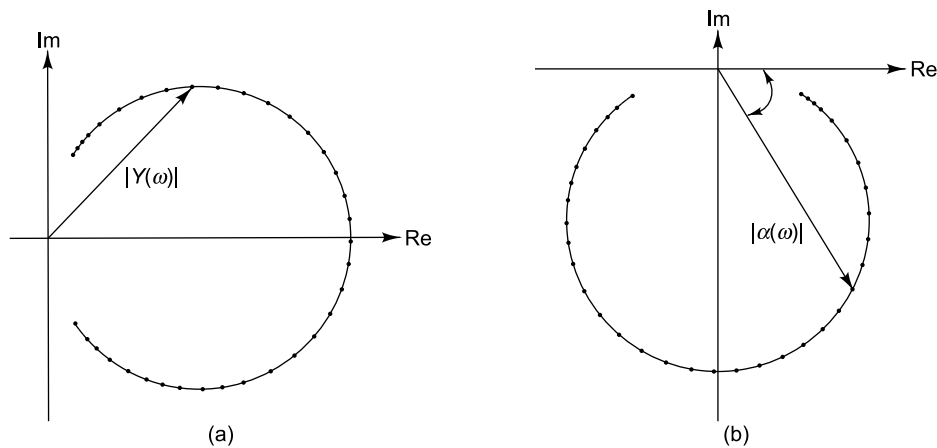
FIGURE 10.12   Nyquist plots for SDOF system: (a) mobility with viscous damping, (b) receptance with structural damping

It is clear from Fig. 10.12 that these FRFs resemble circles. The mobility $Y(\omega)$ with viscous damping traces out an exact circle, while it is the receptance $\alpha(\omega)$ that does so for hysteretic or structural damping. In the other FRFs, the deviation from a circular locus heavily depends on the amount of damping present and becomes negligible for small damping.

10.3.1.4   Properties of SDOF FRF plots   The basic geometric properties of three specific FRF plots of an SDOF system, which are very convenient forms of representation, are described below. They are as follows:

  (i) Log–log plot of modulus of mobility versus frequency (Fig. 10.10b)
 (ii) Nyquist mobility plot for viscous damping (Fig. 10.12a)
(iii) Nyquist receptance plot for hysteretic damping (Fig. 10.12b)

In a log–log plot of modulus of mobility, it can be seen that for light damping (less than 1%, say), the curve is more or less symmetrical about the resonance frequency. As far as the Nyquist plots are concerned, it is seen that two particular cases, namely, mobility for a viscously damped system and receptance for a hysteretically damped system, trace out exact circles as frequency ω is changed from 0 to ∞. This may be proved as shown below for hysteretic damping. Receptance with this damping is described by Eq. (10.12a). It follows from this equation that

$$\text{Re}(\alpha) = \frac{k - \omega^2 m}{\left[\left(k - \omega^2 m\right)^2 + \left(h^2\right)\right]} \tag{10.16a}$$

$$\text{Im}(\alpha) = \frac{h}{\left[\left(k - \omega^2 m\right)^2 + \left(h^2\right)\right]} \tag{10.16b}$$

It is clear that

$$(\text{Re})^2 + \left(\text{Im} + \frac{1}{2h}\right)^2 = \left(\frac{1}{2h}\right)^2 \tag{10.16c}$$

Thus the Nyquist plot of receptance for a hysteretically damped SDOF system traces a circle of radius $(1/2h)$ with centre at $(0, -1/2h)$ as depicted in Fig. 10.13. It is to be noted that Nyquist plots are not useful when damping is very low.
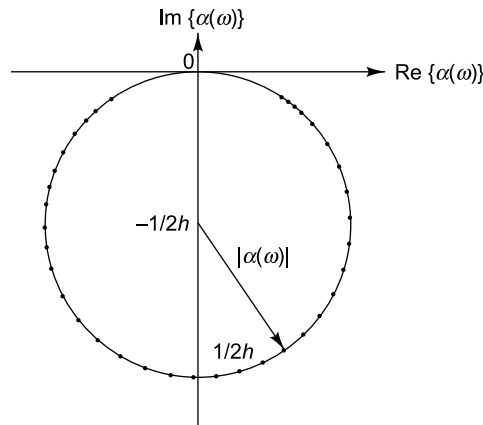
FIGURE 10.13   Nyquist plot of receptance

10.3.2   Characteristics and Presentation of MDOF FRF Data

The general form of the receptance FRF expression for an MDOF system is

$$\alpha(\omega) = \sum_{r=1}^{N} \frac{C_r}{\omega_r^2 - \omega^2 + iD_r} \tag{10.17}$$

where C_r is a constant representing the contribution to the response from the rth mode and D_r is a damping term. This is an extension of the FRF expression for the much simpler SDOF system studied earlier. It is important to understand the properties of this type of function to get an understanding of the modal analysis process. Hence, the various means of displaying the information it contains is explained in this section. There are three main alternatives, using displacement, velocity or acceleration response to produce receptance, mobility or inertance (or accelerance), respectively, as in the case of the SDOF system. The mobility and inertance matrices are related to the receptance matrix as shown below.

$$\left[Y(\omega) \right] = i\omega \left[\alpha(\omega) \right] \tag{10.18}$$

$$\left[A(\omega) \right] = i\omega \left[Y(\omega) \right] = -\omega^2 \left[\alpha(\omega) \right] \tag{10.19}$$

The FRFs of MDOF systems are more complex than their SDOF counterparts. A general receptance term $\alpha_{jk}(\omega)$ is given by

$$\alpha_{jk}(\omega) = \left(\frac{x_j}{f_k} \right); f_m = 0, m = 1, N; \neq k \tag{10.20}$$

These FRFs may also be expressed in the form of the inverses of the standard receptance, mobility and inertance as in the case of the SDOF system. However, deriving these inverse properties from the standard form is not straight forward. The velocity response of a structure to an exciting force may be got in terms of the mobility matrix using the equation

$$[v] = [Y(\omega)]\{f\} \tag{10.21}$$

The above equation may be rewritten in terms of the impedance matrix (inverse of mobility) as

$$\{f\} = [Z(\omega)][v] \tag{10.22}$$

It is to be noted however that a typical element in the mobility matrix is not related through a simple relationship to its counterpart in the impedance matrix, i.e.

$$Y_{jk}(\omega) = Y_{kj}(\omega) \neq \frac{1}{Z_{jk}(\omega)} \tag{10.23}$$

because

$$Y_{kj}(\omega) = \left(\frac{v_k}{f_j} \right); f_m = 0, m = 1, N; \neq j \tag{10.24}$$

and

$$Z_{jk}(\omega) = \left(\frac{f_j}{v_k} \right); v_m = 0, m = 1, N; \neq k \tag{10.25}$$

From a practical point of view, while is possible to measure the mobility FRF by applying excitation force at a single point, keeping the forces at all other points zero, it is almost impossible to measure impedance keeping all locations grounded except one. Thus, only the standard FRFs can be measured directly. Besides, it is to be noted that an element in the impedance matrix is not got as the inverse of the corresponding element in the mobility matrix, but from an inversion of the matrix itself.

Some common terms used in modal analysis parlance are driving point or simply point FRF and transfer FRF. A point receptance is obtained keeping the response and excitation coordinates identical, and is measured using an impedance head which has in-built force and vibration transducers. Point FRFs constitute the diagonal terms of the FRF matrix. A transfer mobility term is one for which the response and excitation coordinates are different and constitutes an off-diagonal term of the FRF matrix. Transfer mobilities are further classified as direct and cross-mobilities; in the former the response and

excitation co-ordinates are the same, say, both along *x*-axis, while in the latter, they are different (for example, one along *x*-axis and the other along *y*-axis).

10.3.2.1 *FRF representation of undamped MDOF system*   The receptance for an undamped MDOF system is given by the equation

$$\alpha_{jk}(\omega) = \sum_{r=1}^{N} \frac{_rA_{jk}}{\omega_r^2 - \omega^2} = \frac{_1A_{jk}}{\omega_1^2 - \omega^2} + \frac{_2A_{jk}}{\omega_2^2 - \omega^2} + \frac{_3A_{jk}}{\omega_3^2 - \omega^2} + \cdots \frac{_NA_{jk}}{\omega_N^2 - \omega^2} \qquad (10.26)$$

Here $_rA_{jk}$ is called the modal constant or modal participation factor for the *r*th mode with excitation at location *k* and response at location *j*; ω_1, ω_2, ω_3,..., ω_r, etc. are the natural frequencies of the 1st, 2nd, 3rd, ... *r*th modes. The participation factor indicates the amount each mode contributes to the total response at a particular point.

   The individual terms in the FRF may be plotted with log–log axes as described earlier. Figure 10.14 shows the receptance plot corresponding to Eq. (10.26); each term in the series expression for the FRF is shown as a separate curve in the figure and the sum of the individual modal contributions is also shown superimposed. It is however, difficult to predict the exact nature of the curve from the individual curves, since only the magnitude information is available in the log plots and the phase information is missing. It is thus unclear from the plot whether in each section of a curve corresponding to one term of Eq. (10.26), the FRF is actually positive or negative in sign and this decides the shape of the total FRF when terms are added together.
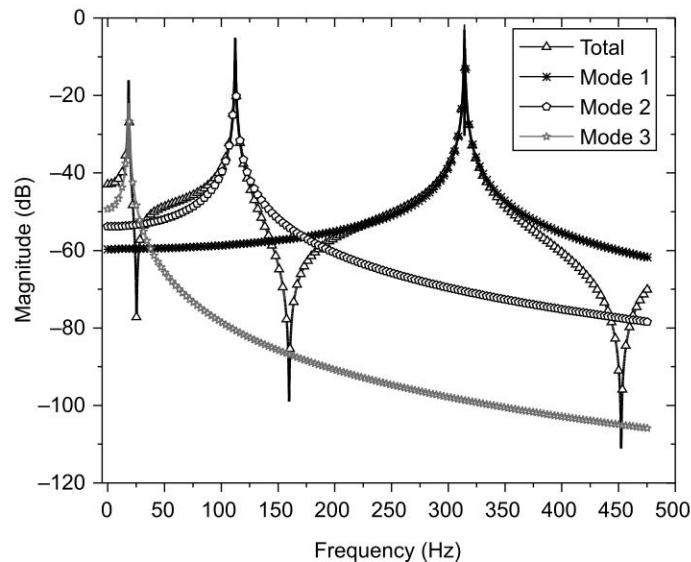
FIGURE 10.14   Typical receptance FRF for MDOF system showing contributions of individual modes and their superposition

Comments on the shape of FRF   The shape of the total FRF at any frequency depends on the signs of all the contributing terms at that frequency, which in turn depend on the signs of the modal constants. It is also interesting to see what determines the sign of the modal constant. It should be remembered that the

modal constant $_rA_{jk} = (_r\psi_j)(_r\psi_k)$, where $(_r\psi_j)$ is the displacement of the jth coordinate (response co-ordinate) in the rth mode and $(_r\psi_k)$ is the displacement of the kth coordinate (excitation co-ordinate) in the rth mode, or in other words, $_rA_{jk}$ is the product of two eigenvector elements, one at the response point and the other at the excitation point. In order to visualize eigenvectors, mode shapes corresponding to the first three modes of a cantilever beam are shown in Fig. 10.15(a). $(_1\psi_j)$ and $(_1\psi_k)$ are marked in the figure corresponding to mode 1 for two locations j and k. So also $(_2\psi_j)$ and $(_2\psi_k)$ are indicated for the second mode and $(_3\psi_j)$ and $(_3\psi_k)$ for the third mode. Based on these, the general trends to be expected in point and transfer FRFs are described below.

Point FRF   For a point mobility, the modal constant for every mode is positive since it is of the form $(_r\psi_j)^2$. This implies that for a point FRF, in the frequency region between two natural frequencies, the signs of the two terms contributing to these modes are negative, leading to a cancellation, and therefore there must compulsorily be an antiresonance (sharp dip) following every resonance (Fig. 10.15b).

Transfer FRF   In the case of a transfer FRF, at frequencies where all terms have the same sign and are thus additive, the magnitude of the total FRF curve is higher than the magnitude of any of the individual components, but since the FRF plot uses a logarithmic scale, the contribution of modes other than the predominant one at these frequencies is relatively small. Hence, the total FRF curve is only slightly higher than that for the predominant term. Thus in the immediate vicinity of a resonant frequency, the contribution of that term is much greater than that of the others, with the result that the total is, in effect almost the same as that one term. This implies that physically, the response of an MDOF system at one of its natural frequencies is totally dominated by that mode, the participation of the other modes being negligible.

If two consecutive modes have the same sign for the modal constants, then at some frequency between the natural frequencies of those two modes, there will be an antiresonance, because one term becomes positive and the other negative, cancelling each other out. The most important feature of the antiresonance is that the magnitude is very small and there is an associated phase change. If on the other hand, the modal constants have opposite signs, then these two adjacent terms in the series will have the same sign leading to addition, and hence there will not be an antiresonance, but just a minimum. In MDOF systems for which the transfer FRF has a large number of terms, the modal constant will sometimes be positive and sometimes negative, depending on the distance between the excitation and response locations. Thus, the FRF plot shows a mixture of antiresonances and minima (Fig. 10.15c). The further apart the two locations in question are, the higher is the possibility that the two eigenvector elements will alternate in sign as one progresses through the modes. Thus it can be expected that a transfer FRF between two locations far apart on the structure will exhibit less antiresonances, than one between two locations physically closer. A clear example of this is given in Fig. 10.15(d).

One important precaution to be taken while making measurements is that neither the excitation, nor the response coordinate, should coincide with a node for any of the modes, giving rise to the condition $_rA_{jk} = (_r\psi_j)(_r\psi_k) = 0$. If this happens, then the mode in question will not appear as a resonance on the FRF plot because the only contribution at or near $\omega = \omega_r$ will be from the off-resonant modes and this can be expected to be very small.

10.3.2.2 *Display of FRFs for damped MDOF systems*   The general appearance of the FRF (Bodé) plot for damped systems is similar to that for undamped systems, except that the resonances and

antiresonances are rounded by the inclusion of damping. Besides, the phase angles are no longer exactly 0° or 180°. Figures 10.15(b–d) shows mode shapes and FRFs obtained from a finite element (FE) model of a damped cantilever beam with 3 DOF, where the force location is near the fixed end. Rayleigh (proportional) damping with α = 0.01 and β = 0 has been assumed for the simulation.
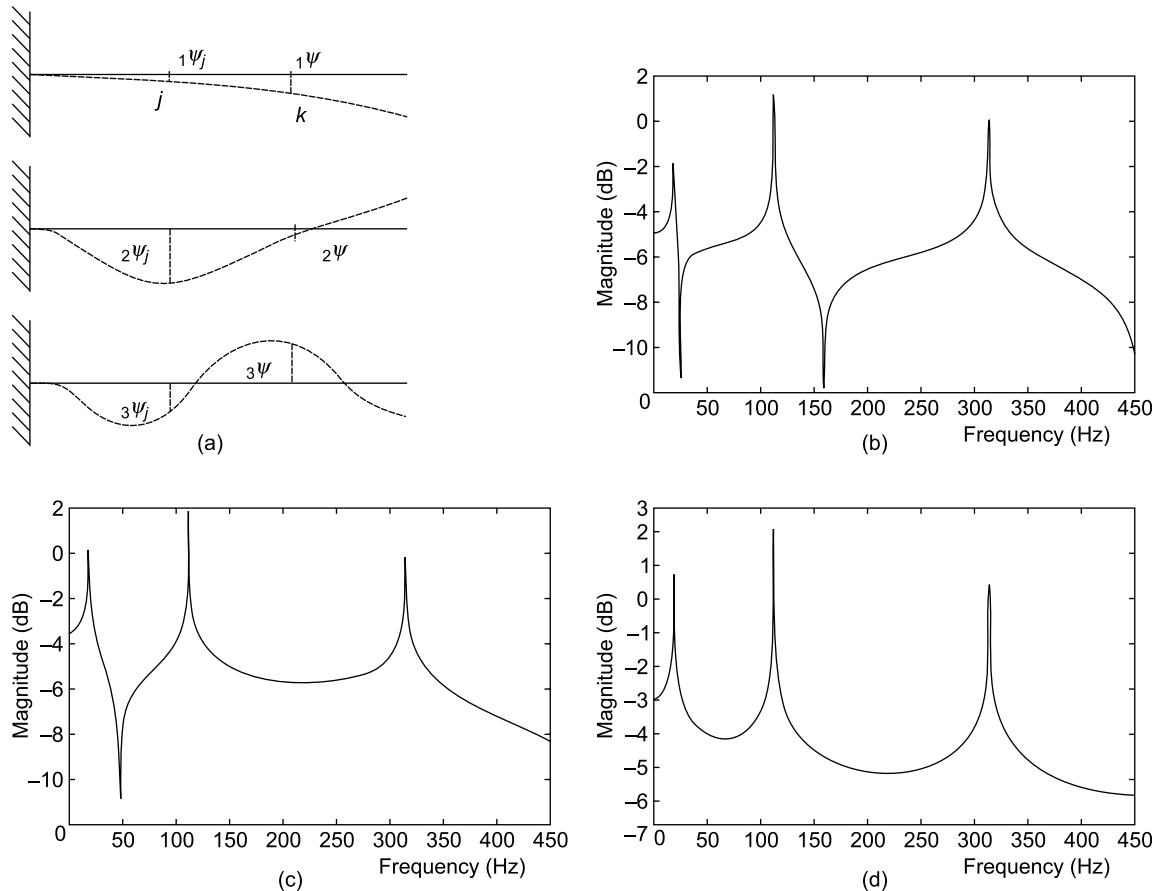
FIGURE 10.15 Point and transfer FRF receptance plots for a cantilever beam with damping: (a) mode shapes corresponding to first three modes, (b) point FRF at a point close to the fixed end of the beam, (c) cross-FRF at an intermediate point, (d) cross-FRF at the free end of the beam

The Nyquist or Argand diagram plot gives an exact circle as for the SDOF case. The mobility FRF is a true circle with viscous damping, while the receptance FRF is a true circle with hysteretic damping. There are typically as many lobes or circles in the FRF as the number of modes. With proportional damping, the modal constants are real quantities, being identical to those for the undamped version and will result in circles symmetric about the Im(FRF) axis. If the signs of the terms of the FRF are the same, then all the circles lie in the upper half or lower half, depending on the sign. If they have opposing signs, then some of the circles will lie in the upper half, while others will lie in the lower half of the complex plane. In the case of non-proportional damping, the individual modal circles are no longer 'upright', but are rotated by an amount depending on the phase of the complex modal constants; due to this, the

resonance points are no longer at the bottom (or top) of the corresponding circles. Figure 10.16 shows a Nyquist plot of the point FRF displayed in Fig. 10.15(a). Plots of the real and imaginary parts of the same as a function of frequency are shown in Fig. 10.17.
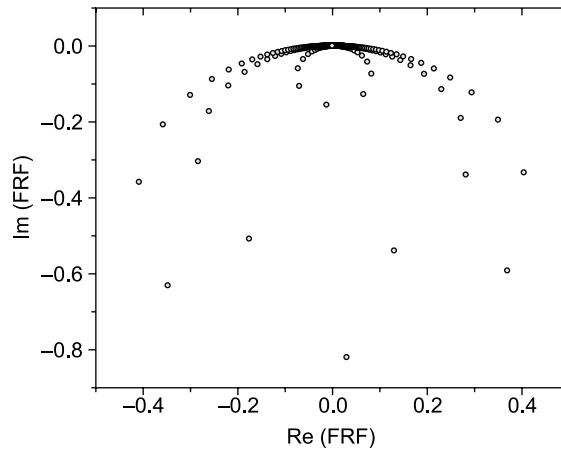
FIGURE 10.16   Nyquist receptance plot for damped 3-DOF system
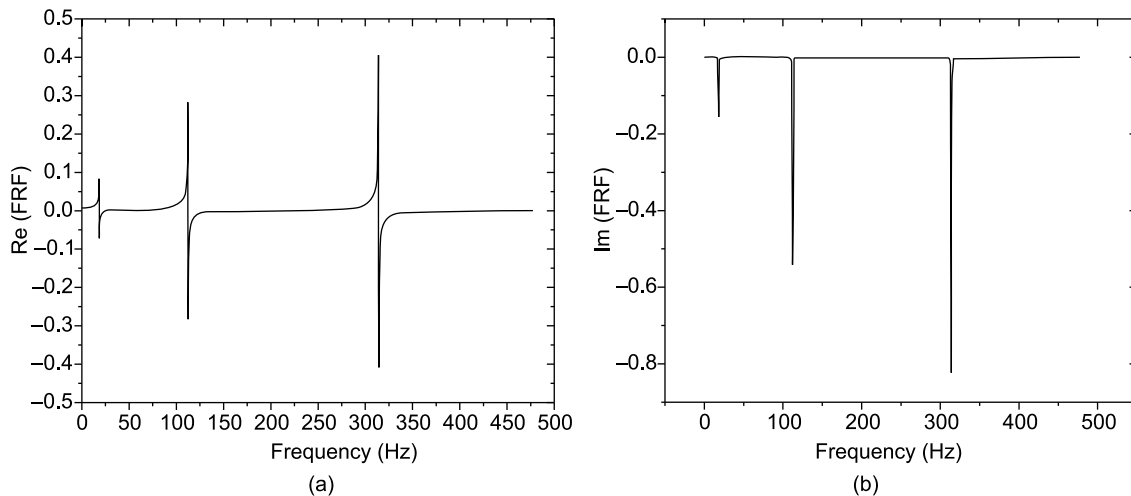
FIGURE 10.17   Real and imaginary receptance plots for damped 3-DOF system

10.4   OBTAINING FRFS WITH TRUE RANDOM EXCITATION

The measurement of an FRF as has been described involves measurement of the input force and output response as a function of frequency with all the phase information intact. In any modal testing involving single or multipoint excitation, the former is simpler, while the latter method has far-reaching implications due to the flexibility it offers in terms of the combination of position and orientation of the exciter and pickup. The input force may be imparted through a shaker or an impact hammer. FRF measurements

have to be made at a sufficiently large number of DOFs to get accurate mode shapes. All modal parameters are extracted from these FRFs and are therefore only as good as the FRFs, even if very elegant parameter estimation techniques are used. Modal parameters are subsequently extracted from the FRFs using any of the curve-fitting techniques available. It has to be ensured that driving-point FRFs (with the excitation and response measurements made at the same DOF) are included, in order to calibrate (scale) the modal model.

When excitation is imparted using an impact hammer, we have what is called roving impact excitation, the hammer being moved around the structure to excite every DOF, with the accelerometer response DOF being fixed and serving as a reference DOF. With shaker excitation, the excitation DOF is usually fixed, serving as the reference DOF. We then have roving response measurements with response transducer being moved around the structure. This will cause a varying mass loading of the structure and corrective action for this must be taken while extracting modal parameters. If instead of one roving accelerometer, however, there are many accelerometers simultaneously fixed at all the DOFs, the mass loading at each DOF will be the same in all the FRFs, providing better consistency of results and minimizing testing time. Figure 10.18(a) shows a typical modal analysis test setup with roving excitation and Fig. 10.18(b) shows the setup with fixed shaker and roving transducer.

10.4.1 Single and Multiple Input System Response Models

Both Single Input Single Output (SISO) and Single Input Multiple Output (SIMO) analyses involve excitation at a single point. Figure 10.19 shows the system configuration for the SIMO case.

Typically a random signal is fed to a shaker which excites the system. Alternately a broadband input may be imparted through an impact hammer. The input force and displacement/velocity/acceleration response signals are measured and fed to a real-time analyser which processes the time domain data to produce FRFs. The equation relating the input force and the output response of a linear system is given by

$$Y_i = H_{ij} X_j \qquad (10.27a)$$

where Y_i is the output spectrum at the ith DOF, X_j is the input spectrum at the jth DOF and H_{ij} is the FRF relating the two. Thus,

$$H_{ij} = \frac{Y_i}{X_j} \qquad (10.27b)$$

with the input forced to zero at all DOFs other than j. There are different classical estimators for estimating the optimal FRF H_{ij}. One of the most common ways of determining the FRF with single point excitation is through the use of the expression given below

$$S_{yy}(\omega) = \left| H(\omega) \right|^2 S_{xx}(\omega) \qquad (10.28)$$

Here $S_{yy}(\omega)$ and $S_{xx}(\omega)$ represent the auto power spectral densities of the response and excitation, respectively and $H(\omega)$ is the FRF linking the output response y of the structure to the input force x. Though this is a convenient expression, it has the drawback that it does not provide a complete description of the random vibration, since it lacks the phase information of the FRF. Alternate ways of describing the response are based on cross-correlation between the excitation and the response and are of the form given below
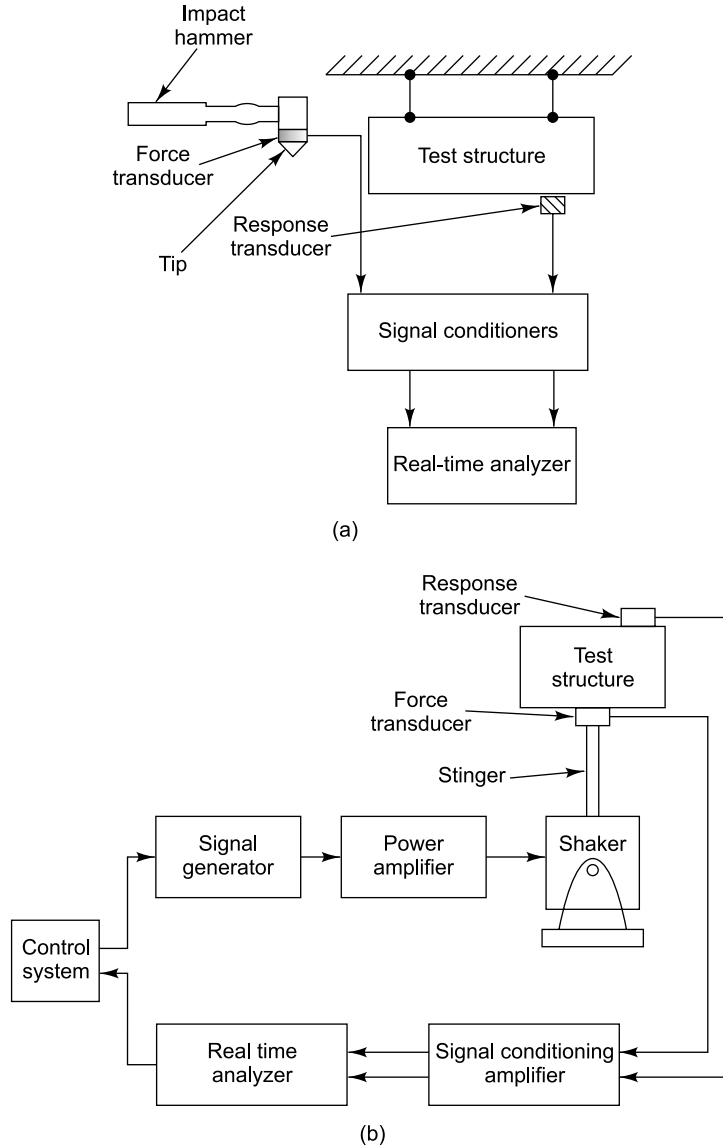
FIGURE 10.18 Typical modal analysis test arrangements: (a) roving hammer excitation, (b) fixed shaker excitation

$$S_{xy}(\omega) = H(\omega)S_{xx}(\omega) \tag{10.29a}$$

or
$$S_{yy}(\omega) = H(\omega)S_{yx}(\omega) \tag{10.29b}$$

In this expression $S_{yx}(\omega)$ is the cross-spectral density between the output and input signals, i.e. the complex conjugate of $S_{xy}(\omega)$. Most commercially available spectrum analysers have the facility to estimate these various quantities. However, in spite of using random signals (which typically require very lengthy

records for complete description) for excitation, their evaluation is based on finite length data and these lead to errors. Errors also creep in from the measured forces and responses. Hence, to understand how much error has crept in, the results may be cross-checked by obtaining different FRF estimates such as $H_1(\omega)$, $H_2(\omega)$, etc. using more than one equation. Details regarding various FRF estimators, and what is called coherence is discussed in the next section.
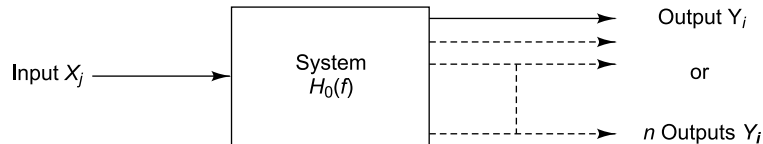
Figure 10.19 System configuration for SIMO

The analytical expressions extended to the case of simultaneous multiple excitations have not only the auto- and cross-spectral densities of and between the different inputs, but also the cross-spectral densities between inputs and outputs as discussed in Section 7.8.3.1. For uncorrelated inputs, the input-output expression would take the matrix form

$$\left[S_{xy}(\omega) \right] = \left[S_{xx}(\omega) \right] \left[H(\omega) \right] \tag{10.30}$$

where $[S_{xx}(\omega)]$ is the matrix of the auto- and cross-spectral densities of and between the different input DOFs, and $[S_{xy}(\omega)]$ is the matrix of the cross-spectral densities between inputs and outputs. The assumption behind this computation is that at each frequency of interest, all the forces are non-zero and that no pair of forces is fully correlated. Figure 10.20 shows the system configuration for the MIMO testing.
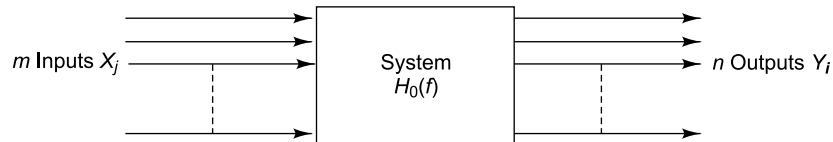
Figure 10.20 System configuration for MIMO testing

10.4.2 Issues Involved in Multiple-Input Multiple-Output (MIMO) Analysis

An important question to be considered in Multiple-Input Multiple-Output (MIMO) analysis is whether we need all the individual FRFs in the $[H]$ matrix or whether a few are sufficient to establish the modal model as shown in Figs. 10.21(a) and (b).

The answer to this question will tell us how many auto- and cross-spectral density terms need to be measured to estimate these FRFs. From a theoretical point of view, only one row or one column of the FRF matrix $[H]$ may suffice, but, from a practical point of view it may not be enough.

10.4.2.1 *Single reference modal analysis* The assumption behind single-reference modal analysis is that the selected reference DOF does not correspond to a nodal position for any of the modes. This DOF thus contains information about all the modes and ensures accurate modal parameter extraction. Hence, in such cases, FRF data are processed with data measured from only one reference DOF, i.e. measurement of only one row or one column of the $[H]$ matrix (Fig. 10.21) is sufficient. The identification

of a proper reference DOF may require some initial trial and error, or results of FE analysis if available, could be used. With a roving impact excitation (Fig. 10.21a), only one response DOF is needed, i.e. only one accelerometer point along a predetermined direction which serves as the reference DOF. In this case we need only determine one row of the FRF matrix. With a fixed shaker test (Fig. 10.21b), the location and direction of the shaker serve as the reference DOF and we need only one column of the FRF matrix to get the complete modal information required. A test in which multiple outputs are sensed in each measurement is called a Single-Input Multiple-Output (SIMO) test configuration.

$$
\begin{Bmatrix} Y_1 \\ \mathbf{Y_2} \\ Y_3 \\ . \\ Y_n \end{Bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} & \cdots & H_{1n} \\ \mathbf{H_{21}} & \mathbf{H_{22}} & \mathbf{H_{23}} & \cdots & \mathbf{H_{2n}} \\ H_{31} & H_{32} & H_{33} & \cdots & H_{3n} \\ . & . & . & \cdots & . \\ H_n & H_n & H_n & \cdots & H_{nn} \end{bmatrix} \begin{Bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \\ \mathbf{X_3} \\ . \\ \mathbf{X_n} \end{Bmatrix}
$$

(a)

$$
\begin{Bmatrix} \mathbf{Y_1} \\ \mathbf{Y_2} \\ \mathbf{Y_3} \\ . \\ \mathbf{Y_n} \end{Bmatrix} = \begin{bmatrix} H_{11} & \mathbf{H_{12}} & H_{13} & \cdots & H_{1n} \\ H_{21} & \mathbf{H_{22}} & H_{23} & \cdots & H_{2n} \\ H_{31} & \mathbf{H_{32}} & H_{33} & \cdots & H_{3n} \\ . & . & . & \cdots & . \\ H_{n1} & \mathbf{H_{n2}} & H_{n3} & \cdots & H_{nn} \end{bmatrix} \begin{Bmatrix} X_1 \\ \mathbf{X_2} \\ X_3 \\ . \\ X_n \end{Bmatrix}
$$

(b)
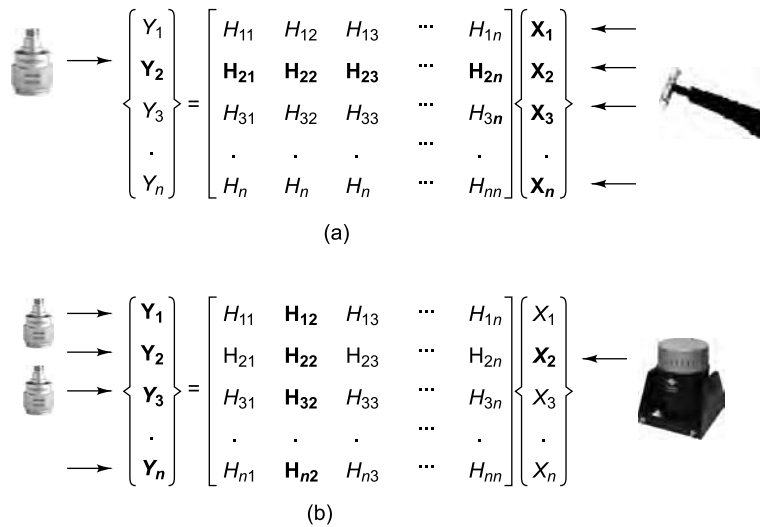
FIGURE 10.21   Experimentally obtaining the *H* matrix: (a) single fixed response measurement, roving impact excitation, (b) single fixed shaker excitation, roving response measurement

10.4.2.2   *Multireference modal analysis*

In cases in which it is not possible to find a proper reference DOF, measurements with more than one reference DOF will be required, i.e. measurements of more than one row or more than one column of the FRF matrix [*H*]. This happens if a structure exhibits local modes with predominant modal deflections at different parts of the structure and in complex structures made of different parts having different structural properties. It then becomes impossible to find a DOF where all the modes have a sufficiently large contribution as to enable proper modal extraction. Multireference testing also becomes necessary in symmetrical structures with more than one mode at the same frequency, requiring the measurement of as many rows or columns as the number of modes (or more) at the same frequency. Here it is also to be ensured that the mode shapes to be extracted look different in the reference DOFs. With measurements over many rows or columns of the FRF matrix, it becomes possible to enhance different modes by computation of linear combinations of these rows or columns. Thus the sum of the two rows or the two columns of a symmetrical structure measured with two reference DOFs at symmetrical locations may help in highlighting the symmetrical modes, while the difference may enhance the antisymmetrical modes. In the case of a roving impact test, as many rows can be obtained as the number of response DOFs measured. With a fixed shaker excitation, a multireference data set could be acquired simply by measuring one column at a time, by keeping the shaker first in one position (one reference DOF), then in another position and so on.

10.4.2.3 *Multiple-input multiple-output modal test*   Instead of obtaining a multireference data set by measuring one column at a time as suggested earlier, a better solution could be obtained through the use of multiple shakers to perform a 'Multiple-Input Multiple-Output (MIMO)' test. The Multiple Reference Impact Testing (MRIT) method or the Multiple Input Multiple Output method has been used successfully, especially in large and complex structures and structures with heavy damping such as aircraft, wind turbines, large industrial fans, bridges, etc. Reduced cost of electronics leading to multiple channel data acquisition systems, inexpensive transduction systems and multiple measurement parameter estimation algorithms have made multipoint excitation possible. The MIMO method uses controlled multiple shaker excitation or multipoint impact testing. Single point excitation does not work here for the reasons that the excitation force applied at a single point may not excite all parts of the structure owing to its large size and if the excitation level is increased in an effort to do so, the system might be driven into the non-linear range of vibration, resulting in deterioration in the estimation of the FRFs. Besides, results got from finite element analysis of large structures also require verification and for this, modal tests based on multipoint excitation (MIMO tests) come in handy. Many exciters (reference DOfs) are used to simultaneously excite the structure at multiple DOFs, resulting in simultaneous measurements of many columns of the FRF matrix. The response transducers must be roved around or if there are many transducers available, simultaneous measurements may be made at all the response DOFs. Another aspect to be looked into in MIMO studies is that the excitation forces applied to the structure are random and uncorrelated in order to satisfy the mathematical assumptions made in the analysis. Typically uncorrelated random excitation signals such as pure random, burst random or periodic random signals are used for excitation. Signal generators with multiple signal sources allow one to perform multiple point random modal tests. System non-linearities can also give rise to low coherence and their effects can be minimized by appropriate choice of excitation levels and sufficient averaging. Generally for MIMO testing 10–20 exciter positions or roving impact hammer positions and up to 100s or 1000s of response positions may be used depending on the requirement. It is required to have a multichannel FFT analyser or modal analysis system, with the basic minimum being a 2 channel analyser. Triaxial or uniaxial accelerometers may be used to make FRF measurements.

The most attractive feature of MIMO testing is that it provides a better distribution of the input force energy over the structure, as compared to single point excitation, resulting in a more uniform vibration response over the structure. Excitation at multiple locations also simulates more closely the excitation that the structure experiences during real-life operation. Simultaneous measurement of FRFs at more than one column provides increased consistency in the data set which is an important consideration in polyreference modal extraction procedures or when linear superpositions of columns are computed for modal enhancement. Besides, this method results in shorter measurement time, as compared to sequential column measurements.

There is another method called multi-point sine dwell test or normal mode testing method that uses fixed frequency sinusoidal excitation with multiple shakers. Signal generators with multiple signal sources allow one to tune the excitation frequency of the system to a resonance frequency, thus analysing one mode at a time. The amplitude and phase of the force signals should be set such that they match those of the mode being excited, i.e. the force signals are set according to the mode shape of the mode under consideration. This exercise is very time consuming and is therefore not very popular.

Yet another form of modal analysis is the so called operational/ambient/output only modal analysis. This method involves extraction of modal parameters from response data alone without knowledge of the input excitation forces. This method is convenient for parameter estimation of large structures like bridges and towers where it is difficult to artificially excite the structure and determine the responses.

10.4.3 Estimation of Frequency Response Functions

Ideally the FRF should give a very accurate representation of the actual structural response. But quite often the FRFs are distorted by noise in the input or output or both and hence a lot of research has been done on obtaining good FRFs. Most FFT analysers use the conventional H_1 estimator to estimate the FRF. There is also the H_2 FRF which is an improvement over the H_1 estimator. There are other ways of estimating FRFs as well, such as H_c and H_v. Four methods of estimating FRFs are presented in this section. Let us first start with the derivation of the H_1 FRF. For this, consider a model of the measurement system shown in Fig. 10.22 which contains noise in both the input and output.
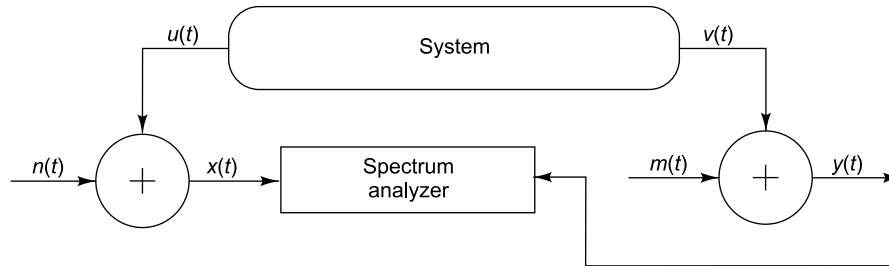
FIGURE 10.22 System measurement model

We shall use the following notations for the subsequent derivations. Let

$u(t)$ = the true input signal to the system (not measured signal)

$v(t)$ = the true output of the system (not measured signal)

$n(t)$ = the uncorrelated noise in the input

$m(t)$ = the uncorrelated noise in the output

$x(t)$ = the input signal measured by the FFT analyser, which is the sum of the true input and uncorrelated noise

i.e. $x(t) = u(t) + n(t)$ (10.31)

$y(t)$ = the output signal measured by the FFT analyser, which is the sum of the true output and uncorrelated noise.

i.e. $y(t) = v(t) + m(t)$ (10.32)

S_{uu} = power spectral density of the true input

S_{vv} = power spectral density of the true output

S_{nn} = power spectral density of random measurement noise in the input

S_{mm} = power spectral density of random measurement noise in the output

S_{xx} = power spectral density of the measured input

i.e. $S_{xx} = S_{uu} + S_{nn}$ (10.33)

S_{yy} = power spectral density of the measured output

i.e. $S_{yy} = S_{vv} + S_{mm}$ (10.34)

S_{xy} = cross-spectral density between measured input x and measured output y

S_{uv} = cross-spectral density between the true input u and the true output v

The true FRF is denoted by $H_0(f)$ and is defined as shown below.

$$H_0(f) = \frac{S_{uv}}{S_{uu}} = \frac{U^*\{f\}V\{f\}}{U^*\{f\}U\{f\}} = \frac{V\{f\}}{U\{f\}} \tag{10.35a}$$

Here the superscript * denotes the complex conjugation operator.

Alternately

$$H_0(f) = \frac{S_{vv}}{S_{vu}} = \frac{V^*\{f\}V\{f\}}{V^*\{f\}U\{f\}} = \frac{V\{f\}}{U\{f\}} \tag{10.35b}$$

For any measurement system contaminated by noise, it is obvious that it is impossible to measure the true FRF. Therefore, one needs to get estimates of the FRF, which are as close as possible, to the true FRFs.

10.4.3.1   *H_1 frequency response function*   This is the conventional way of estimating the FRF. Here the cross-spectral density between the input and output is divided by the auto spectral density of the input. The noise signals $n(t)$ and $m(t)$ in the input and output, respectively are assumed to be uncorrelated with each other and with the true input $u(t)$ and true output $v(t)$.

$$H_1(f) = \frac{S_{xy}}{S_{xx}} \tag{10.36}$$

Here

$$S_{xy} = X^*(f)Y(f) = \left[U^*(f) + N^*(f)\right]\left[V(f) + M(f)\right] \tag{10.37}$$

Since uncorrelated noise $n(t)$ and $m(t)$ get cancelled out after averaging over a sufficiently large number of records, S_{xy} reduces to S_{uv}. Therefore H_1 is expressed mathematically as

$$H_1(f) = \frac{X^*(f)Y(f)}{X^*(f)X(f)} = \frac{U^*(f)V(f)}{X^*(f)X(f)} = \frac{S_{uv}}{S_{xx}} \tag{10.38}$$

However, by definition, the true FRF is as defined in Eq. (10.35). Dividing the numerator and denominator of Eq. (10.38) by S_{uu} and substituting for $H_0(f)$ from Eq. (10.35a), we obtain the following simplification.

$$H_1(f) = \frac{S_{uv}}{S_{xx}} = \frac{S_{uv}}{S_{uu} + S_{nn}} = \frac{(S_{uv}/S_{uu})}{1 + (S_{nn}/S_{uu})} = H_0(f)\left[\frac{1}{1 + (S_{nn}/S_{uu})}\right] \tag{10.39}$$

The above equation relates the H_1 estimate to the true FRF H_0 and is seen to be sensitive to input noise S_{nn}. $H_1(f)$ is not good around resonance since S_{uu} itself becomes quite small in this region (of the order of S_{nn}). However, at antiresonance, it is to be noted that S_{nn}/S_{uu} becomes small and hence $H_1(f)$ is very good in this region.

10.4.3.2   *H_2 frequency response function*   This method was proposed as an alternative to $H_1(f)$ which is contaminated by input measurement noise, as revealed by Eq. (10.39). This inverse method for calculating an FRF is given as

$$H_2(f) = \frac{S_{yy}}{S_{yx}} = \frac{S_{vv} + S_{mm}}{S_{vu}} \qquad (10.40)$$

Dividing numerator and denominator by S_{vv}, the relationship between $H_2(f)$ and the true FRF $H_0(f)$ is obtained as

$$H_2(f) = \left(1 + \frac{S_{mm}}{S_{vv}}\right)\frac{S_{vv}}{S_{vu}} = H_0(f)\left(1 + \frac{S_{mm}}{S_{vv}}\right) \qquad (10.41)$$

Most commercially available analysers provide only one of these two forms of FRFs, and generally $H_1(f)$ is more common because it is slightly easier to compute. The $H_2(f)$ FRF estimate, unlike the conventional $H_1(f)$, is insensitive to input noise. Besides, since the input force signal is usually low relative to the measurement system noise at the resonant frequencies, i.e. since S_{nn}/S_{uu} is large, H_2 is a better estimator in this region. If H_1 is used in this situation, erroneous results will be obtained; on the other hand, H_2 remains unaffected. In addition, at resonance, the true output signal is very large compared to the output noise, i.e. S_{mm}/S_{vv} is very small compared to errors in H_1 around resonance. Therefore, H_2 yields an almost perfect estimate of the true H_0 in the region of resonance. However, at antiresonance, the output power spectrum drops towards output noise S_{mm}, while the input power spectrum remains high with respect to the input noise power spectrum. Under these conditions, H_2 fails to yield acceptable results because of its sensitivity to output noise.

10.4.3.3 Coherence function
The coherence function γ_{xy}^2 is used to detect errors in estimating an FRF. Coherence is defined as

$$\gamma_{xy}^2 = \frac{|S_{xy}|^2}{S_{xx}S_{yy}} \qquad (10.42a)$$

The coherence function is always less than or equal to 1.0. One can compute the coherence function in alternative way, as

$$\gamma_{xy}^2 = \frac{H_1(f)}{H_2(f)} = \frac{(S_{xy}/S_{xx})}{(S_{yy}/S_{yx})} = \frac{S_{xy}S_{xy}^*}{S_{xx}S_{yy}} = \frac{|S_{xy}|^2}{S_{xx}S_{yy}} \qquad (10.42b)$$

If the measurements are made accurately, coherence must be very close to unity. If coherence is small, the FRF should be discarded and it must be determined why it is small. A small frequency resolution in the FFT analyser may be a possible cause or there may be noise in either the excitation or the response signals, which are possibly degrading the measured spectra. One of the most likely sources of low coherence in lightly damped structures is poor frequency resolution in the analyser. If the resolution is not fine enough to describe adequately the rapid amplitude changes around resonances and antiresonances, the resulting FRF data are poor. The coherence function also gives an indication if the measured response can be fully attributed to the measured inputs. A value of the coherence function (Fig. 10.23) less than 1 over a frequency range indicates that response is only partially due to the measured inputs in this range and may be the result of unknown inputs. If the number of averages taken under such circumstances is increased, this situation does not change, and therefore the coherence function also does not change, but will only be less noisy. The only way to improve the quality of coherence functions is to take all possible

precautions to eliminate unknown sources. This can be achieved by ensuring proper boundary conditions, minimizing shaker–structure interaction, etc. When the coherence is low in spite of all such precautions, a zoom measurement may be made to enhance the accuracy of the measurements around resonance, but at the cost of increased measurement time. It is a good practice to make several measurements and to accumulate a running average of the corresponding FRF estimates and the coherence.
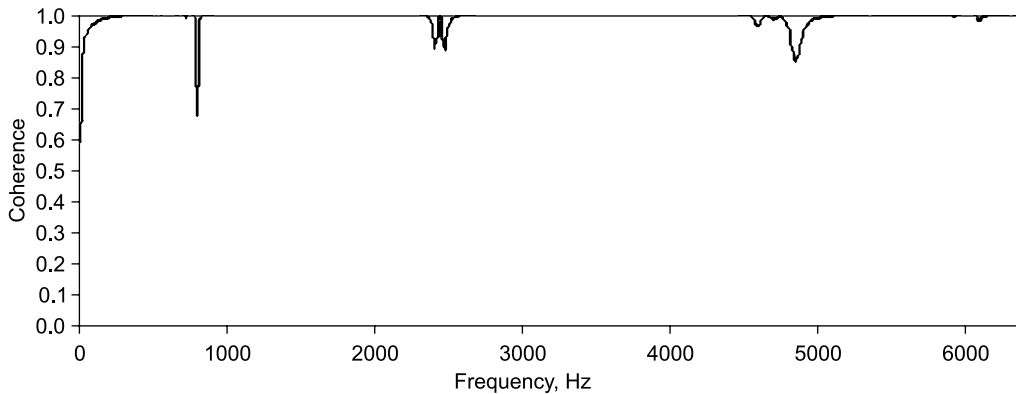
FIGURE **10.23**   Typical coherence plot

While conducting MIMO studies, it has to be ensured that the excitation forces applied to the structure are random and uncorrelated with each other in order to satisfy the mathematical assumptions made in the analysis. Actual excitation forces are in fact uncorrelated. However, at the natural frequencies, the structure loads the input drive systems, making the multiple excitation forces correlated and leading to degradation in the quality of FRFs and hence the coherence function. Constant current power amplifiers may be used to drive the exciters, instead of the usual constant voltage amplifiers to reduce these problems. Another problem is that the response measured at a point cannot be completely attributed to the force which is measured. This happens if the coupling between the shaker and the structure is too stiff and a lateral or rotational constraint is inadvertently applied to the test structure, giving rise to low coherence. System non-linearities can also give rise to low coherence.

10.4.3.4 H_3 *frequency response function* From previous discussions, it can be concluded that the H_1 estimator minimizes the error due to noise in the output, but is very sensitive to noise in the input, resulting in an underestimation of the true FRF, H_0, around resonance. The H_2 estimator, on the other hand, minimizes the error due to the noise in the input, but is very sensitive to noise in the output. This results in an overestimation of the true FRF H_0. Since H_1 is a lower bound estimator and H_2 an upper bound estimator, the true FRF H_0 must lie somewhere between the two estimates. This suggests a third method of estimating an FRF, called H_3 as the average or arithmetic mean of H_1 and H_2 as follows:

$$H_3(f) = \frac{H_1(f) + H_2(f)}{2} \tag{10.43}$$

This estimator has contamination from the input and output noise. However, the high and low estimators tend to cancel each other, thus yielding a better estimator of the FRF.

10.4.3.5 H_c *frequency response function* The H_c estimator is an unbiased FRF estimator. It attempts to eliminate the errors in FRF estimation caused by non-linearities and shaker–structure interaction during electrodynamic random shaker excitation as shown in Fig. 10.24.
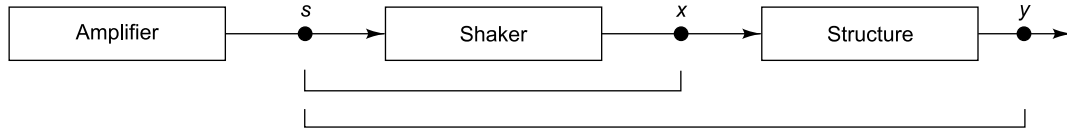
FIGURE 10.24 Shaker–structure interaction

Here

s = driving or actuating signal

x = input signal (reference)

y = output signal (response)

The complex version of this frequency response estimator is given by the following formula.

$$H_c = \frac{S_{ys}}{S_{xs}}$$

(10.44)

where

S_{ys} = Cross-spectral density of response with actuating signal

S_{xs} = Cross-spectral density of reference signal with actuating signal

The subscript c in H_c is an indication that the estimator is obtained from the ratio of complex cross-spectral estimates. Both the cross-spectra are with reference to the driving or source signals. Thus, all the signals within the force applied to the structure and the structural responses which are uncorrelated with the input driving signal are averaged out in the cross-spectral averaging process, irrespective of whether or not one is in the vicinity of resonance or in the antiresonance regions of the FRF.

10.5 SIGNAL PROCESSING PROBLEMS AND SOLUTIONS IN MODAL ANALYSIS

There are many signal processing techniques which may be used for improving the quality and accuracy of measurements and for reducing measurement errors; they are autoranging, averaging, windowing and zooming, to name a few.

10.5.1 Autoranging

It is required to obtain a high SNR and to avoid signal overloads (to prevent clipping), especially with impact excitation, in order to get meaningful FRFs. For this, the dynamic range of the analog-to-digital converter (ADC) of each channel may be optimized using a feature called autoranging which is available in most analysers. If the input range of the ADC is set too low, clipping will occur; if, on the other hand, it is set too high, the sensitivity of the measurements will be poor. The ADC range of each input channel must therefore be set according to the maximum anticipated voltage levels on the channel. Ideally, the peak amplitude of the measured signal should fall within the upper half of the input range. This is easy to set with continuous signals, but not so with impact or other transient signals. Hence, the ADC range for impact testing may be set manually or the modal analysis system may be trained by repeatedly impacting the structure while the analysis system autoranges until no overloads are detected. If the impacts are consistent, then the autoranging may be successful with a smaller number of averages, if not a large number is required. This may be a time-consuming procedure, especially with a large number of input channels, but definitely yields good results. Another point to be noted while autoranging is that it

should be done at the highest available sampling frequency, otherwise frequency components of the impact signal above the test frequency range could cause overloading. Often the range settings of the input channels may be used for several measurement locations if impacts are made in the same direction.

10.5.2   Removing Noise and Distortion from Measurements

In the process of obtaining transfer functions, power spectrum averaging is done as has been seen. This helps remove uncorrelated extraneous noise from the measurement. It also helps reduce the statistical variance of responses to random excitation functions and to reduce the effects of system non-linearities. Averaging measured spectra will thus decrease noise in FRFs and in the measured coherence function, but will not change the shape of the latter. However, increasing the number of averages will increase statistical confidence in the measured FRFs; thus the results will be acceptable even with low coherence. Once a sufficient number of records of data have been averaged together, the transfer function is computed by using H_1 or H_2 or any of the other estimates described earlier. Thus unbiased estimates of the transfer function may be obtained in the presence of noise.

Distortion, associated with non-linear motion, is an unwanted contaminant of the measured vibration signals. While making measurements, we must ensure that the data obtained accurately describe linear motion of the structure, and not any non-linear motion, since the dynamic model on which all modal testing theory is based is linear. Distortion can occur if the range of the ADC or amplifier is not set sufficiently high or if there is an unwanted vibration stop in the test setup. Distortions due to non-linearities in the structure show up as random noise in the spectrum and can be got rid of by a sufficiently large number of averages, like any other extraneous noise.

10.5.3   Windowing

Though FFT techniques are highly useful for fast computation of spectra, they have the limitation that they are based on periodicity of the time signals in the analysis window considered, resulting in leakage as seen in Section 6.9. In practice, due to leakage effects, the resulting FRF gets smeared, leading to a broadening of the resonance peaks. Therefore in FRF measurements, leakage leads to underestimation of the magnitude at the FRF peaks and distortion of phase and drop in coherence at the peaks. This leads to errors in estimates of the modal parameters obtained from signals which are non-periodic in the window, but does not cause a problem with transient signals such as an impact force or impulse response function which decay well before the end of the window. The periodic random signal contains the desirable properties of being random, yet repetitive in the time window, thus removing distortions from the measurement. The only drawback of this excitation signal is that it is two to three times slower than the other random methods.

10.6   MODAL PARAMETER EXTRACTION METHODS FOR SDOF AND MDOF SYSTEMS

10.6.1   Preliminary Checks of FRF Data

Before embarking on a modal parameter extraction procedure, it has to be ascertained that the FRFs to be used for the same are good. The checks are typically made using log–log Bodé plots of the measured receptance, mobility or inertance FRFs. The points to be checked are listed as follows:

   (i) The characteristics of the FRF plot at very low frequencies, below the first natural frequency, should be checked to see if the features corresponding to the test boundary conditions are reflected. A fixed support condition should result in a stiffness-line asymptote characteristic at low frequencies, the magnitude of this stiffness corresponding to the static stiffness of the structure at the clamp. A free support condition, on the other hand, should result in a mass-line asymptote in this low frequency region, with its magnitude deducible from rigid body considerations. If these low frequency characteristics do not match the expected behaviour, it can be concluded that measurements have not been made at sufficiently low frequencies or that the achieved support condition deviates from what was aimed at.

  (ii) Another verification can be made near the upper end of the desired frequency range. If the point mobility measurements show the curve becoming asymptotic to a mass line or a stiffness line (which is more common), it indicates that the excitation has been applied to a point of very high mass or flexibility. Under such circumstances, it would be desirable to obtain the FRFs again using a different excitation point.

 (iii) A third set of checks can be made to see if antiresonances and minima between resonances occur as expected. For this of course, the FRFs should show clear resonance and antiresonance characteristics in the first place. A point mobility plot should reveal an antiresonance after every resonance; a transfer mobility plot between two points far apart on the structure should show more minima than antiresonances. Besides, the resonance peaks and antiresonance dips should exhibit the same degree of sharpness on a log–log plot; if not, the implication is that the quality of measurements is poor because the measured vibration levels are too low or because the chosen frequency resolution on the FFT analyser is not fine enough. Measurements may have to be repeated in such a case also.

 (iv) If the FRF representations are in the form of Nyquist plots, then it is to be checked if each resonant region traces out a circular arc. For a system with well-separated modes, each of these arcs will correspond to a major portion of a circle, while for closely spaced modes, the arcs will subtend only 45° or 60° at the centres of the circles. If the arcs are not smooth, it implies that the measurements are poor.

10.6.2   SDOF Modal Analysis

There exist a number of modal analysis methods such as peak-pick method, circle-fit method, inverse method, etc. All these are frequency domain methods in the sense that they operate on the FRF to get the modal parameters and all of them assume that in the region of a resonance, the natural mode under consideration contributes to most of the response, the contributions from other modes being negligible. The methods vary as to whether they assume all the response being attributed to that single mode or whether the contribution of other modes is represented by a simple approximation. There are time domain methods also and these operate on the impulse response function, but these are not discussed in this book.

10.6.2.1   Peak-pick method   The simplest of these methods is called the 'peak-amplitude' or 'peak-pick' method. The assumption behind this method is that the FRFs of the structures considered have certain characteristics: (i) they have well-separated modes, (ii) the modes are not too lightly damped as to make it difficult to obtain accurate measurements around resonance and (iii) the modes are not so

heavily damped that the response at a resonance has contributions from more than one mode. Though these characteristics limit the applicability of the method, it can be used to obtain initial estimates to the modal parameters. Subsequently for obtaining better estimates, other curve-fitting procedures may be used. The peak-pick procedure works on one mode after another and is as follows:

   (i) Frequencies corresponding to peak values of response in the FRF curve (Fig. 10.25a) are identified as being the natural frequencies of various modes. Let us denote the frequency of maximum response as the natural frequency ω_r of a particular mode.

  (ii) The peak value of the response corresponding to this mode is denoted as $|\hat{\alpha}|$. The half-power or 3 dB frequencies ω_1 and ω_2 (Fig. 10.25b) with a response level of $|\hat{\alpha}|/\sqrt{2}$ are determined on either side of resonant frequency ω_r and the half power bandwidth $\Delta\omega = \omega_2 - \omega_1$ is computed.

 (iii) The damping of the mode under consideration can be got from the following formula.

$$\eta_r = \frac{\left(\omega_2^2 - \omega_1^2\right)}{\omega_r^2} \cong \frac{\Delta\omega}{\omega_r} \tag{10.45a}$$

$$\zeta_r = \eta_r / 2 \tag{10.45b}$$

  (iv) An estimate of the frequency dependent modal constant of the mode under consideration may be obtained based on the SDOF assumption that the total response in this frequency region is due to only one term in the general FRF series representation.

$$\alpha_{jk}(\omega) = \sum_{r=1}^{N} \frac{(_r\varphi_j)(_r\varphi_k)}{\omega_r^2 - \omega^2 + i\eta_r\omega_r^2} = \sum_{r=1}^{N} \frac{_rA_{jk}}{\omega_r^2 - \omega^2 + i\eta_r\omega_r^2} \tag{10.46}$$

Here $_rA_{jk}$ is the modal constant corresponding to the response of the *j*th degree of freedom when the system is excited at the *k*th degree of freedom in the *r*th mode. Thus

$$|\hat{\alpha}| = \frac{A_r}{\left(\omega_r^2\eta_r\right)} \tag{10.47a}$$

or

$$A_r = |\hat{\alpha}|\omega_r^2\eta_r \tag{10.47b}$$

This method has certain shortcomings due to the assumptions made. The estimates of both damping and modal constant depend on the peak value of FRF level $|\hat{\alpha}|$. There is a lot of uncertainty in measurements around the resonance region, especially if measurements have been made with a poor frequency resolution and if the structure has lightly damped modes. This method ends up measuring the peak value based entirely on a single point in the FRF. Besides, this method gives estimates of real modal constants, implying that we are considering systems with real modes, or proportionally damped structures. In addition, the single-mode assumption is a major drawback leading to inaccuracy of the estimate. Even in systems with well-separated modes, it is often found that the neighbouring modes do contribute to the total response at the resonance under consideration. The circle-fit method is a refinement of this technique and does away with these problems. Figure 10.25(c) shows a Nyquist plot which might give the modulus plot shown in Fig. 10.25(b).

(a)

(b)

(c)
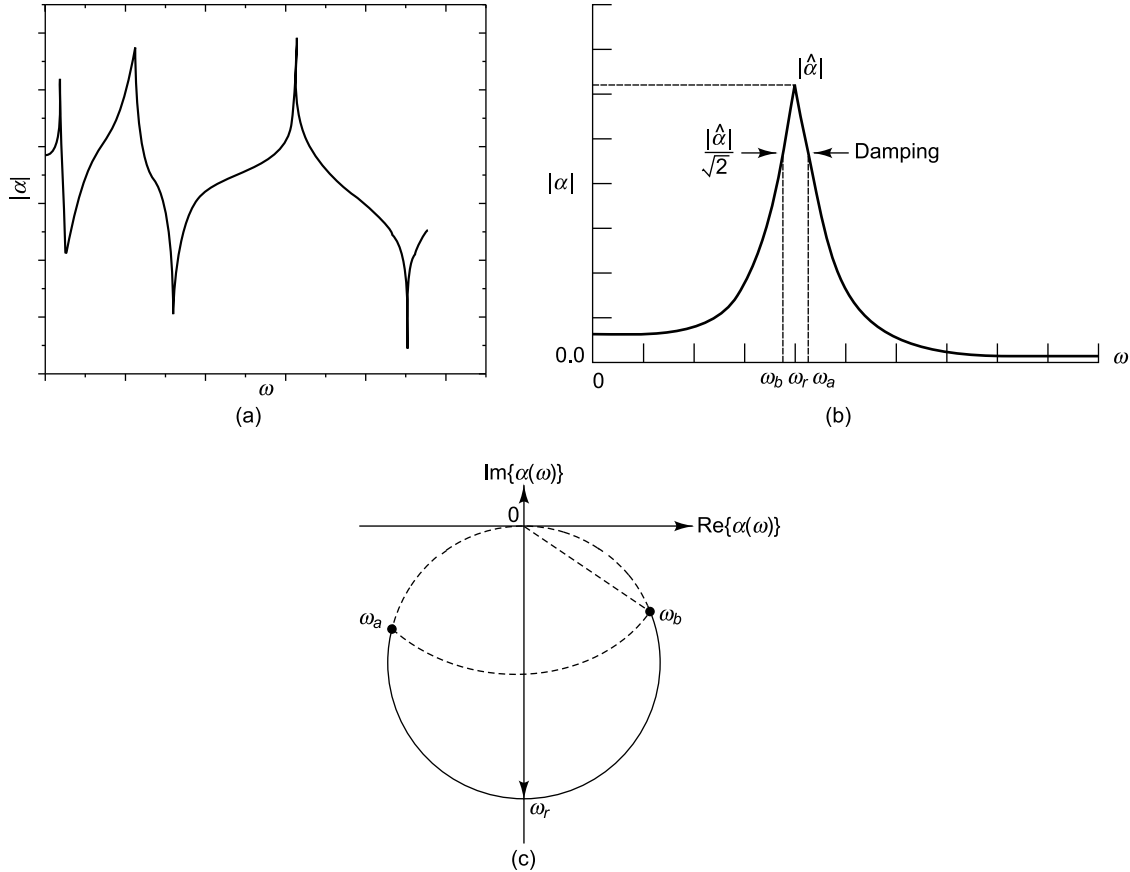
FIGURE 10.25   Peak-pick method of modal parameter extraction: (a) measured FRF plot, (b) determination of half-power frequencies, (c) Nyquist FRF plot

The above procedure may be applied to a plot of the real part of the receptance FRF, instead of the modulus plot as shown in Fig. 10.26. This curve indicates that the positions and values of the maximum and minimum values yield good estimates of the half-power points. Besides, a more refined estimate of the natural frequency itself can be obtained as the point midway between the maximum and minimum on the imaginary plot. Thus,

$$A_r = \left(|MX| + |MN| \right) \omega_r^2 \eta_r \qquad (10.48)$$

10.6.2.2 *SDOF circle-fitting method*   We know that for a general SDOF system, a Nyquist plot of frequency response produces circle-like curves and that specifically the receptance FRF gives a true circle with a structural damping
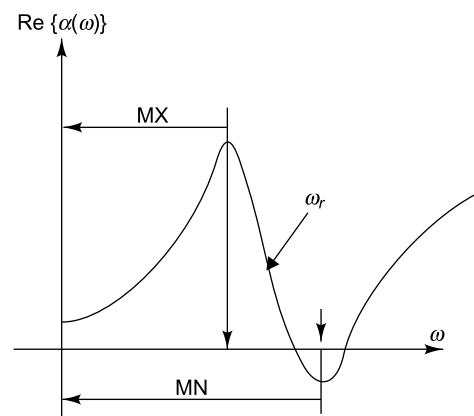
FIGURE 10.26   Real FRF plot

model and the mobility FRF with a viscous damping model (the circle is rotated by 90° in the complex plane). Further, we have seen for MDOF systems that Nyquist plot of FRF data show sections of near-circular arcs in the regions around resonances. This feature provides the basis for the SDOF circle-fitting method. The derivations in this section are based on a system with structural damping and thus the receptance FRF data are used. Many of the modal analysis packages offer both types of damping and take the appropriate FRF representation for the circle-fitting procedure.

As implied by the name, this method makes use of the fact that the response of a system around resonance, is predominantly due to a single mode. This implies that the magnitude of the FRF is mainly due to one term in the series, i.e. that due to the mode being analysed. Equation (10.46) can be rewritten, without simplification, as

$$\alpha_{jk}(\omega) = \frac{{}_r A_{jk}}{\omega_r^2 - \omega^2 + i\eta_r \omega_r^2} + \sum_{s=1, \neq r}^{N} \frac{{}_s A_{jk}}{\omega_s^2 - \omega^2 + i\eta_s \omega_s^2} \tag{10.49}$$

By virtue of making the SDOF assumption, it can be said that over a small range of frequencies around natural frequency ω_r of mode r, the frequency dependence of the second term in Eq. (10.49) is negligible and hence it simplifies to

$$\alpha_{jk}(\omega)_{\omega = \omega_r} = \frac{{}_r A_{jk}}{\omega_r^2 - \omega^2 + i\eta_r \omega_r^2} + {}_r B_{jk} \tag{10.50}$$

This is illustrated in Fig.10.27. The figure shows each of the two terms in Eq. (10.50), as well as the total receptance over the same frequency range. It is seen that the first term (contribution of the mode being analysed) varies considerably through the resonance region, sweeping out a circle, while the second term (contribution of all other modes) is a constant, independent of frequency. Thus, the total receptance is a circle with the same properties as that of the mode under consideration, but displaced from the origin of the Argand plane by the constant contributed to by all the other modes.
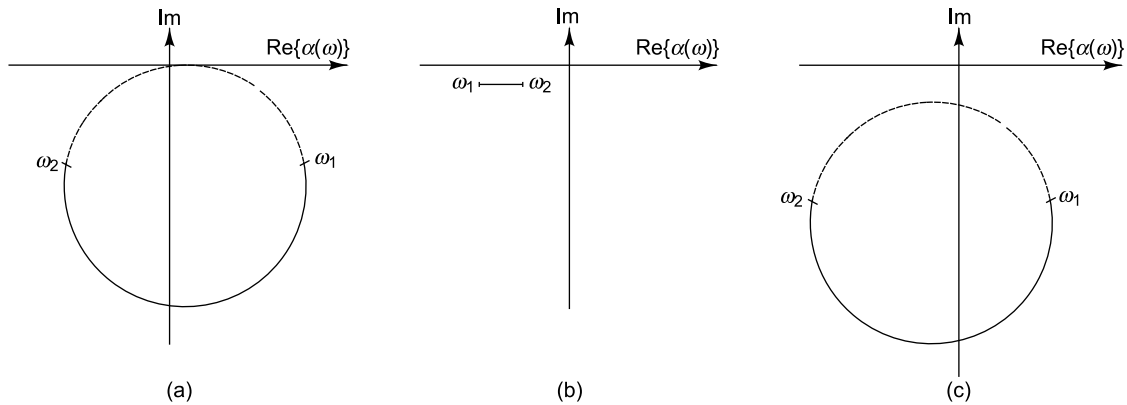
FIGURE 10.27   Nyquist FRF plot

Properties of the modal circle with structural damping   The relevant properties of the modal circle which are useful in extracting the required modal parameters are described in this section. The basic FRF is of the form

$$\alpha = \frac{1}{\omega_r^2 \left[1 - (\omega/\omega_r)^2 + i\eta_r \right]}$$

(10.51)

$$\tan \gamma = \frac{\eta_r}{[1 - (\omega/\omega_r)^2]}$$

(10.52a)

$$\tan(90° - \gamma) = \tan\left(\frac{\theta}{2}\right) = \frac{[1 - (\omega/\omega_2)^2]}{\eta_r}$$

(10.52b)

If the modal constant $_rA_{jk}$ is included in the numerator of Eq. (10.51), it scales the size of the circle by $|_rA_{jk}|$ and rotates it by $\angle_rA_{jk}$ so that the principal diameter (the one passing through the natural frequency) is oriented at an angle $\angle_rA_{jk}$ to the negative imaginary axis. From Eq. (10.52c) we obtain

$$\omega^2 = \omega_r^2 [1 - \eta_r \tan(\theta/2)]$$

(10.53)

Differentiating Eq. (10.53) with respect to θ, we get

$$\frac{d\omega^2}{d\theta} = \left(\frac{-\omega_r^2 \eta_r}{2} \right) \left\{ 1 + \frac{[1 - (\omega/\omega_r)^2]^2}{\eta_r^2} \right\}$$

(10.54)

The reciprocal of this quantity is a measure of the rate at which the locus sweeps around the circular arc; the maximum sweep rate occurs when $\omega = \omega_r$. Thus

$$\left(\frac{d}{d\omega} \right) \left(\frac{d\omega^2}{d\theta} \right) = 0 \quad \text{when} \quad \left(\omega_r^2 - \omega^2 \right) = 0$$

(10.55)

The sweep rate parameter provides an estimate of damping since

$$\left(\frac{d\theta}{d\omega^2} \right)_{\omega=\omega_r} = \frac{-2}{\left(\omega_r^2 \eta_r \right)}$$

(10.56)

This equation gives a hint as to where exactly the natural frequency is; the relative spacing of the measured data points around the circular arc near each resonance helps determine its value. In Fig. 10.28, let ω_b and ω_a correspond to two frequencies on the modal circle, ω_b being below the natural frequency and ω_a being above. Then

$$\tan\left(\frac{\theta_b}{2}\right) = \left[\frac{1 - (\omega_b/\omega_r)^2}{\eta_r} \right]; \quad \tan\left(\frac{\theta_a}{2}\right) = \left[\frac{(\omega_a/\omega_r)^2 - 1}{\eta_r} \right]$$

(10.57)

From these two equations, the exact equation for damping of the mode is

$$\eta_r = \frac{\left(\omega_a^2 - \omega_b^2 \right)}{\{\omega_r^2 [\tan(\theta_a/2) + \tan(\theta_b/2)]\}}$$

(10.58a)

For light damping (say, less than 2–3%), the expression above becomes

$$\eta_r = \frac{2\left(\omega_a - \omega_b \right)}{\{\omega_r [\tan(\theta_a/2) + \tan(\theta_b/2)]\}}$$

(10.58b)

For the two half-power points for which $\theta_a = \theta_b = 90°$, we get

$$\eta_r = \frac{(\omega_2 - \omega_1)}{\omega_r} \tag{10.59a}$$

If the damping is not light we get

$$\eta_r = \frac{(\omega_2^2 - \omega_1^2)}{\omega_r^2} \tag{10.59b}$$
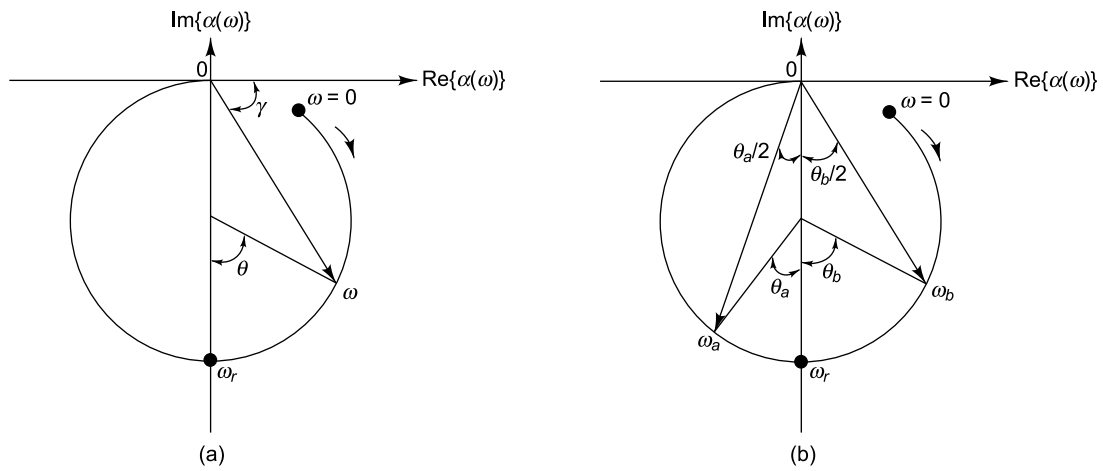
FIGURE 10.28 Modal circle

Finally the diameter of the circle is got as $1/(\omega_r^2 \eta_r)$. If this equation is scaled by a modal constant in the numerator, the diameter of the circle is

$$_r D_{jk} = \frac{|_r A_{jk}|}{(\omega_r^2 \eta_r)} \tag{10.60}$$

Properties of the modal circle with viscous damping   With viscous damping, mobility FRF gives a true circle and hence should be used instead of receptance. Thus

$$Y(\omega) = \frac{i\omega}{(k - \omega^2 m) + i\omega c} \tag{10.61}$$

From the real and imaginary parts of Eq. (10.61), we get

$$\tan\left(\frac{\theta}{2}\right) = \frac{\omega(k - \omega^2 m)}{\omega^2 c} = \frac{[1 - (\omega/\omega_r)^2]}{(2\zeta\omega/\omega_r)} \tag{10.62a}$$

$$\tan\left(\frac{\theta_b}{2}\right) = \frac{[1 - (\omega_b/\omega_r)^2]}{(2\zeta\omega_b/\omega_r)}; \quad \tan\left(\frac{\theta_a}{2}\right) = \frac{[(\omega_a/\omega_r)^2 - 1]}{(2\zeta\omega_a/\omega_r)} \tag{10.62b}$$

These equations give

$$\zeta = \frac{\left(\omega_a^2 - \omega_b^2\right)}{2\omega_r[\omega_a \tan(\theta_a / 2) + \omega_b \tan(\theta_b / 2)]} \tag{10.63a}$$

For lightly damped structures,

$$\zeta \cong \frac{\left(\omega_a - \omega_b\right)}{\omega_r[\tan(\theta_a / 2) + \tan(\theta_b / 2)]} \tag{10.63b}$$

For the half-power frequencies θ_a = θ_b = 90° and therefore we have

$$\zeta = \frac{\left(\omega_2 - \omega_1\right)}{2\omega_r} \tag{10.63c}$$

This is an exact equation independent of the level of damping.

Circle-fitting procedure   The objective behind this analysis procedure is to extract the necessary coefficients in Eq. (10.46), and from them the modal parameters. The following five-step procedure is described for the case of structural damping. A similar approach is used for viscous damping also.

  (i) *Selection of points to be used:* A fixed number of points (at least six) should be chosen on either side of any identified resonance in the FRF. If the system is user-interactive in nature, the analyst may use his/her discretion to separate true modes from spurious perturbations and to reject suspect data points. The selected points should not be influenced much by adjacent modes. It is preferable to have a circular arc of at least 180° and, if possible, more.

 (ii) *Fitting circle and calculating quality of fit:* This is usually done using one of a host of curve-fitting algorithms. It entails finding a circle with the minimum least-squares deviation for the selected points. The most accurate way of doing this is to minimise the deviations of the measured points from their actual locations on the circle. Another possible method involves minimising the deviations of points from the nearest point on the circle and is not as accurate; nevertheless, it is more common since it is easier to implement. The centre and radius of the circle are obtained at the end of this exercise, along with an error which is nothing but the mean square deviation of the chosen points from the circle. These deviations are typically of the order of 1–2%.

(iii) *Locating natural frequencies and obtaining damping estimates:* One of the most common definitions used for the natural frequency in order to locate it is that it is the frequency at which the sweep rate around the circle reaches a maximum. To determine this, radial lines are constructed from the centre of the circle to the points around resonance and the angles they subtend with each other are noted, using which the sweep rate through the region is estimated and thence the natural frequency. If, the frequency points used are spaced at regular intervals, then a finite difference method may be used. Other definitions of the natural frequency are also used; they are: (i) the frequency of maximum response, (ii) the frequency of maximum imaginary receptance and (iii) the frequency of zero real receptance.

(iv) *Calculation of multiple damping estimates and scatter:* A number of damping estimates are generally calculated using all combinations of selected data points, two at a time with one on either side of resonance (Eq. 10.63a). It is a common practice to evaluate the mean, as well as the deviation of the estimates. A deviation of less than 4–5% indicates a successful analysis, but if the scatter is 20

or 30%, the values used should be re-examined. Random variations in damping estimates could be due to measurement errors; systematic variations, on the other hand, could be due to poor experimental setup, contributions from neighbouring modes, non-linear behaviour of the system, etc. Such errors will not get averaged out.

(v) *Determination of modal constant:* This involves finding the magnitude and phase of the modal constant from the diameter of the circle and from its orientation in the Argand plane. This calculation is straightforward once the natural frequency has been located and the damping estimates obtained.

After the entire circle-fitting analysis procedure is over, a theoretically regenerated FRF plot may be obtained and compared with the original measured data. At this juncture, it will be necessary to add the contribution of the other modes to the resonant mode under consideration. This contribution ($_r B_{jk}$ in Eq. 10.50) may be measured off as the distance from the top of the principal diameter to the origin. Then, using this equation and the modal parameters extracted, a curve based on the model obtained may be plotted.

10.6.2.3   SDOF modal analysis using inverse method

The inverse method is based on the fact that a function, which generates a circle when plotted in the complex plane, will trace out a straight line when plotted as a reciprocal (Fig. 10.29). Thus, if we were to plot the reciprocal of receptance of an SDOF system with structural damping, it would produce a straight line as can be seen from the expressions below.

$$\alpha(\omega) = \frac{(k - \omega^2 m) - ih}{(k - \omega^2 m)^2 + h^2} \tag{10.64a}$$

and

$$\frac{1}{\alpha(\omega)} = (k - \omega^2 m) + ih \tag{10.64b}$$
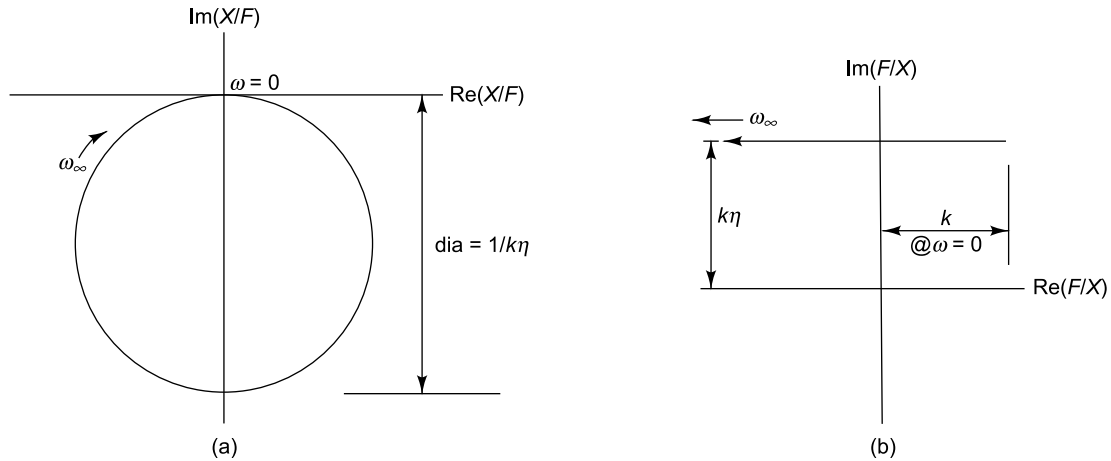
FIGURE 10.29   Nyquist plots of receptance: (a) standard receptance, (b) inverse receptance

The procedure is as follows. The inverse FRF is first computed and a best-fitting straight line is constructed through the data points. The intercept of this line with the imaginary axis gives an estimate for the damping parameter. The deviations of the data points from the line itself give an indication of the reliability of the estimate. If the points are randomly scattered on either side of the line, it is possible that there are experimental errors, but if the deviations are systematic, then there is a source of bias in the data and this needs investigation. A least-squares operation indicating the deviation between the real part of the measured data points and that of the theoretical model gives estimates for the mass and stiffness parameters in the theoretical model. This method gives best results for systems with real modes and relatively well-separated modes; corrections have to be applied if the FRF has contributions from more than one mode. A very attractive feature of this method is that the straight line can readily be obtained with data points far away from resonance and not necessarily exactly at the natural frequency (where the real part of the inverse receptance is zero). The method described here for structural damping can be extended to the case of viscous damping as before.

10.6.3 MDOF Curve-Fitting Methods

The SDOF approach to modal analysis is inappropriate under certain circumstances such as when the system has closely coupled modes or extremely light damping, making measurements at resonance inaccurate and difficult to obtain. Under these conditions, alternative methods called MDOF curve fits may be used. By closely coupled modes are meant those modes for which the response at resonance is not due to just one mode (or term in the FRF series) because the natural frequencies are very closely spaced, or because they have relatively heavy damping, causing an overlapping of modes.

A regenerated FRF obtained by repeated application of any of the SDOF curve-fitting procedures around each of the resonances in the frequency range of interest reveals, on comparison with the FRF from original measurements, that the match is poor, especially in regions away from resonances. This is because the regenerated curve has taken into account the frequency range f_1 to f_2 in computing modes m_1 to m_2, whereas the actual measured curve has contributions from modes in the frequency ranges below f_1 and above f_2 also. Thus the regenerated curve can be corrected by accounting for the effects of these modes also. This is done by including two simple extra terms called residuals.

Consider the case where the FRF is regenerated using a formula of the form

$$\alpha_{jk}(\omega) = \sum_{r=m_1}^{m_2} \frac{{}_rA_{jk}}{\omega_r^2 - \omega^2 + i\eta_r\omega_r^2} \tag{10.65}$$

Here modes from m_1 to m_2 alone have been included. But the fact remains that the modes outside this frequency range affect the measured FRFs. Hence, the equation most closely representing the measured data is

$$\alpha_{jk}(\omega) = \sum_{r=1}^{N} \frac{{}_rA_{jk}}{\omega_r^2 - \omega^2 + i\eta_r\omega_r^2} \tag{10.66a}$$

This may be rewritten as

$$\alpha_{jk}(\omega) = \sum_{r=1}^{m_1-1} + \sum_{r=m_1}^{m_2} + \sum_{r=m_2+1}^{N} \left(\frac{{}_rA_{jk}}{\omega_r^2 - \omega^2 + i\eta_r\omega_r^2} \right) \tag{10.66b}$$

The first term in this equation corresponds to the low frequency modes and the third term to the high frequency modes, both of which are not included in the regenerated FRF. It is only the second term which is included. Figure 10.30(a)–(d) shows the contribution of all modes in the range 1 to N, modes from $m_2 + 1$ to N, modes from m_1 to m_2 and those from 1 to $m_1 - 1$, respectively.
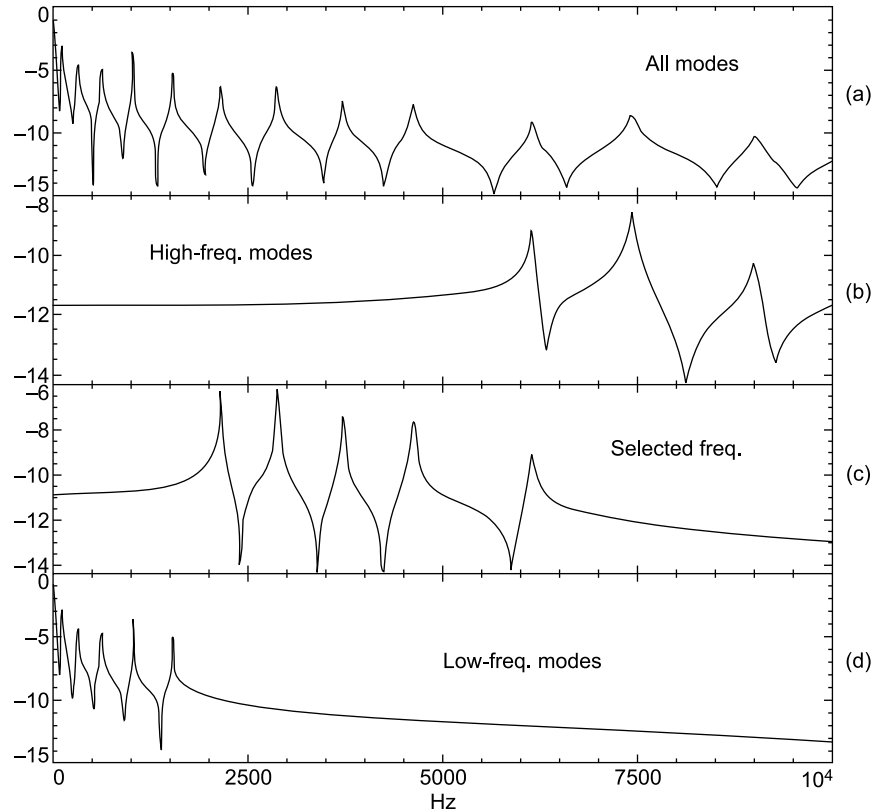
FIGURE 10.30 Receptance FRF of MDOF system page: (a) superposition of all modes, (b) modes corresponding to high-frequency range, (c) modes corresponding to identified frequency range, (d) modes corresponding to low-frequency range

From these figures it is clearly seen that the contribution of modes 1 to $m_1 - 1$ is that of a mass-like behaviour in the selected frequency range, while the contribution of modes $m_2 + 1$ to N is seen as a stiffness-like behaviour and these effects may be accounted for in the regenerated FRF using residual mass and stiffness terms as shown below.

$$\alpha_{jk}(\omega) = -\frac{1}{\omega^2 {}_M R_{jk}} + \sum_{r=m_1}^{m_2} \left(\frac{{}_r A_{jk}}{\omega_r^2 - \omega^2 + i\eta_r\omega_r^2} \right) + \frac{1}{{}_K R_{jk}} \qquad (10.67)$$

Here ${}_M R_{jk}$ and ${}_K R_{jk}$ represent the residual mass and stiffness terms, respectively for the frequency range considered. If the frequency range of analysis is changed, these terms also change. The regenerated curve is compared with the measured FRF at the lowest frequencies and from the difference between the

two, $_M R_{jk}$ is obtained. In a similar fashion, the regenerated and actual curves are compared at the highest frequencies and the difference gives $_K R_{jk}$. This process is refined iteratively in case the addition of the stiffness residual upsets the effectiveness of the mass term and the addition of the mass term upsets that of the stiffness term.

There are many different methods available for MDOF modal parameter extraction. In this section three frequency-domain MDOF curve-fitting methods are described. All of them are similar in that they perform a curve fit to the entire FRF measurement in one step. As for the SDOF cases, the structurally damped system is described and may be extended to the viscously damped case. The three approaches are as follows:

  (i) an extension of the SDOF method outlined in Section 10.6.2.1,

 (ii) a general approach to multimode curve-fitting and

(iii) a method meant for very lightly damped structures.

10.6.3.1   Extension of SDOF method

In the SDOF circle-fitting modal analysis procedure, it was assumed that near the resonance under consideration, the effect of all the other modes could be represented by a constant. The present method is an extension of that assumption, enabling a more accurate analysis of the data. The receptance FRF in the frequency range of interest may be written taking into account the residual mass and stiffness terms as

$$\alpha_{jk}(\omega) = \sum_{s=m_1}^{m_2} \frac{_s A_{jk}}{\omega_s^2 - \omega^2 + i\eta_s\omega_s^2} + \frac{1}{_K R_{jk}} - \frac{1}{\omega^2 \, _M R_{jk}} \tag{10.68a}$$

This can be rearranged as

$$\alpha_{jk}(\omega) = \frac{_r A_{jk}}{\omega_r^2 - \omega^2 + i\eta_r\omega_r^2} + \sum_{s=m_1,\neq r}^{m_2} \frac{_s A_{jk}}{\omega_s^2 - \omega^2 + i\eta_s\omega_s^2} + \frac{1}{_K R_{jk}} - \frac{1}{\omega^2 \, _M R_{jk}} \tag{10.68b}$$

The first term gives the contribution of the rth mode, the second term the contribution of all modes from m_1 to m_2 save the rth mode, the third term the contribution of all high frequency modes that have been left out of the FRF regeneration process and the fourth term the contribution of all the left out low frequency modes. The main difference between the SDOF circle-fitting procedure and the present MDOF method is that while the second term was assumed to be a constant in the former curve-fitting procedure, it is considered as frequency dependent in the present method. To simplify the curve-fitting procedure, the magnitude of the co-efficients in the second term in Eq. (10.68b) can be got from estimates which have already been obtained from an SDOF analysis. Let $\alpha_{jk}^m(\omega)$ represent a set of measured FRF data points around the resonance at ω_r. We can now regenerate FRFs at each frequency for which we have a measured FRF value. The difference between the two can be used as correction factors to get adjusted FRF data points which should then represent true SDOF behaviour as shown below.

$$\alpha^m{}_{jk}(\omega) - \left(\sum_{s=m_1,\neq r}^{m_2} \frac{_s A_{jk}}{\omega_s^2 - \omega^2 + i\eta_s\omega_s^2} + \frac{1}{_K R_{jk}} - \frac{1}{\omega^2 \, _M R_{jk}} \right) = \frac{_r A_{jk}}{\omega_r^2 - \omega^2 + i\eta_r\omega_r^2} \tag{10.69}$$

This procedure should be repeated iteratively for all the modes in the frequency range of interest till convergence is reached. The method offers a drastic improvement in cases where the modes are strongly coupled, though this not true for systems with relatively weakly coupled modes.

10.6.3.2 *General MDOF curve-fitting procedure* The previous method was an extension of SDOF analysis method. The general MDOF curve-fitting methods requiring sophisticated computational methods are newer developments in the field. Let an individual FRF measured data be represented as

$$\alpha_{jk}^m \left(\Omega_j \right) = \alpha_j^m \tag{10.70}$$

The corresponding regenerated values are denoted by

$$\alpha_{jk} \left(\Omega_j \right) = \alpha_j = \sum_{s=m_1}^{m_2} \frac{{}_s A_{jk}}{\omega_s^2 - \Omega_j^2 + i\eta_s \omega_s^2} + \frac{1}{{}_K R_{jk}} - \frac{1}{\Omega_j^2 \, {}_M R_{jk}} \tag{10.71}$$

Here the values of ${}_1 A_{jk}, {}_2 A_{jk}, \ldots, \omega_1, \omega_2, \ldots, \eta_1, \eta_2, \ldots {}_K R_{jk}$ and ${}_M R_{jk}$ are to be found. Let ε_j denote an individual error as

$$\varepsilon_j = \left(\alpha_j{}^m - \alpha_j \right) \tag{10.72}$$

The total error can be expressed as

$$E_j = \left| \varepsilon_j^2 \right| \tag{10.73}$$

A weightage may be assigned to each frequency ω_j in the frequency range of interest. The curve-fitting procedure has to be done such that the total error computed as shown below is minimized.

$$E = \sum_{j=1}^{N} \omega_j E_j \tag{10.74}$$

This may be done by differentiating Eq. (10.74) with respect to each unknown in turn, leading to as many equations as unknowns, many of which are non-linear in the ω_s and η_s parameters. Hence, there are many algorithms for solving these equations, each with its own simplifications and assumptions and giving rise to a different procedure.

10.7 Damage Detection from Modal Parameters

Modal parameters are used in damage detection on the basis that a defect will alter the stiffness, mass and damping of the system. These changes show up in the FRFs in the form of natural frequency shifts and changes in damping coefficients. Recent research has shown that the changes are very small and generally noticeable only for large defects. Researchers have also tried detection based on mode shape curvature obtained by numerically differentiating the relevant mode shape vectors. But this is a stupendous task for complex structures. One method which has been used successfully in helicopter blades is based on the beat frequencies obtained between the time record of a healthy blade (baseline signature) and the signature obtained at some other point of time. Any small differences in frequencies between the two signals are blown up when they beat. For this, inexpensive embedded piezoceramic materials are typically used for both excitation and sensing. This method has been very successful in detecting very small amounts of damage (<0.1% change in mass). Actuators and transducers based on smart materials such as piezoelectric patch and stack actuators, shape memory alloys and electro-rheological fluids can conveniently be embedded into large inaccessible structures to provide cheap diagnostic solutions based on vibration signals and will form the future diagnostic systems.

FURTHER READINGS

1. Allemang, R.J., Rost, R.W. and Brown, D.L., Dual input estimation of frequency response functions for experimental modal analysis of aircraft structures, *Proceedings of the 1st International Modal Analysis Conference*, Orlando, Florida, 333–340, 1983.

2. Bendat, J.S. and Piersol, A.G., *Random Data*: *Analysis* & *Measurement Procedures*, Wiley-Interscience, New York, 2000.

3. Clough, R.W. and Penzien, J., *Dynamics of Structures*, McGraw Hill Book Co., New York, 1993.

4. Conciauro, G., Guglielmi, M. and Sorrentino, R., *Advanced Modal Analysis*, John Wiley, New York, 2000.

5. De Silva, C.W., *Vibration Damping, Control, and Design*, CRC Press, Boca Raton, Florida, 2007.

6. De Silva, C.W., *Vibration Monitoring, Testing, and Instrumentation*, CRC Press, Boca Raton, Florida, 2007.

7. De Silva, M., Júlio M. and Maia, N.M.M. (Eds.), Modal Analysis and Testing, *Proceedings of the NATO Advanced Study Institute*, Sesimbra, Portugal, 3–15 May, 1998, Series: NATO Science Series E, Vol. 363, Springer, 1999.

8. Døssing, O., *Structural Testing. Part 1: Mechanical Mobility Measurements*, Brüel & Kjær Theory and Application Booklet, BR 0458-12, Denmark.

9. Døssing, O., *Structural Testing. Part 2: Modal Analysis and Simulation*, Brüel & Kjær Theory and Application Booklet, BR 0507-11, Denmark.

10. Elliott, K.B. and Mitchell, L.D., The Improved Frequency Response Function and its Effect on Modal Circle Fits, *ASME Journal of Applied Mechanics*, 51, 657–663, 1984.

11. Ewins, *D.J., Dynamic Testing Agency Handbook on Guidelines to Best Practice: Modal Testing*, Vol. 3, Dynamic Testing Agency, United Kingdom, 1993.

12. Ewins, D.J., *Modal Testing: Theory and Practice*, Research Studies Press Ltd., England, 2003.

13. Gade, S., Herlufsen, H. and Hansen, H.K., *How to Determine the Modal Parameters of Simple Structures*, Brüel&Kjær Application Note, Bo042, Brüel&Kjær, Denmark.

14. Ginsberg, J.H., *Mechanical and Structural Vibrations: Theory and Applications*, John Wiley & Sons, Inc., New York, 2001.

15. Hansen, C.H. and Snyder, S.D., *Active Control of Noise and Vibration*, Taylor & Francis, London, 1997.

16. Herlufsen, H., *Dual Channel FFT Analysis. Parts 1 and 2*, Brüel & Kjær Technical Reviews No.1 & 2, BV0013-11 & BV 0014-11, Denmark, 1984.

17. Herlufsen, H., *Modal Analysis using Multi-Reference and MIMO Techniques*, Brüel & Kjær Application Note BT 0001-12, Denmark, 1985.

18. http://literature.agilent.com/litweb/pdf/5954-7957E.pdf, The Fundamentals of Modal Testing, Hewlett Packard Application Note 243-3, USA.

19. Inman, D.J., *Engineering Vibration*, Prentice Hall, Upper Saddle River, NJ, 2001.

20. Inman, D.J., *Vibration with Control*, John Wiley & Sons Ltd., Chichester, England, 2006.

21. International Organization for Standardization, ISO 7626-1:1986, Vibration and Shock –Methods for the Experimental Determination of Mechanical Mobility, 1986.

22. Jimin He and Zhi-Fang Fu, *Modal Analysis*, Butterworth-Heinemann, Oxford, 2001.

23. Maia, N.M.M. and e Silva, M., *Theoretical and Experimental Modal Analysis*, John Wiley and Sons. Inc., Somerset, NY, 1997.

24. McConnell, K,G. and Varoto, P.S., *Vibration Testing: Theory and Practice*, Wiley-Interscience, New York, 1995.

25. Mitchell, L.D., Improved methods for the Fast Fourier Transform (FFT) calculation of the Frequency response function, *ASME Journal of Mechanical Design*, l04, 277–279, 1982.

26. Rocklin, G. T., Crowley, J. and Vold, H., A comparison of H_1, H_2 and H_v frequency response functions, *Proceedings of IMAC III*, Vol. l, pp. 272–278, 1985.

27. Rossing, T. D. and Fletcher, N.H., *Principles of Vibration and Sound*, Springer-Verlag, New York, 2004.

28. Slater, J.C., *Vibration Testing, with Modal Testing and Health Monitoring*, John Wiley and Sons, New York, 2007.

29. Thomson, W.T., *Theory of Vibration with Applications*, Prentice-Hall, USA, 1998.

30. Zaveri, K., *Modal Analysis of Large Structures—Multiple Exciter Systems*, Brüel & Kjær, BT 0001-12, Denmark.

Chapter 11

Fundamentals of Acoustics

Acoustics is concerned with the study of sound and mechanical waves in gases, liquids and solids and has its origin in the study of vibrations and the radiation of these vibrations as acoustic waves. It deals with all aspects of production, propagation, control, transmission, reception and effects of sound. These are applicable to sounds created and received by human beings, machines and measuring instruments. Sound is defined as any pressure variation over and above the mean atmospheric pressure impinging on the ear drum; it encompasses all sounds that the ear can detect, from the weakest sounds which are barely audible to sounds which cause pain and damage hearing.

This chapter deals with fundamental aspects of acoustics without a knowledge of which the acoustic measurement techniques which follow in the next two chapters cannot be fully appreciated. Fundamental definitions for describing sound and aspects related to the physics of sound, as well as terminology used in acoustic measurements are described. The theory behind sound propagation, transmission and the behaviour of sound in enclosed spaces has also been discussed. In this chapter complex quantities are shown in bold face, a notation convenient for equations in acoustics, and not followed in the other chapters.

11.1 HUMAN PERCEPTION OF SOUND

We necessarily have to deal with sound in our day to day life, whether it be spoken communication, music or noise produced by machines. The study of acoustics includes what is called sound, as well as noise. From the point of view of acoustics, both sound and noise are caused by pressure fluctuations about the mean atmospheric value. There is a big difference though, in the way the two are sensed and this is also highly subjective. What one person perceives as sound may very well be perceived as noise by somebody else. In general, noise is defined as any unpleasant or undesirable sound. Typically sounds which have a structured waveform or spectrum are perceived by the ear as being pleasant and would be called music, while those which are unstructured sound unpleasant to the ear and are called noise. Music may be considered to be sound, not noise, since it is pleasant to most people; noise from machinery on the other hand gives pleasure to none other than the owner of the machinery, to whose ears it may sound like music since it is bringing him money! The level of annoyance felt due to noise is highly subjective, varying from person to person, time of the day, health condition, mood and so on. It also depends on the pitch, loudness and tonal quality of the sound. Higher the sound intensity, higher is the perceived loudness. Typically higher sound levels are tolerated better during the day than at night. Sounds can also cause damage, examples being sonic booms from aircraft or the music of a soprano singer which can shatter windows and glasses.

Noise has become an unavoidable part of day to day life with noise emanating from machines, factories, traffic, etc. For the study of human exposure to noise and for the establishment of noise criteria, it is required to have a good understanding of the way the human ear responds to it. Noise is increasingly being perceived as a pollutant and a serious health hazard these days. Studies have shown that in addition to inducing a threshold shift (hearing loss, possibly restricted to distinct frequency ranges), noise also causes irritability, reduces concentration, increases blood pressure, affects blood circulation, causes cardiac problems and changes resistance of the skin to electric potential. It is therefore important that steps towards noise control are taken. In connection with this drive against noise, legislations have been made to control specific noise-making activities, based on different kinds of zones. In India, allowable noise limits for each zone are set by the Central Pollution Control Board of India as described in Section 13.9.1.

11.1.1 Sound Pressure and Sound Pressure Level

The sensation of sound is caused by pressure fluctuations or oscillations in air. The response of the human ear to sound or noise depends on the sound frequency, the sound pressure level (SPL) and the pressure waveform; besides there is a high degree of subjectivity involved. For a young human ear which is in good health and which has not been affected by too much exposure to excessively loud sounds, the range of these pressure amplitudes for audio acoustics is very large, from 2×10^{-5} N/m^2 or 20 μPa, the threshold of audibility (roughly the sound of a mosquito flying 3 m away) to 20–63 N/m^2, the threshold of pain. The acoustic pressures involved in audio-acoustics are thus seen to be very small fluctuations about the mean atmospheric pressure which has a value of 1.013×10^5 Pa at sea level. Due to the remarkably large range of pressures involved (more than a million to 1, or 6 decades), some kind of compressed scale or a logarithmic representation of this quantity (using ratios) is more appropriate than a linear representation which would mean using very large and unwieldy numbers. Besides, the ear responds logarithmically, not linearly, to acoustic stimuli, in terms of both intensities and frequencies. For these reasons, quantities like acoustic pressure, intensity and power (to be discussed later) are represented in terms of the deciBel (dB), which is the logarithm of the ratio of any of the said quantities to a reference level. SPL L_p gives an indication of the loudness of a sound and is the most important acoustic quantity that is measured. It is defined as

$$L_p = 10 \log_{10} \left(\frac{p_{\text{rms}}^2}{p_{\text{ref}}^2} \right) = 20 \log_{10} \left(\frac{p_{\text{rms}}}{p_{\text{ref}}} \right) \tag{11.1}$$

where p_{ref} is the reference sound pressure corresponding to the threshold of audibility of human hearing at the acoustic reference frequency of 1000 Hz and is equal to 2×10^{-5} N/m^2 or 20 μPa; p_{rms} refers to the root mean square value of the pressure being measured. A pressure equal to the reference value works out to 0 dB, while 1 Pa is equivalent to about 94 dB. Table 11.1 gives the SPLs due to some common activities.

Table 11.2 gives the relationship between pressure ratio, power (proportional to pressure squared) ratio and change in SPL in dB. Though the loudness perceived by human beings is correlated to the SPL in dB, which is related to intensity or pressure squared, there is no linear relationship between the two. An increase in SPL of 3 dB, corresponding to a pressure ratio of 1.4 or a doubling of sound energy (pressure squared) is just perceptible to the normal human ear. Thus if a manufacturer makes claims that

his product produces a sound pressure only half that of his rival's product, one should be wary; the change in SPL would be noticeable, but not significantly lower. A change of 10 dB (corresponding to a pressure ratio of 3.16) is perceived as twice as loud. Table 11.3 summarizes the subjective perception of changes in SPL in dB.

TABLE 11.1   Typical sound pressure levels L_p

Sound source	Sound pressure (Pa)	Sound pressure level L_p (dB)
Sound wave at 1 atmosphere (theoretical limit)	101,325	194
Large rocket engine (nearby)	20,000	180
Jet engine at 30 m	630	150
Threshold of pain/Deepavali atom bomb	63	130
Jet takeoff (60 m)	20	120
Jack hammer at 1 m/discotheque	2	100
Major road traffic at 10 m	6×10^{-1}	90
Moving automobile at 10 m	6×10^{-2}	70
Normal conservation	2×10^{-3}–2×10^{-2}	40–60
Quiet residential area	2×10^{-3}	40
Soft whisper (5 m)	6×10^{-4}	30
Normal breathing	6×10^{-5}	10
Threshold of audibility for a healthy young person	2×10^{-5}	0

TABLE 11.2   Relationship between power ratio, pressure ratio and change in dB

Power ratio n	Pressure ratio $n^{1/2}$	Change in sound pressure level L_p (dB) = $10log_{10} n$ = $20log_{10}n^{1/2}$
10,000	100	40
100	10	20
4	2	6
2	1.4	3
1	1	0
0.5	0.7	–3
0.25	0.5	–6
0.01	0.1	–20
0.0001	0.01	–40

TABLE 11.3   Perception of loudness related to L_p (dB)

Change in SPL (dB)	Change in perceived loudness
1	Insignificant
3	Just perceptible
5	Clearly perceptible
10	Twice or half as loud
15	Significant change
20	Four times or 1/4 as loud

11.1.2   Frequencies of Interest and Frequency Weighting

Normally, healthy young human beings can detect sounds with frequencies in the range 20 to 20,000 Hz. With age, the sensitivity to higher frequencies reduces. Pitch is related to frequency; low frequencies are perceived as being low-pitched, while high frequencies are identified as being high-pitched. Pitch is thus the subjective response of the human ear to the spectral content of sound. The frequencies involved in the study of sound encompass the following ranges:

 (i) Sonic: 20 Hz–20,000 Hz or the audio-frequency range.
 (ii) Ultrasonic: above 20,000 Hz; this is not considered damaging at levels below 105 dB.
(iii) Infrasonic (subsonic): below 20 Hz; these sounds are felt, but not heard and are damaging at levels above 120 dB. Infrasounds as well as ultrasounds affect human senses and cause discomfort.

   The sensitivity of the human ear to different frequencies varies, with minimum sensitivity to extremely low and extremely high frequencies. For example, a pure tone of 1000 Hz with an SPL of 40 dB would sound louder than a pure tone of 80 Hz at 50 dB, and a 1000 Hz tone at 70 dB would give the same subjective perception of loudness as a 50 Hz tone at 85 dB. The minimum audible sound level occurs at about 4000 Hz. The frequency range between 500 Hz and 2 kHz is most crucial to human beings since speech signals are in this range.

   Frequency analysis is required for noise source identification. There are a number of instruments available for carrying out spectrum analysis of acoustic signals as described in Chapter 13. For the evaluation of human exposure to noise, the sound measuring system must take into account the varying sensitivity of the ear over the audio frequency range. This is done using internationally standardized frequency weighting networks such as the *A*, *B*, *C* and *D* networks (as per IEC 60651 which is now obsolete) that weight the contributions of the different frequencies to the subjective perception of loudness of sound while arriving at the overall SPL. The two weighting networks most commonly used are the *A* and *C* networks, which correlate with the frequency response of the human ear for different sound levels. Their characteristics are specified in IEC 61672 and have been described in Section 13.2.1. The B and D networks are obsoltete now and are not specified in this standard. While frequency analysis of vibration signals is done using constant bandwidth narrow band filters or FFT algorithms, constant percentage bandwidth analyzers are used for acoustic signals for the reason that the human ear senses frequencies in a logarithmic fashion. In order to maintain uniformity of measurements using different instruments for frequency analysis, 1/1 and 1/3 octave bands have been standardized by the International Standards Organization; these are described in Section 13.2.2.

11.2   SOUND WAVE PROPAGATION IN 1-D

When a sound source such as a tuning fork vibrates, it causes pressure variations in the surrounding air. These fluctuations spread out from the source and are similar to the ripples in a pond caused by a stone thrown in the water. Though the fluctuations move, the air itself does not move away from the source. Such vibrational disturbances that propagate in a fluid or gaseous medium are called acoustic waves. In such wave motion, pressure disturbances may be generated by a vibrating surface or by turbulent fluid flow and they propagate as longitudinal or compressional waves in the elastic medium (or stress waves in a solid). The particles oscillate back and forth in the medium in the direction of the waves, resulting in alternate compressions and rarefactions of the particles of the medium as the sound wave passes a given location (Fig. 11.1). The rate at which these disturbances travel though the medium depends on the

speed with which the molecules transfer energy from one to another. The study of sound wave propagation with air as the medium constitutes audio acoustics, while that in water is called under water acoustics. Three basic elements are involved in sound wave propagation: source, medium and receiver. It is to be noted that sound cannot propagate in a vacuum. If the medium is infinite, the waves will propagate in all directions and get attenuated rather quickly. If the sound waves are constrained to travel in one direction only, as happens when they propagate in a fluid-filled narrow tube, the waves travel without alteration in shape or size if the disturbances are small in the first place and if the dissipation in the medium is negligible.
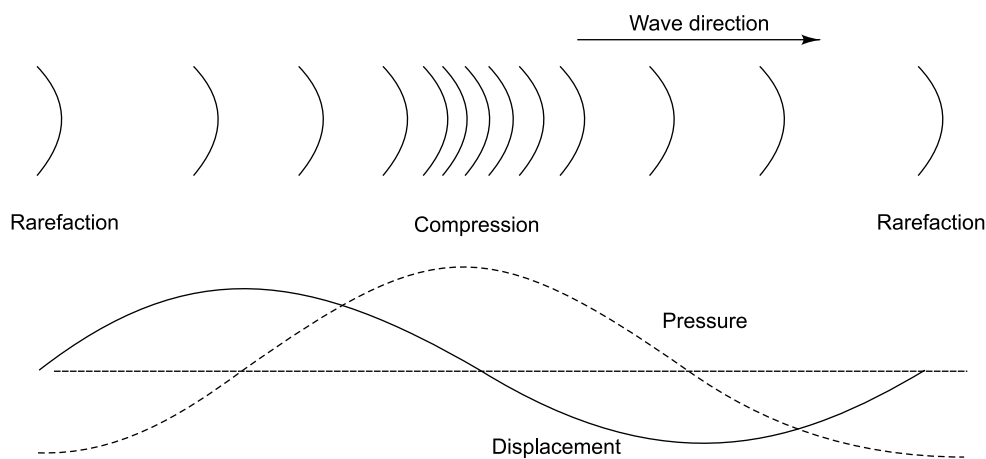
FIGURE 11.1   Longitudinal wave

11.2.1   Longitudinal Waves in a Column of Gas

The most popular model for analyzing wave motion is the particle model. A particle is defined as an exceedingly small element consisting of several million molecules with uniform density, velocity and temperature throughout the element. What typically happens in wave propagation is that a disturbance imparts momentum to some adjacent particles. These particles collide with other particles, transferring momentum to them and setting them into motion. The speed with which the wave propagates is known as the characteristic speed of sound in the medium. Every particle, after colliding with an adjacent particle, oscillates about its equilibrium position and the amplitude of oscillation can be sensed by measuring the sound pressure, which is nothing but the variation over and above the static pressure in the fluid. The equation representing such disturbances is called the wave equation in 1-D. The assumptions made in deriving this equation are:

  (i) The fluid in which the wave propagates is homogeneous and isotropic.
 (ii) The fluid is in thermodynamic equilibrium, except near the propagating wave.
(iii) The fluid is an ideal gas.
(iv) It obeys Hooke's law (has linear stress-strain behaviour).
 (v) The body force acting on the fluid is negligible.

(vi) Damping forces are negligible.

(vii) The thermodynamic process involved in wave propagation is adiabatic and reversible.

(viii) The motion involved in propagation is so small that products resulting in second and higher order terms are negligible.

Let us derive the equation of motion for longitudinal waves in a gas. Let us assume that the tube in which it is propagating is so narrow that at any instant, the displacements at all points on any cross-sectional area S are the same. It is also assumed that there is no friction along the sides of the tube. Figure 11.2 shows the displacement of an element of gas in a tube of cross-sectional area S due to longitudinal wave propagation.

The wave equation may be obtained using Newton's law of motion and the elastic behaviour of the gas. In this figure x and $x + \xi$ indicate the original and displaced positions of the left plane of the element along the tube. AB indicates the undisturbed element and $A'B'$ that due to the passage of the longitudinal wave. The increase in the volume of the element is $S(\partial \xi / \partial x)dx$.

FIGURE 11.2 Displacement of an element due to wave propagation

(i) Newton's law of motion

The total force acting on the displaced element $A'B'$ is

$$F = (P_{A'} - P_{B'})S \tag{11.2a}$$

Here P denotes the total pressure which can be written as a dynamic pressure p superimposed over the static pressure P_0. Therefore the force in the positive x direction is

$$F = -\frac{\partial p}{\partial x}dxS \tag{11.2b}$$

It is seen that P_0 disappears from the differential equation. From the law of conservation of mass, we have

Mass of element $A'B'$ = mass of element $AB = S\rho_0 dx$ \hfill (11.3)

where ρ_0 is the mean density of gas. The acceleration of the element is $\partial^2 \xi / \partial t^2$. Substituting this in Eq. (11.2b), we can write Newton's law as

$$S\rho_0 dx \frac{\partial^2 \xi}{\partial t^2} = -\frac{\partial p}{\partial x}dx\,S$$

$$\frac{\partial p}{\partial x} = -\rho_0 \frac{\partial^2 \xi}{\partial t^2} = -\rho_0 \frac{\partial u}{\partial t} \tag{11.4}$$

Here u is the particle velocity.

(ii) Elastic behaviour

The change in pressure of the gas results in a change in its volume due to its compressibility. Since the frequencies of pressure fluctuations associated with acoustic waves are high, in the range 20Hz–20 kHz for audio-acoustics, the thermodynamic process is assumed to be adiabatic. Hence,

$$P_0 V_0^\gamma = K \qquad (11.5)$$

where V_0 is the undisturbed volume, γ is the ratio of the specific heats of the gas and K is a constant. Differentiating Eq. (11.5), we get

$$\gamma P_0 V_0^{\gamma-1} dV + V_0^\gamma dP = 0 \qquad (11.6)$$

dP in the above equation is the same as the dynamic pressure p. Hence, the above equation may be written as

$$p = -\gamma P_0 \frac{dV}{V_0} \qquad (11.7)$$

Here the undisturbed volume V_0 is equal to Sdx and the change in volume dV is $S(\partial\xi/\partial x)dx$. Therefore

$$p = -\gamma P_0 \frac{\partial\xi}{\partial x} \qquad (11.8)$$

Differentiating the above equation with aspect to x and combining with Eq. (11.4) to eliminate p, we get

$$\gamma P_0 \frac{\partial^2\xi}{\partial x^2} = \rho_0 \frac{\partial^2\xi}{\partial t^2}$$

or

$$\frac{\partial^2\xi}{\partial x^2} = \frac{1}{c_0^2}\frac{\partial^2\xi}{\partial t^2} \qquad (11.9)$$

where

$$c_0^2 = \gamma P_0/\rho_0 \qquad (11.10)$$

It has been assumed here that the fluctuating part of wave speed is negligible. The uni-directional wave equation described above is in terms of dynamic displacement ξ. However, the quantity of interest in acoustics is dynamic pressure p which is related to the displacement by Eq. (11.8). Substituting $\rho_0 c_0^2$ for γP_0 in this equation, we get

$$p = -\rho_0 c_0^2 \frac{\partial\xi}{\partial x} \qquad (11.11)$$

The wave equation may also be obtained directly in terms of pressure by differentiating Eq. (11.4) with respect to x as shown below.

$$\frac{\partial^2 p}{\partial x^2} = -\rho_0 \frac{\partial^2}{\partial t^2}\left(\frac{\partial\xi}{\partial x}\right) \qquad (11.12)$$

Substituting for $\partial\xi/\partial x$ from Eq. (11.11) into the above equation gives

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{c_0^2}\left(\frac{\partial^2 p}{\partial t^2}\right) \qquad (11.13)$$

Thus the 1-D wave equation has the general form

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{1}{c_0^2} \frac{\partial^2 \phi}{\partial t^2} \tag{11.14}$$

The quantity ϕ may represent the pressure amplitude or particle displacement. This equation is the same as that for the transverse vibration of a string where the displacement is perpendicular to the direction in which the wave is travelling. It is the wave equations in terms of particle pressure (Eq. 11.13) and displacement (Eq. 11.9) which have exactly the same form for 1-D propagation; the former is the most commonly used form of wave equation in acoustics.

Equation (11.14) may be solved by separation of variables. It may be written as shown below, to give two separate differential equations.

$$\phi = F(x)\,G(t) \tag{11.15}$$

The general solution for wave propagation would then take the form

$$\phi = g(c_0 t - x) + h(c_0 t + x) \tag{11.16}$$

Here g and h are arbitrary, independent functions. The first term in the solution represents a wave of constant shape moving in the positive x direction with velocity c_0, while the second term represents a wave of constant shape travelling in the negative x direction with velocity c_0. Let us consider a harmonic wave, which is the simplest solution, travelling in the positive x direction. It may be represented as

$$\phi = a_1 \cos k\,(c_0 t - x) \tag{11.17}$$

where a_1 may represent the peak pressure and k is the wavenumber and is related to the wavelength λ as shown below.

$$k = \frac{2\pi}{\lambda} \tag{11.18a}$$

The period T of the wave is given by

$$T = \frac{2\pi}{kc_0} = \frac{\lambda}{c_0} \tag{11.18b}$$

and the frequency in Hz is

$$f = \frac{1}{T} = \frac{c_0}{\lambda} \text{ or } c_0 = f\lambda \tag{11.18c}$$

It is good to have a rough idea regarding correspondence between wavelength and frequency. Considering that the speed of sound in air is 343 m/s, at 1 kHz the wavelength is close to 34 cm, at 20 Hz it is close to 17 m and at 20 kHz it is only 1.7 cm.

Writing Eq. (11.17) in terms of angular frequency $\omega = 2\pi f$ in rad/s, we have

$$\phi = a_1 \cos\,(\omega t - kx) \tag{11.19}$$

A more general solution to the forward travelling wave may be written in terms of a combination of sine and cosine waves; this can describe any desired harmonic waveform by an appropriate choice of a_1 and a_2.

$$\phi = a_1 \cos\,(\omega t - kx) + a_2 \sin(\omega t - kx) \tag{11.20}$$

A general solution consisting of both the forward and backward travelling waves is

$$\phi = a_1 \cos(\omega t - kx) + a_2 \sin(\omega t - kx) + a_3 \cos(\omega t - kx) + a_4 \sin(\omega t - kx) \tag{11.21a}$$

This harmonic solution may be written in terms of complex exponential functions as shown below.

$$\boldsymbol{\phi} = \boldsymbol{A} e^{j(\omega t - kx)} + \boldsymbol{B} e^{j(\omega t + kx)} \tag{11.21b}$$

Here A and B are, in general, complex quantities and are shown in boldface. It is only the real part of the solution which is used, the imaginary part being ignored though it is used in computation.

Particle velocity is the velocity of a particle in the medium as it transmits a wave and is not the same as the speed of sound. The particle velocity u can be obtained from particle pressure by writing Eq. (11.4) in the form shown below.

$$\frac{\partial u}{\partial t} = -\frac{1}{\rho_0}\left(\frac{\partial p}{\partial x}\right) \tag{11.22}$$

Integrating the above equation over t, we get

$$u = -\frac{1}{\rho_0}\int \frac{\partial p}{\partial x} dt \tag{11.23}$$

The solution to the wave equation in terms of particle pressure for the forward travelling wave only may be written as

$$p_+ = a_1 \cos(\omega t - kx) + a_2 \sin(\omega t - kx) \tag{11.24}$$

Differentiating w.r.t. x, we get

$$\frac{\partial p_+}{\partial x} = ka_1 \sin(\omega t - kx) - ka_2 \cos(\omega t - kx) \tag{11.25}$$

Integrating the above equation w.r.t. t, we get

$$\int \frac{\partial p_+}{\partial x} dt = -\frac{k}{\omega} a_1 \cos(\omega t - kx) - \frac{k}{\omega} a_2 \sin(\omega t - kx) \tag{11.26}$$

$$= -\frac{k}{\omega} p_+$$

From Eq. (11.23) it follows that

$$u_+ = \frac{k}{\omega \rho_0} p_+ = \frac{p_+}{\rho_0 c_0} = \frac{\partial \xi_+}{\partial t} \tag{11.27a}$$

$$u_- = -\frac{k}{\omega \rho_0} p_- = -\frac{p_-}{\rho_0 c_0} = \frac{\partial \xi_-}{\partial t} \tag{11.27b}$$

These expressions relate particle pressure and particle velocity. Thus the total particle pressure and velocity are as shown below.

$$p = p_+ + p_- \text{ and } u = u_+ + u_- \tag{11.28}$$

Figure 11.3 shows the phase relationships between p, u, ξ and $\partial \xi/\partial x$ for the waves travelling in the positive and negative x-directions.

Two common boundary conditions which are considered while solving are as follows:

(i) closed tube for which $(u)_x = 0$ at the closed end for all time t and

(ii) open tube for which $(p)_x = 0$ at the open end for all time t. For the open tube, the disturbances extend slightly beyond the open end, and hence the boundary condition may be applied at the imaginary end. It has been shown in literature that the end correction is $0.6r$ for a tube of radius r with no flange and $8r/3\pi$ for a tube with a large flange.
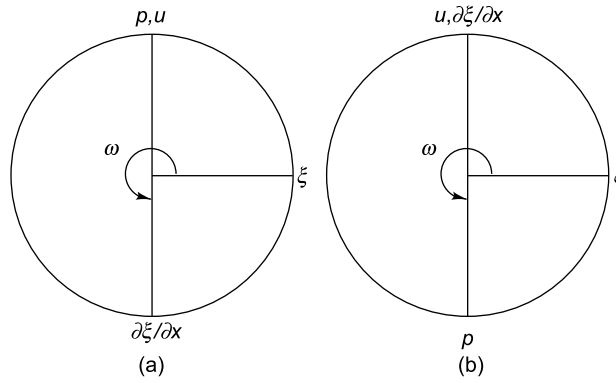
FIGURE 11.3　Phase relationships: (a) forward travelling wave, (b) backward travelling wave

Let us consider an unflanged tube of length L, closed at one end and open at the other end. The pressure and particle velocity are expressed by the following equations.

$$p = a_1 \cos(\omega t - kx) + a_2 \sin(\omega t - kx) + a_3 \cos(\omega t + kx) + a_4 \sin(\omega t + kx) \qquad (11.29)$$

$$u = \frac{a_1}{\rho_0 c_0} \cos(\omega t - kx) + \frac{a_2}{\rho_0 c_0} \sin(\omega t - kx) - \frac{a_3}{\rho_0 c_0} \cos(\omega t + kx) - \frac{a_4}{\rho_0 c_0} \sin(\omega t + kx) \qquad (11.30)$$

Applying the first boundary condition, i.e. $(u)_x = 0$ for all t, we get

$$0 = a_1 \cos(\omega t) + a_2 \sin(\omega t) - a_3 \cos(\omega t) - a_4 \sin(\omega t) \qquad (11.31a)$$

Hence, $\qquad\qquad\qquad a_3 = a_1; \; a_4 = a_2 \qquad\qquad\qquad\qquad\qquad\qquad (11.31b)$

Substituting these values, the expression for pressure in Eq. (11.29) gets simplified to

$$p = 2 \cos kx \, (a_1 \cos \omega t + a_2 \sin \omega t) \qquad (11.32)$$

This represents a stationary waveform with a pressure antinode (maximum) at the rigid end near the origin. Applying the second boundary condition, namely that the tube is open at $x = L$, or at $x = L'$ to be more precise with the application of end correction, we have for all values of t

$$(p)_{x=L'} = 0 \qquad (11.33a)$$

where $\qquad\qquad\qquad L' = L + 0.6r \qquad\qquad\qquad\qquad\qquad\qquad (11.33b)$

Hence, $\qquad\qquad\qquad \cos kL' = 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad (11.33c)$

i.e. $\qquad\qquad\qquad kL' = (2n - 1)\pi/2, \; n = 1, 2, 3... \qquad\qquad (11.33d)$

Therefore,

$$\lambda = \frac{4L'}{(2n-1)} \quad \text{or} \quad f = \frac{(2n-1)c_0}{4L'} \qquad (11.34)$$

Figure 11.4 depicts the pressure variation in an unflanged pipe of length L closed at one end and open at the other for the first four modes.
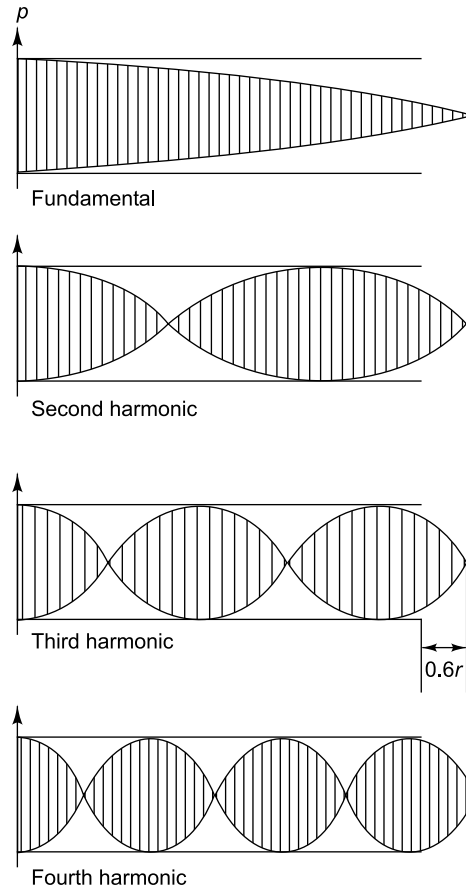
FIGURE 11.4   Pressure variation in a pipe closed at one end and open at the other

As in the case of a vibrating system, the pipe sustains natural vibrations or resonances only at certain frequencies. These natural frequencies are found to be functions of the speed of sound and the length of the tube. The first natural frequency is called the fundamental and the higher natural frequencies the harmonics. Due to the nature of the boundary conditions imposed in this case, all harmonics are odd integral multiples of the fundamental frequency.

11.3   SOUND PROPAGATION IN 3-D SPACE: THE 3-D WAVE EQUATION

Most practical acoustic problems involve propagation in 3-D; hence the plane wave theory discussed in Section 11.2 needs to be extended to the 3-D case. In this section, the general equation for propagation in 3-D is obtained in Cartesian coordinates as an extension of the 1-D equation. For applications like spherical radiation studies, spherical coordinates would be more appropriate and for cylindrical radiation, cylindrical coordinates. With rectangular coordinates, the equilibrium position of the particle may be defined in terms of coordinates x, y and z. The particle displacements in the three directions are ξ, η and

ζ and the corresponding particle velocity components are $\partial\xi/\partial t$, $\partial\eta/\partial t$ and $\partial\zeta/\partial t$, respectively. We start off with a gas element in 3-D as shown in Fig. 11.5.
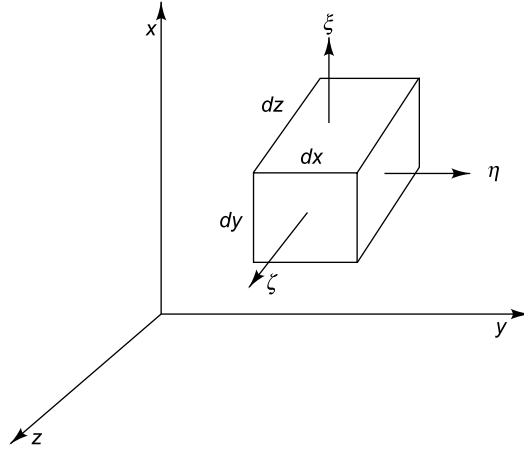
Figure 11.5   Gas element in 3-D

Since the acceleration in each direction is proportional to the force in that direction alone, the motion in each direction may be written as

$$\frac{\partial p}{\partial x} = -\rho_0 \frac{\partial^2 \xi}{\partial t^2}; \ \frac{\partial p}{\partial y} = -\rho_0 \frac{\partial^2 \eta}{\partial t^2}; \ \frac{\partial p}{\partial z} = -\rho_0 \frac{\partial^2 \zeta}{\partial t^2} \tag{11.35}$$

The three equations may be differentiated, each with respect to its own coordinate and then added to get

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = -\rho_0 \frac{\partial^2}{\partial t^2} \left(\frac{\partial \xi}{\partial x} + \frac{\partial \eta}{\partial y} + \frac{\partial \zeta}{\partial z} \right) \tag{11.36}$$

Equation (11.7) developed for the 1-D case remains unaltered. The increased volume of the element due to the passage of the sound wave is

$$V + dV = \left[\left(1 + \frac{\partial \xi}{\partial x} \right) \right] \left(1 + \frac{\partial \eta}{\partial y} \right) \left(1 + \frac{\partial \zeta}{\partial z} \right) dxdydz \tag{11.37}$$

where $V_0 = dxdydz$ is the original volume.

and

$$\frac{dV}{V_0} = \left(\frac{\partial \xi}{\partial x} + \frac{\partial \eta}{\partial y} + \frac{\partial \zeta}{\partial z} \right) \tag{11.38}$$

Therefore,

$$p = -\gamma P_0 \left(\frac{\partial \xi}{\partial x} + \frac{\partial \eta}{\partial y} + \frac{\partial \zeta}{\partial z} \right) \tag{11.39}$$

Substituting the above equation into Eq. (11.36) gives

$$\frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} + \frac{\partial^2 p}{\partial z^2} = \frac{1}{c_0^2} \frac{\partial^2 p}{\partial t^2} \tag{11.40}$$

or

$$\nabla^2 p = \frac{1}{c_0^2} \frac{\partial^2 p}{\partial t^2} \tag{11.41a}$$

where

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{11.41b}$$

This is very similar to the 1-D wave equation. The particle velocity in terms of the particle pressure may be obtained as

$$u = \left(\frac{\partial \xi}{\partial t} + \frac{\partial \eta}{\partial t} + \frac{\partial \zeta}{\partial t} \right) = -\frac{1}{\rho_0} \int \left(\frac{\partial p}{\partial x} + \frac{\partial p}{\partial y} + \frac{\partial p}{\partial z} \right) dt \tag{11.42}$$

or

$$u = -\frac{1}{\rho_0} \int \mathrm{grad}(p) \, dt \tag{11.43}$$

In the case of spherical waves, it may be shown that

$$\left(\frac{\partial^2 p}{\partial r^2} + \frac{2}{r} \frac{\partial p}{\partial r} \right) = \frac{1}{c_0^2} \frac{\partial^2 p}{\partial t^2} \tag{11.44}$$

or

$$\left(\frac{\partial^2 (rp)}{\partial r^2} \right) = \frac{1}{c_0^2} \frac{\partial^2 (rp)}{\partial t^2} \tag{11.45}$$

This equation represents an omni-directional wave radiating out from the source. It is a function of time t and radial distance r from the source and its solution may be written as

$$p(r, t) = \frac{g(c_0 t - r)}{r} + \frac{h(c_0 t + r)}{r} \tag{11.46}$$

Here the first term represents the wave radiating outward from the source; it gets attenuated as a function of $1/r$. The second term represents the backward wave travelling towards the source. For spherical waves which are harmonic, $p(r, t)$ in the most general complex form can be written as

$$p(r, t) = \frac{P_+}{r} e^{i(\omega t - kr)} + \frac{P_-}{r} e^{i(\omega t + kr)} \tag{11.47}$$

where P_+ and P_- represent the peak amplitudes of the forward and backward travelling waves, respectively.

On the lines of the equation $u = -\dfrac{1}{\rho_0}\displaystyle\int \dfrac{\partial p}{\partial x}\,dt$ for longitudinal waves in a column of gas, we have the relation between acoustic pressures and particle velocity for the outward travelling spherical wave.

$$u(r,\,t) = -\frac{1}{\rho_0}\int \frac{\partial p(r,\,t)}{\partial r}\,dt \tag{11.48}$$

Substituting for the term corresponding to the forward travelling wave into the above equation and noting that $\rho_0\omega = \rho_0 c_0 k$, we get

$$u(r,\,t) = \frac{p(r,\,t)}{\rho_0 c_0}\left(1 + \frac{1}{ikr}\right) \tag{11.49}$$

There are two regions of interest with regard to the equation relating $u(r,t)$ and $p(r,t)$. They are the regions described by: (i) $kr \ll 1$ and (ii) $kr \gg 1$ as discussed in Section 11.4.7.

11.4 Some Important Acoustic Quantities and Relations

In this section quantities of importance like velocity of sound, acoustic impedance, intensity, power and radiation fields of a sound source are described. Sound pressure, intensity and power levels are defined and the relationships between them are obtained. The additive effects of sources and the radiation fields of a source are also described.

11.4.1 Velocity of Sound

For small disturbances, the velocity of sound in a gas is

$$c_0 = \left(\frac{\gamma P_0}{\rho_0}\right)^{1/2} \tag{11.50a}$$

where γ is the ratio of specific heats of the gas, P_0 is the mean pressure and ρ_0 is the mean density. γ appears in the equation because the thermodynamic process involved is adiabatic due to the rapid fluctuations of pressure. P_0/ρ_0 is almost constant for all gases and the velocity of sound is independent of pressure.

Substituting $P_0/\rho_0 = RT/M$ where R is the gas constant, M is the molecular weight and T is the absolute temperature in Kelvin (K), we get

$$c_0 = \left(\frac{\gamma RT}{M}\right)^{1/2} \tag{11.50b}$$

The velocity of sound is thus seen to be proportional to the square root of the absolute temperature. For propagation in liquids

$$c_0 = \left(\frac{B_a}{\rho_0}\right)^{1/2} \tag{11.51}$$

where B_a is the adiabatic bulk modulus of elasticity and ρ_0 is the density. For propagation in solids

$$c_0 = \left(\frac{E}{\rho_0} \right)^{1/2} \tag{11.52}$$

where E is the modulus of elasticity and ρ_0 is the density.

11.4.2 Characteristic Impedance and Specific Acoustic Impedance

When a 1-D wave strikes a flat surface, the impedance it faces is called the specific acoustic impedance z, which is defined as the ratio of acoustic pressure to particle velocity.

$$z = \frac{p}{u} = r + ix \tag{11.53}$$

where r is called the specific acoustic resistance and x is called the specific acoustic reactance. Since in the general case of plane wave propagation in a medium, the forward and backward travelling waves are present, p and u in the above equation will have contributions from both terms and will be a function of position. For a wave impinging on a rigid wall, the particle velocity normal to the wall is zero, or the specific acoustic impedance of that component of the wave perpendicular to the wall is infinite. For standing waves or for diverging waves, specific acoustic impedance is a complex quantity. However, in the case of plane wave propagation in an infinite medium (plane progressive waves), we have only the forward travelling wave and the specific acoustic impedance is a real constant of magnitude $\rho_0 c_0$ as shown below.

$$z_0 = \frac{p}{u} = \pm \rho_0 c_0 \tag{11.54}$$

z_0 is a function purely of the medium and has greater significance as a characteristic property of the medium than either ρ_0 or c_0 alone. It is therefore called characteristic impedance and has the unit kg/m^{2}s or MKS Rayl. The velocity c_0 and the wavenumber k are in general, complex quantities, the imaginary parts of which contribute to the resulting absorption in the medium as the wave propagates. If absorption is small, the velocity and wavenumber are wholly real and the characteristic impedance is also wholly real. Table 11.4 shows the velocity of sound in some common materials that we come across and their characteristic impedances.

TABLE 11.4   Speed of sound in different media

Material	Speed (m/s)	Characteristic impedance $\rho_0 c_0$ (MKS Rayls)
Air (at STP)	335	428
Lead	1128	23.2×10^6
Concrete	3109	8.0×10^6
Fresh water	1385	1.48×10^6
Steel	4925	47×10^6
Soft wood	3417	1×10^6
Glass	4771	12.9×10^6

11.4.3 Energy Density and Intensity

Acoustic pressure is the most meaningful quantity since our ears and acoustic transducers sense this scalar quantity. Besides it is SPL which is used for evaluation of the harmfulness and annoyance of noise

sources. But sound energy being a much more fundamental quantity, it becomes imperative to derive a relationship between the two. Energy density D is defined as the energy per unit volume of the fluid and is a function of both space and time and consists of two parts: (i) kinetic energy of the particles and (ii) potential energy of the gas. Knowing the particle velocity, u, the kinetic energy is obtained as

$$D_{KE} = \frac{1}{2}\rho_0 u^2 \tag{11.55}$$

When gas of volume V_0 undergoes an expansion dV, the potential energy it gains can be expressed as

$$D_{PE} = \frac{1}{V_0}\int p\,dV \tag{11.56}$$

It is to be remembered that in the course of sound wave propagation, the gas undergoes consecutive adiabatic compressions and rarefactions. Differentiating the adiabatic law $PV^\gamma = K$, we get $\dfrac{dV}{V_0} = -\dfrac{1}{\gamma}\dfrac{dp}{P_0}$ as described in Eq. (11.7). Substituting this into Eq. (11.56), we get the expression for potential energy as

$$D_{PE} = \frac{1}{\gamma P_0}\int p\,dp = \frac{1}{2}\frac{p^2}{\gamma P_0} = \frac{1}{2}\frac{p^2}{\rho_0 c_0^2} \tag{11.57}$$

Hence, the total energy density is

$$D(x,t) = \frac{1}{2}\left(\rho_0 u^2 + \frac{p^2}{\rho_0 c_0^2}\right) \tag{11.58}$$

The sound intensity I may be defined as the rate at which work is done on a conducting medium by an advancing sound wave and is thus the rate of flow of sound energy through unit area of the gas normal to the intensity vector; its unit is W/m^2. This quantity is again a function of space and time and it is a vector since the direction of flow of energy is defined (away from the source). The main use of sound intensity is for location of noise sources. Intensity can be thought of as the rate at which the acoustic pressure does work and is the product of the pressure and the component of particle velocity which is in phase with it.

$$I(x,t) = pu \tag{11.59}$$

The instantaneous pressure of a plane progressive wave for the forward travelling component may be written in the simplest form as

$$p = a_1 \cos(\omega t - kx) \tag{11.60}$$

The instantaneous particle velocity for the forward travelling component is given by

$$u = \frac{p}{\rho_0 c_0} \tag{11.61}$$

It may be noted that particle pressure and velocity are in phase. The energy density at any instant of time is got by substituting Eqs (11.60) and (11.61) into (11.58).

$$D(x,t) = \frac{1}{2}\left(\frac{p^2}{\rho_0 c_0^2} + \frac{p^2}{\rho_0 c_0^2}\right)$$

$$= \frac{a_1^2}{\rho_0 c_0^2} \cos^2(\omega t - kx) \tag{11.62}$$

From the instantaneous value of energy density described by the above equation, the average value may be got by averaging over one period, i.e. integrating over one period and dividing by that period. The energy density will be the same at all points in space for a plane progressive wave. Hence

$$D(x,t) = \frac{a_1^2}{\rho_0 c_0^2} \frac{1}{T} \int_0^T \cos^2(\omega t - kx)\, dt = \frac{a_1^2}{2\rho_0 c_0^2} \tag{11.63}$$

Writing the above equation in terms of the RMS pressure rather than in terms of the peak value, we get

$$p^2{}_{\text{rms}} = \frac{1}{T} \int_0^T a_1^2 \cos^2(\omega t - kx)\, dt \;=\; \frac{a_1^2}{2} \tag{11.64}$$

The average energy density in terms of the RMS pressure is

$$D = \frac{p_{\text{rms}}^2}{\rho_0 c_0^2} \tag{11.65}$$

The intensity is got by substituting for pressure and particle velocity in Eq. (11.59) as

$$I(x,t) = \frac{a_1^2}{\rho_0 c_0} \cos^2(\omega t - kx) \tag{11.66}$$

On averaging this over one period, we get

$$I = \frac{a_1^2}{2\rho_0 c_0} \;=\; \frac{p_{\text{rms}}^2}{\rho_0 c_0} \tag{11.67}$$

Comparing Eqs (11.65) and (11.67) it is seen that
$$I = Dc_0 \tag{11.68}$$

11.4.4 Sound Power

Acoustic pressure and intensity depend on the distance from the source, especially for 3-D or spherical waves. For a point source the sound pressure drops to half its value when the distance from the source is doubled and this corresponds to a drop in SPL of 6 dB. Hence, while specifying these quantities, the distance from the source must be mentioned. Besides, these quantities also depend on the reflective properties of the spaces surrounding the source. Hence, it is desirable to have a quantity that describes the acoustical characteristics of a source and which can be used to get intensity or pressure in different acoustical environments and at different distances from the source. Sound power happens to be such a quantity and is used for the noise rating of machines; its value does not depend on the distance or location around a source.

Sound power W is defined as the total sound energy (Joules) emitted by a source per unit time and is measured in Watts. It is obtained by integrating the sound intensity over an imaginary surface area S surrounding a source. If $\bar{I}$ represents the intensity vector, then the sound power radiated by the source is

$$W \;=\; \int_S \bar{I} \cdot d\,\bar{S} \tag{11.69}$$

The dot product is used since the area over which intensity is integrated is perpendicular to the flow of acoustical energy. For a sound source producing uniform spherical waves (or radiating equally in all directions), a spherical surface is most convenient, but sometimes other surfaces are chosen, as dictated by the circumstances for the particular case considered. The acoustic power for a source radiating spherically can be found knowing the intensity I at distance r from the source. Thus

$$W = \int_S \frac{p_{rms}^2(r)dS}{\rho_0 c_0} = \int_S \frac{p_{rms}^2(r)4\pi r^2}{\rho_0 c_0} = 4\pi r^2 I \tag{11.70}$$

For a surface composed of irregular subareas, the acoustic power is obtained is

$$W = \sum_{i=1}^{n} I_i S_i \tag{11.71}$$

where I_i is the intensity of the ith segment and S_i is the surface area perpendicular to the energy flow through the ith segment. Sound power can thus be computed from measured values of sound pressure or sound intensity levels knowing the area over which the measurements were made.

11.4.5 Levels

In view of the fact that sound pressures, intensities and powers encompass a large number of orders of magnitude, these quantities are expressed in terms of levels in dB.

11.4.5.1 Sound intensity level The range of acoustic intensities that the human ear is sensitive to is 10^{-12} W/m^2, corresponding to the threshold of audibility to 10^2 W/m^2, or a range of 14 decades. The intensity level L_I is defined as

$$L_I = 10 \log_{10}\left(\frac{I}{I_{ref}}\right) \text{dB} \tag{11.72a}$$

Here
$$I_{ref} = \frac{p_{ref}^2}{(\rho_0 c_0)_{ref}} \tag{11.72b}$$

$I_{ref} = 10^{-12}$ W/m^2 and $(\rho_0 c_0)_{ref} = 400$ MKS Rayls.

11.4.5.2 Sound power level The sound power level, L_w may be defined as follows:

$$L_I = 10 \log_{10}\left(\frac{W}{W_{ref}}\right) \tag{11.73a}$$

Here the power, W, is measured in Watts with W_{ref} the internationally agreed reference power being 10^{-12} W.

$$W_{ref} = I_{ref} S_{ref}, \text{ where } S_{ref} = 1 \text{ m}^2. \tag{11.73b}$$

Table 11.5 gives the sound power and power levels of some common sources.

11.4.5.3 Relation between sound pressure, intensity and power levels Amongst the three levels mentioned, it is only SPL which can be measured easily; sound intensity and power levels can be calculated from it. The relation between L_p, L_I and L_W can be obtained starting with Eq. (11.72) for sound intensity.

TABLE 11.5   Sound power and sound power levels of some sound sources

Sound source	Sound power (Watts)	Sound power level L_W dB; re 10^{-12} W
Rocket engine	10^6	180
Turbojet engine	10^4	160
Siren	10^3	150
Heavy truck engine	100	140
Machine gun	10	130
Trumpet	0.3	115
Chain saw	0.1	110
Helicopter	0.01	100
Loud speech	10^{-3}	90
Usual talking	10^{-5}	70
Refrigerator	10^{-7}	50
Auditory threshold	10^{-12}	0

$$\frac{I}{I_{\text{ref}}} = \frac{p_{\text{rms}}^2}{p_{\text{ref}}^2} \cdot \frac{(\rho_0 c_0)_{\text{ref}}}{\rho_0 c_0} \tag{11.74}$$

Taking logarithm of both sides of the above equation

$$10 \log_{10}\left(\frac{I}{I_{\text{ref}}}\right) = 10 \log_{10}\left(\frac{p_{\text{rms}}^2}{p_{\text{ref}}^2}\right) + 10 \log_{10}\left[\frac{(\rho_0 c_0)_{\text{ref}}}{\rho_0 c_0}\right] \tag{11.75}$$

This can be simplified to

$$L_I = L_p - 10 \log_{10}\left[\frac{\rho_0 c_0}{(\rho_0 c_0)_{\text{ref}}}\right] \tag{11.76}$$

$(\rho_0 c_0)$ for air at 20 °C and 50% Relative Humidity and at a standard atmospheric pressure of 1.013×10^5 Pa is 415 MKS Rayls and $(\rho_0 c_0)_{\text{ref}}$ = 400 Rayls. Therefore,

$$L_1 = L_p - 0.16 \tag{11.77}$$

For the atmospheric pressure and temperature conditions stated above, L_I and L_p are almost the same, but not for other conditions. The equation for sound power level may be derived as shown below.

$$\frac{W}{W_{\text{ref}}} = \frac{p_{\text{rms}}^2 \, 4\pi r^2}{\rho_0 c_0} \frac{(\rho_0 c_0)_{\text{ref}}}{p_{\text{ref}}^2 \, 4\pi r_{\text{ref}}^2} = \frac{p_{\text{rms}}^2}{p_{\text{ref}}^2} \frac{r^2}{r_{\text{ref}}^2} \frac{(\rho_0 c_0)_{\text{ref}}}{\rho_0 c_0} \tag{11.78}$$

Taking logarithm on both sides,

$$10 \log_{10}\left(\frac{W}{W_{\text{ref}}}\right) = 10 \log_{10}\left(\frac{p_{\text{rms}}^2}{p_{\text{ref}}^2}\right) + 10 \log_{10}\left(\frac{r^2}{r_{\text{ref}}^2}\right) - 10 \log_{10}\left[\frac{\rho_0 c_0}{(\rho_0 c_0)_{\text{ref}}}\right] \tag{11.79}$$

With $S_{\text{ref}} = 1\,\text{m}^2$ $r_{\text{ref}} = 0.282\,\text{m}$. Hence, for standard atmospheric temperature and pressure,

$$L_W = L_p + 20 \log_{10}\left(\frac{r}{r_{\text{ref}}}\right) - 0.16 \tag{11.80}$$

11.4.6   Additive Effects of Sound

Often, the total SPL due to sounds from many sources has to be evaluated. In general, the sounds are of different frequencies with random phases between them and the sources are then said to be incoherent or uncorrelated. An example of uncorrelated sources is the sound from two typewriters operated by different people. However, when the additive effect of sounds of the same frequency is to be found, the phase between them must be considered in the calculation. The sources are said to be fully correlated if they are producing the same waveform as a function of time, i.e. they have the same frequency and phase at a particular point. The cumulative effect can then be found by simply adding the pressures. Sometimes what happens is that all the waves arrive in phase, adding together to produce double the pressure at some points, while at other points they cancel each other out to produce zero pressure. If the sources are coherent, in phase and less than one wavelength apart, the sound pressures may be added to get the total sound pressure from which the total SPL may be got. An example of correlated sources is two stereo speakers which are placed within one wavelength of each other and fed with the same signal. The total sound pressure due to n correlated sources producing SPLs L_{I1}, L_{I2}, L_{I3},..., L_{In} is obtained as follows:

$$L_{p,\text{total}} = 20\log_{10}(10^{L_{p1}/20} + 10^{L_{p2}/20} + 10^{L_{p3}/20} + \ldots 10^{L_{pn}/20}) \tag{11.81}$$

To find the overall effect due to multiple sound sources which are uncorrelated, the individual sound intensities (or p_{rms}^2 values) have to be added to get the total sound intensity, from which the total sound intensity level may be found. Thus if two correlated sound sources produce 60 dB each at a listener's location, then $L_{p,\text{total}}$ = 60 dB + 60 dB = 66 dB. If two uncorrelated sound sources produce 60 dB each, then $L_{I,\text{total}}$ = 60 dB + 60 dB = 63 dB. But since L_I is almost equal to L_p for normal atmospheric conditions, we can say that $L_{p,\text{total}}$ = 63 dB in the latter case. Thus, the combination of two random noise sources with both sources being identical, results in an increase of 3 dB over the SPL of one source. If the two independent SPLs are different, the combined level will exceed the higher of the two levels by less than 3 dB. When the difference between the two SPLs exceeds 10 dB, the contribution of the less noisy source to the overall noise level is negligible. The total sound intensity level due to n uncorrelated sources with sound intensity levels L_{I1}, L_{I2}, L_{I3},...,L_{In} is obtained as follows:

$$L_{I,\text{total}} = 10\log_{10}(10^{L_{I1}/10} + 10^{L_{I2}/10} + 10^{L_{I3}/10} + \ldots 10^{L_{In}/10}) \tag{11.82}$$

Sometimes it is necessary to subtract one noise level from another; as when background noise has to be subtracted from the combination of background and machine noise to obtain the SPL produced by a machine alone. The procedure to be followed is similar to the addition of levels.

11.4.7   Radiation Fields of a Sound Source

Considering the simplest form of a sound source, the energy emitted will propagate as spherical waves in all directions. The radiation fields of an acoustic source are the near field, far field, free field, reverberant field and diffuse field. Figure 11.6(a) clearly shows the free field and reverberant field, while Fig. 11.6(b) shows the variation in SPL as a function of distance from the source or in different fields. The salient features of these fields are described below.

11.4.7.1   Near field   The near field is the region very close to the source where the SPL may vary significantly with a small change in distance from the source. The extent of the near field depends on the

frequency and radiation characteristics of the source, as well as its characteristic dimensions. For a spherical sound source, near field extends to about 2 wavelengths of the lowest frequency (largest wavelength) emitted by the source; for an irregularly shaped and rather large source with a definite radiation pattern (which does not radiate uniformly in all directions), the near field extends from about two to five times the largest characteristic dimension of the source. Sound pressure measurements should be avoided in this region.

The intensity and sound pressure at a given point, at a distance r from the source are related as shown in Eq. (11.49). In this equation, the region in which $kr \ll 1$ is called the acoustic near field. In this region

$$u(r, t) = -i \frac{p(r, t)}{\rho_0 \omega r} \tag{11.83}$$

(a)

(b)
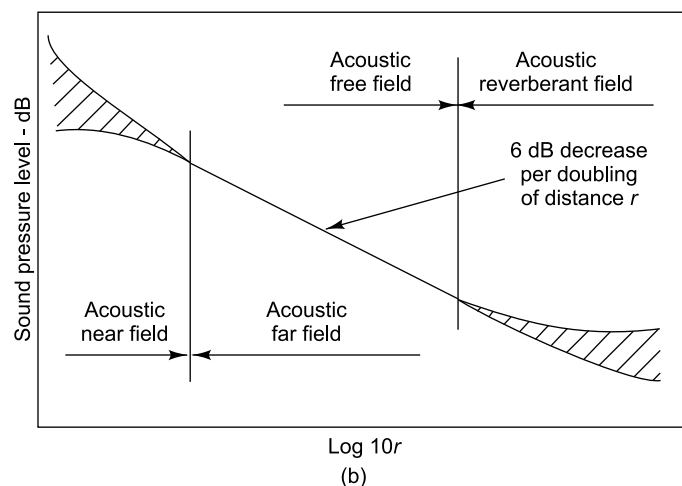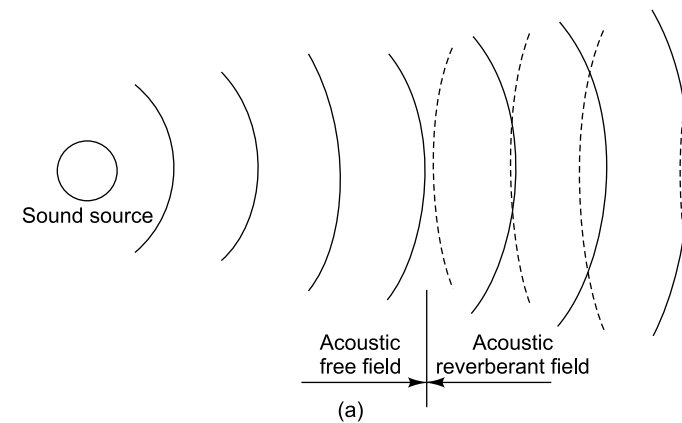
FIGURE 11.6 Radiation fields of a source: (a) free field and reverberant field, (b) variation in SPL with field

The region corresponding to $kr \gg 1$ is called the acoustic far field. For 1-D spherical waves, $kr = 10$ is often used as a demarcation between the near field and the far field. Since $\lambda = c_0/f$ where f is the

frequency in Hz, we have $k = 2\pi/\lambda$. Hence, the distance to the demarcation between the near and far fields corresponding to $kr = 10$ is $r = 1.6\lambda$. In the near field as can be seen from Eq. (11.83) the particle pressure and velocity are 90° out of phase. Besides, in this region as seen in Fig. 11.6(b) the sound field does not decrease by 6 dB each time the distance from the source is doubled (as it does in the far field). The sound pressure and particle velocity must be in phase for sound power to be radiated.

11.4.7.2   Far field   The region far away from the source in called the far field. The far field begins where the near field ends and extends to infinity. The transition from near to far field is gradual and the demarcation is assumed to be $kr = 10$. The equation relating particle pressure and velocity in the far field is as shown below.

$$u(r,t) = \frac{p(r,t)}{\rho_0 c_0} = \frac{p(r,t)}{z_0} \tag{11.84}$$

Here acoustic pressure and particle velocity are in phase as seen from the above equation and this phase relationship is required for wave propagation. The far field comprises the free field and the reverberant field. In the free field, the sound pressure is inversely proportional to the distance r away from the source and therefore the SPL will decay at the rate of 6 dB per octave or 6 dB for every doubling of distance from the source.

11.4.7.3   Free field/direct field   A free field is constituted by a plane wave which propagates in one defined direction. The free field is a region in space where only the directly radiated sound waves from the sound source are present and where there are no reflected waves. Such a field exists when a sound source is located far away from reflecting surfaces or when nearby surfaces are highly absorbent, thus preventing reflected waves. A small sound source (point source) may create a satisfactory plane wave at the measurement location, provided the location is sufficiently far away. The distance between the source and the measurement site should be at least five to ten times the largest dimension of the source and of the microphone or object which is to be placed in the field. It has already been mentioned that in free field conditions, the SPL decreases by 6 dB for doubling of distance from the source. Ideal free fields are difficult, if not impossible, to realize in practice. However, free fields which may be used for instrument verification and for calibration may be created in the laboratory in specially fabricated anechoic chambers, or outdoors away from reflecting surfaces. The anechoic chamber is a room with such highly absorbent walls that all the sound energy striking the walls is absorbed.

Directivity: The SPL at a point due to radiation from a source will depend on the presence of a reflecting surface near a source which will affect not only the radiated sound, but also the apparent directional properties of the source. In the direct field, the equation relating SPL and sound power level must be modified to account for the presence of such reflecting surfaces. This is done by introducing a directivity factor, Q which may also be used to characterise the directional sound radiation properties of a source. We have seen that when a spherical sound source of constant power output (not affected by reflecting surfaces) radiates into 3-D space or whole space

$$W = 4\pi r^2 I \tag{11.85}$$

where I is the constant intensity over a spherical surface of radius r. If a sound source is placed on a hard floor, the radiation will be over a hemisphere as the sound source is reflected from the plane. The source is then said to radiate sound into half space and the expression for sound power is

$$W = 2\pi r^2 I \tag{11.86}$$

It is also possible that the source is in a corner where a wall and floor or a wall and ceiling meet. With two such reflecting planes, the radiation will be over 1/4 of a sphere, or into quarter space. The source may be in a corner where two walls and a floor or ceiling meet; with three reflecting planes, sound is radiated into eighth space or 1/8 of a sphere as shown in Table 11.6. A common expression for sound power, independent of the nature of radiation space may be obtained by defining directivity factor Q and a corresponding Directivity Index (DI) in dB may be incorporated in the expression for sound power level. Table 11.6 shows the directivity indices for difference spaces. The radiated power in terms of directivity factor is

$$W = \frac{4\pi r^2 I}{Q} \tag{11.87}$$

Directivity index is defined as

$$DI = 10 \log_{10} Q \tag{11.88}$$

The sound power level of a spherical source shown in Eq. (11.80) may be modified to the form shown below to account for directivity index.

$$L_W = L_p + 20 \log_{10} \left(\frac{r}{r_{\text{ref}}} \right) - 0.16 - DI \tag{11.89}$$

TABLE 11.6 Directivity factors and indices

Space Q	Pictorial representation	Q	DI (dB)
Whole		1	0
1/2		2	3
1/4		4	6
1/8		8	9

11.4.7.4 Reverberant fields The reverberant field of a source is defined as that part of the sound field radiated by a source which has experienced at least one reflection from the walls, ceiling or floor of the room or enclosure containing the source. Thus it is the field consisting of both directly radiated and reflected sound waves. Such a field exists when there are many reflecting surfaces surrounding the source; the incident and reflected waves then get superimposed to produce a reverberant field. The SPL at a point in the reverberant field will depend on the number of reflections, as well as the absorption coefficient of the reflecting boundaries. In this field also there is no 6 dB reduction with doubling of distance away from the source.

11.4.7.5 Diffuse field In a reverberant field if there are many reflected waves arriving simultaneously from all directions with equal probability and level, the resulting field is called diffuse field; such conditions are created in the laboratory in a reverberation chamber, a room in which all surfaces are hard and highly

reflecting and which essentially contain no sound absorbing materials. The diffuse-field or 'random incidence' sensitivity of a microphone refers to this type of field, even if, in most cases, the sensitivity is calculated from measurements performed under free-field conditions. Diffuse-field spectra created in such rooms may deviate from those of an ideal field due to resonances in the room and due to the sound absorption of air. In cases where diffuse-fields are to be used for technical purposes, the influence of these effects may be reduced by applying more than one sound source and by mounting reflecting panels which are moved continuously in order to vary the dominating room resonances. Sound fields with a close resemblance to a diffuse-field may be found in environments such as factories where many simultaneous sound or noise sources exist or in buildings with hard walls, for example, in halls or churches.

11.4.7.6   Pressure field   A pressure field is characterized by a sound pressure which has the same magnitude and phase at any position within the field. Pressure fields may be found in enclosures or cavities which are small in dimensions compared to the wavelength of sound, making the pressure uniform in the enclosure. Microphone "pressure sensitivity" refers to this type of field. Pressure field is used in acoustic calibrators and in piston phones where an exact sound pressure field occurs in the couplers which are used for testing earphones or calibrating microphones by the reciprocity technique.

11.5   Sound Transmission from One Medium to Another with Normal Incidence

When sound waves from a source strike a surface representing a change from one physical medium to another across an interface, a part of the sound energy is absorbed in the second medium and a part of it reflected back into the first medium, the two quantities depending on the characteristic impedances of the two media. This is shown in Fig. 11.7. Let medium 1 have a characteristic impedance $\rho_1 c_1$ and medium 2, $\rho_2 c_2$. Let the wavenumbers in the two media be $k_1 = \omega/c_1$ and $k_2 = \omega/c_2$. It should be noted that the frequency of the wave in the two media will be the same.
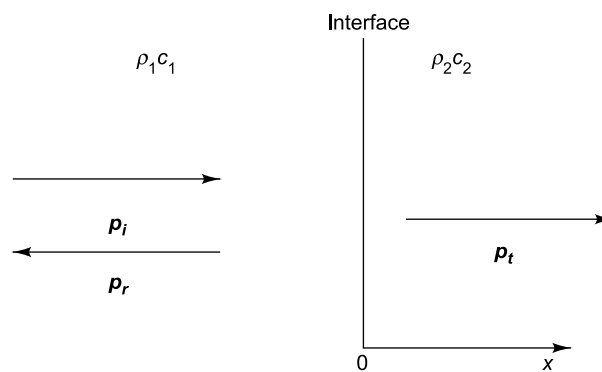
Figure 11.7   Transmission with normal incidence

Let the incident wave be represented as

$$p_i = A_i e^{i(\omega t - k_1 x)} \tag{11.90}$$

Here boldface is used to represent complex quantities as before.

The reflected and transmitted waves may be represented as

$$p_r = A_r e^{i(\omega t + k_1 x)} \tag{11.91}$$

$$p_t = A_t e^{i(\omega t - k_2 x)} \tag{11.92}$$

Two continuity conditions must be satisfied at the boundary separating the two media.

(i) Continuity of pressure: The total pressure in media 1 and 2 should be equal at $x = 0$.

$$(p_i + p_r)_{x=0} = (p_t)_{x=0} \tag{11.93a}$$

$$A_i + A_r = A_t \tag{11.93b}$$

Here A_i, A_r and A_t are in general complex to allow for phase differences between the three waves.

(ii) Continuity of particle velocity: It is assumed that the two media are not physically separated at $x = 0$. This implies that the total normal particle velocity in medium 1 is equal to that of medium 2 at $x = 0$. Thus,

$$(u_i + u_r)_{x=0} = (u_t)_{x=0} \tag{11.94}$$

The particle velocities corresponding to the incident, reflected and transmitted waves may be shown to be

$$u_i = \frac{p_i}{\rho_1 c_1}$$

$$u_r = -\frac{p_r}{\rho_1 c_1} \tag{11.95}$$

$$u_t = \frac{p_t}{\rho_2 c_2}$$

It follows from Eq. (11.94) that

$$\frac{p_i - p_r}{\rho_1 c_1} = \frac{p_t}{\rho_2 c_2} \tag{11.96}$$

Eliminating A_t from Eqs. (11.94) and (11.96), we obtain the relation between the reflected and incident waves as

$$\frac{A_r}{A_i} = \frac{\rho_2 c_2 - \rho_1 c_1}{\rho_2 c_2 + \rho_1 c_1} \tag{11.97}$$

Similarly A_r may be eliminated to give the relation between the transmitted and incident waves as

$$\frac{A_t}{A_i} = \frac{2\rho_2 c_2}{\rho_2 c_2 + \rho_1 c_1} \tag{11.98}$$

Alternately, the two boundary conditions can be combined into a single condition that the specific acoustic impedance is continuous at the boundary and is equal to z_s. Then

$$\left(\frac{p_i + p_r}{u_i + u_r}\right)_{x=0} = z_s = \left(\frac{p_t}{u_t}\right)_{x=0} \tag{11.99a}$$

or

$$\frac{A_i + A_r}{A_i - A_r}\rho_1 c_1 \; = z_s = z_2 \tag{11.99b}$$

Equation (11.99b) may be rearranged to the form

$$\frac{A_r}{A_i} = \frac{z_s - \rho_1 c_1}{z_s + \rho_1 c_1} \tag{11.100}$$

The sound power reflection coefficient may be defined as

$$\alpha_t = \frac{I_r}{I_i} = \frac{|A_r|^2 / 2\rho_1 c_1}{|A_i|^2 / 2\rho_1 c_1} = \left|\frac{A_r}{A_i}\right|^2$$

$$= \left(\frac{\rho_2 c_2 - \rho_1 c_1}{\rho_2 c_2 + \rho_1 c_1}\right)^2 \tag{11.101}$$

The sound power transmission coefficient may be defined as

$$\alpha_t = \frac{I_t}{I_i} = \frac{|A_t|^2 / 2\rho_2 c_2}{|A_i|^2 / 2\rho_1 c_1} = \frac{\rho_1 c_1}{\rho_2 c_2}\left|\frac{A_2}{A_1}\right|^2$$

$$= \frac{4\rho_1 c_1 \rho_2 c_2}{\left(\rho_2 c_2 + \rho_1 c_1\right)^2} \tag{11.102}$$

Equations (11.101) and (11.102) show that sound reflection and transmission phenomena are independent of frequency. We can consider three specific cases:

(i) $\rho_1 c_1 << \rho_2 c_2$

From Eqs (11.101) and (11.102), it can be seen that $\alpha_t = 4\rho_1 c_1 / \rho_2 c_2$ or $\alpha_t \approx 0$ and $\alpha_r \approx 1$, implying that almost all the incident sound energy is reflected back into the originating medium, with very little energy getting transmitted to the second medium. When sound is transmitted from air to water which have characteristic impedances of 415 and 1.48×10^6 MKS Rayls, respectively, $\alpha_t = 1.12 \times 10^{-3}$ or only 0.112% of the incident energy is transmitted into water.

Figure 11.8(a) shows the standing wave patterns for such a case.

(ii) $\rho_1 c_1 = \rho_2 c_2$

With this condition $\alpha_t = 1$ and $\alpha_r = 0$ implying that all the incident sound energy is transmitted into the second medium.

(iii) $\rho_1 c_1 >> \rho_2 c_2$

With this condition, we get $\alpha_t = 4\rho_2 c_2 / \rho_1 c_1$ or $\alpha_t \approx 0$ and $\alpha_r \approx 1$ implying that all the incident energy is reflected back into the first medium. Figure 11.8(b) depicts the standing wave patterns for this case.
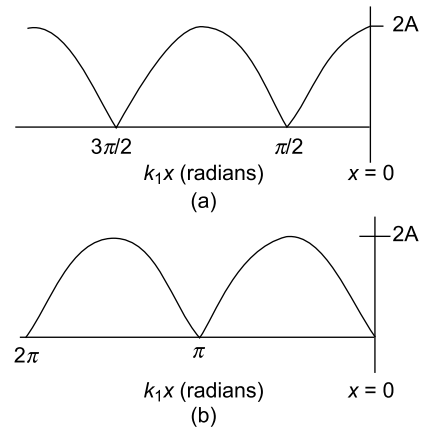
FIGURE 11.8 Standing wave patterns: (a) $\rho_1 c_1 << \rho_2 c_2$, (b) $\rho_1 c_1 >> \rho_2 c_2$

11.5.1 Sound Transmission through a Solid Barrier with Normal Incidence

The equations for this case may be derived on similar lines as for transmission from one medium into another. A quantity of great practical significance stemming from a derivation of this kind is the sound power transmission coefficient of a partition or panel. Such panels find use in noise control applications.

The transmission coefficient is obtained by assuming that we have air as the first and third media in Fig. 11.9 with the barrier as the second medium. The barrier is assumed to be rigid and to have a finite thickness.
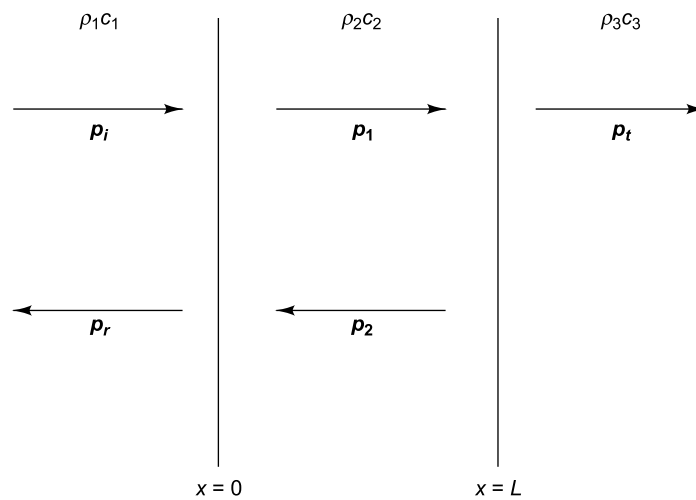
FIGURE 11.9   Sound transmission through a barrier

The sound power transmission coefficient τ, for such a barrier is defined as the ratio of the transmitted acoustic power to the incident power. Since $W = IS$, S being the surface area common to the two media and since $I_i = \dfrac{p_{\text{rms}(i)}^2}{\rho_0 c_0}$ and $I_t = \dfrac{p_{\text{rms}(t)}^2}{\rho_0 c_0}$, the sound power transmission coefficient is given as

$$\tau = \left| \frac{A_t}{A_i} \right|^2 \tag{11.103}$$

and the transmission loss (TL) of a sound wave through the partition is defined as

$$\text{TL} \;=\; 10 \log_{10} \left(\frac{1}{\tau} \right) \tag{11.104}$$

11.6   Acoustics of Enclosed Spaces

The behaviour of sound in enclosed spaces needs to be studied since this throws light on the design of auditoria, class rooms, churches, opera halls, factories and other such spaces. If a sound source is placed in any such enclosure, a part of the energy radiated will get absorbed and a part of the energy will get

reflected from the boundary surfaces of the room. A hard, compact, smooth surface will reflect most of the acoustic energy, while a porous, soft surface will absorb most of it. If the surfaces of the room are reflective in nature, the sound produced by a source will undergo multiple reflections at the boundaries. The larger the reflection coefficients, the larger will be the contribution of the reflected sound to the total sound in the enclosed space. The build up of sound energy will continue even after the source has been turned off. The SPL in such a reverberant sound field will depend on the acoustic power radiated, the size of the room and the acoustic absorption properties of the boundaries. If the surfaces are highly absorptive in nature, the contribution of the reflected sound becomes less and almost a free field condition will exist where most of the sound is the direct sound. In general most work spaces may be considered to be semi-reverberant. Before attempting to obtain the expression for the SPL in an enclosure, it is worthwhile looking at the resonant behaviour of rooms.

11.6.1 Acoustic Field in a Small Rectangular Room

The 3-D wave equation expressed in terms of sound pressure is given by

$$c_0^2 \nabla^2 p(x,y,z,t) - \frac{\partial^2 p(x,y,z,t)}{\partial t^2} = 0$$

where
$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{11.105}$$

If a harmonic solution is assumed, the sound pressure can be written as

$$p(x, y, z, t) = P(x, y, z) e^{jwt} \tag{11.106}$$

Substituting this equation in Eq. (11.105), we get

$$\nabla^2 P(x,y,z) + k^2 P(x,y,z) = 0 \tag{11.107}$$

where $k = \dfrac{\omega}{c_0}$ as has been defined earlier.

This is the 3-D Helmholtz equation and may be solved by a separation of variables as shown.

$$P(x, y, z) = P(x)P(y)P(z) \tag{11.108}$$

Substituting in Eq. (11.107) and dividing by $P(x)P(y)P(z)$, we get

$$\frac{1}{P(x)}\frac{d^2 P(x)}{dx^2} + \frac{1}{P(y)}\frac{d^2 P(y)}{dy^2} + \frac{1}{P(z)}\frac{d^2 P(z)}{dz^2} + k^2 = 0 \tag{11.109a}$$

On separating the variables we get

$$\frac{d^2 P(x)}{dx^2} + k_x^2 P(x) = 0 \tag{11.109b}$$

$$\frac{d^2 P(y)}{dy^2} + k_y^2 P(y) = 0 \tag{11.109c}$$

$$\frac{d^2 P(z)}{dz^2} + k_z^2 P(z) = 0 \qquad (11.109d)$$

where

$$k_x^2 + k_y^2 + k_z^2 = k^2 \qquad (11.110)$$

The solutions to the above equations are

$$P(x) = A_x \cos k_x x + B_x \sin k_x x \qquad (11.111a)$$

$$P(y) = A_y \cos k_y y + B_y \sin k_y y \qquad (11.111b)$$

$$P(z) = A_z \cos k_z z + B_z \sin k_z z \qquad (11.111c)$$

The constants A_x, A_y, A_z, B_x, B_y and B_z are obtained by applying the boundary conditions associated with the rectangular enclosure. The particle velocity at the rigid walls must be zero. Stating this in terms of pressure from the momentum Eq. (11.4), the boundary conditions are

$$\left.\frac{dP(x)}{dx}\right|_{x=0,\,a} = 0, \quad \left.\frac{dP(y)}{dy}\right|_{y=0,\,b} = 0, \quad \left.\frac{dP(z)}{dz}\right|_{z=0,\,c} = 0 \qquad (11.112)$$

Taking the derivative of Eq. (11.111a) with respect to x gives

$$\frac{dP(x)}{dx} = -k_x A_x \sin k_x x + k_x A_x \cos k_x x \qquad (11.113)$$

Applying the boundary conditions at $x = 0$, $B_x = 0$
Similarly applying at $x = a$,

$$k_x A_x \sin k_x a = 0 \text{ or } k_x = \frac{l\pi}{a},\ l = 1,\ 2,\ 3... \qquad (11.114)$$

In a similar fashion $B_y = B_z = 0$

and

$$k_y = \frac{m\pi}{b},\ m = 1,\ 2,\ 3... \qquad (11.115)$$

$$k_z = \frac{n\pi}{c},\ n = 1,\ 2,\ 3... \qquad (11.116)$$

Substituting Eq. (11.111) into Eq. (11.108), we get

$$P(x,\ y,\ z) = A_{lmn} \cos k_{x(1)} x \cdot \cos k_{y(m)} y \cdot \cos k_{z(n)} z \qquad (11.117)$$

Substituting values of k_x, k_y, k_z into Eq. (11.110), we get

$$k_{lmn} = \sqrt{\left(\frac{l\pi}{a}\right)^2 + \left(\frac{m\pi}{b}\right)^2 + \left(\frac{n\pi}{c}\right)^2} \qquad (11.118)$$

Equations (11.117) and (11.118) describe the mode shapes and the wavenumbers of the natural modes that can exist in a rectangular room of dimensions a, b and c. The associated natural frequencies are given by

$$\omega_{lmn} = c_0 \pi \sqrt{\left(\frac{1}{a}\right)^2 + \left(\frac{m}{b}\right)^2 + \left(\frac{n}{c}\right)^2} \tag{11.119}$$

The total sound pressure solution in a rectangular enclosure is

$$p(x, y, z, t) = \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} A_{lmn} \cos k_{x(l)} x . \cos k_{y(m)} y . \cos k_{z(n)} z e^{i\omega_{lmn}t} \tag{11.120}$$

It is essentially the largest dimension which determines the lowest natural frequency.

11.6.2 Sound Pressure Level in a Large Enclosure

For obtaining the SPL in a large room, the following aspects are to be considered: the shape of the enclosed area, the sound absorption and reflection characteristics of the different surfaces of the enclosure and in the case of large areas, the effects of sound absorption by air. Figure 11.10 shows a room with a sound source S and a receiver R. There are two types of sound fields associated with the sound source: the direct field consisting of sound waves travelling directly from the source to the receiver and the reverberant field, consisting of sound waves that reach the receiver after at least one reflection. The reverberant sound field results in a considerable increase in the SPL at the receiver. The steady state sound level is reached when the total sound energy added to the reverberant field is equal to the energy absorbed by the reflecting surfaces.
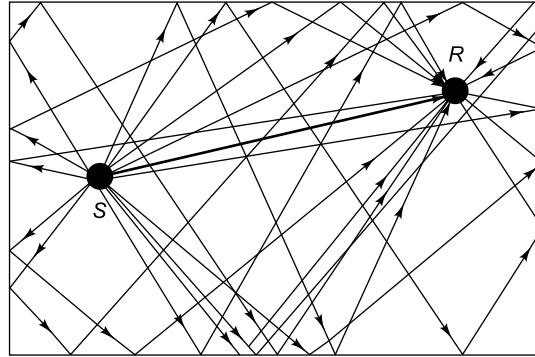
FIGURE 11.10   Direct and reverberant sound fields

The energy density in a large enclosure due to a 1-D spherical wave travelling away from a sound source is composed of the direct field and the reverberant field.

11.6.2.1   *Energy density in the direct field* (D_0)   The energy density D_0 due to the direct field can be computed from the energy flowing through a small volume ΔV as shown in Fig. 11.11 during a small duration Δt. The total energy E in volume $\Delta V = \Delta x \Delta S$ is

$$E = D_0 \Delta S \Delta x \tag{11.121}$$

The acoustic intensity I is given by
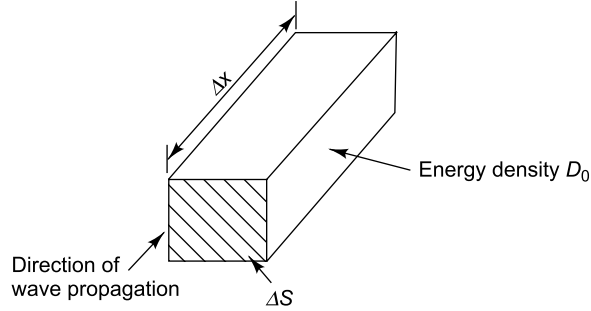
FIGURE 11.11 Energy density in the direct field (D_0)

$$I = \frac{W}{\Delta S} \qquad (11.122)$$

where W = sound power directly radiated by the sound source and ΔS = unit area.

The sound power may also be computed as the energy contained in volume ΔV divided by the time Δt taken for the sound wave to travel distance Δx. Hence, the power radiated from the source is

$$W = \frac{E}{\Delta t} \qquad (11.123)$$

Substituting Eqs (11.121) and (11.123) into (11.122), we get

$$I = \frac{D_0 \Delta S \Delta x}{\Delta S \Delta t} = D_0 c_0 \qquad (11.124)$$

where $\dfrac{\Delta x}{\Delta t} = c_0$, the speed of sound in air.

For a 1-D spherical wave propagation, for any space (full or half or one-fourth or one-eighth) with directivity factor Q

$$I = \frac{WQ}{4\pi r^2} \qquad (11.125)$$

Hence, the expression for energy density in the direct field becomes

$$D_0 = \frac{WQ}{4\pi r^2 c_0} \qquad (11.126)$$

11.6.2.2 *Energy associated with reverberant field* (D_R)   To determine the energy density D_R due to the reverberant field of the enclosure, the sound power associated with the reverberant field that is incident upon a small surface area ΔS is found out. The small energy dE contained in a thin shell-like volume dV a distance r away from ΔS and centred at ΔS (Fig. 11.12) is

$$dE = D_R dV = D_R dr dS, \qquad (11.127)$$

where $\qquad dV = dSdr$ and $dS = r^2 \sin\theta d\theta d\psi \qquad (11.128)$
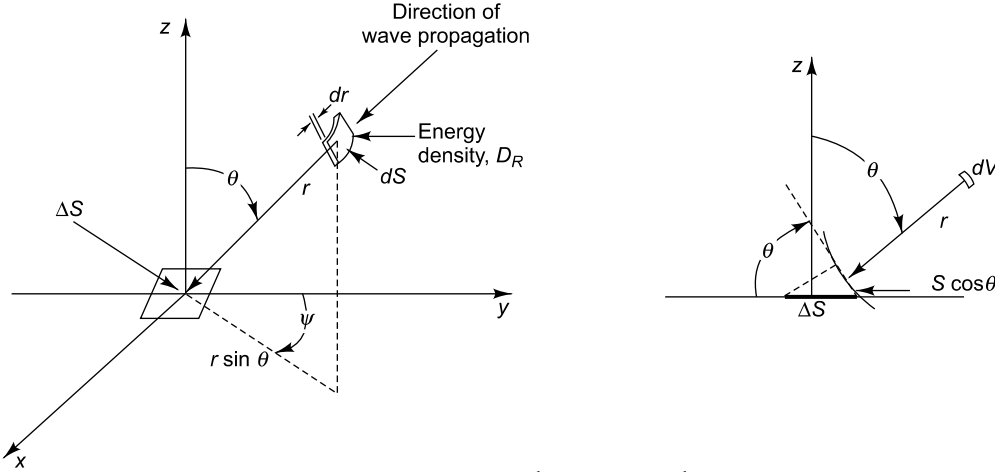
FIGURE 11.12 Acoustic radiation on to element

The sound power dW radiated towards the surface ΔS from the volume dV is

$$dW = \frac{dE}{dt} = D_R \, dS \frac{dr}{dt} = D_R c_0 \, dS \tag{11.129}$$

Acoustic power is assumed to radiate spherically from dV towards ΔS as depicted in Fig. 11.12. The acoustic intensity dI associated with the sound power dW at ΔS is

$$dI = \frac{D_R c_0 \, dS}{4\pi r^2} \tag{11.130}$$

The small power dW_i incident upon the surface ΔS is

$$dW_i = dI \Delta S \cos\theta = \frac{D_R c_0 \Delta S \cos\theta \, dS}{4\pi r^2} \tag{11.131}$$

where $\Delta S \cos\theta$ is the area element of the sphere of radius r projected on to ΔS normal to r.

Due to multiple reflections it may be assumed that a diffuse field exists. Hence, the total sound power incident on the surface element ΔS is obtained by integrating $\mathrm{d}W_i$ over the surface of the hemisphere. Therefore,

$$W_i = \frac{D_R c_0 \Delta S}{4\pi r^2} \int_0^{2\pi} \int_0^{\pi/2} r^2 \, \sin\theta \, \cos\theta \, d\theta \, d\psi = \frac{D_R c_0 \Delta S}{4} \tag{11.132}$$

The sound power W_a absorbed by the surface element ΔS is

$$W_a = W_i \alpha = \frac{D_R c_0 \Delta S \alpha}{4} \tag{11.133}$$

The total sound power absorbed by the surface which is equal to the sound power lost from the reverberant field is obtained by adding the power over all surface areas as shown below.

$$W_a = \frac{D_R c_0 S \bar{\alpha}}{4} \tag{11.134}$$

where $\bar{\alpha}$ is the average absorption coefficient of all such absorbing surfaces $S_1, S_2, S_3,..., S_n$ with sound absorption coefficients $\alpha_1, \alpha_2, \alpha_3, ..., \alpha_n$.

$$\bar{\alpha} = \frac{S_1\alpha_1 + S_2\alpha_2 + ... + S_n\alpha_n}{S_1 + S_2 + ... + S_n} = \frac{S_1\alpha_1 + S_2\alpha_2 + ... + S_n\alpha_n}{S} \qquad (11.135)$$

From the law of conservation of energy, the sound power supplied to the reverberant field should equal the sound power absorbed by the surfaces of the enclosure under steady state conditions. Besides, the sound power supplied to the reverberant field should equal the remaining power after the sound waves undergo their first reflections. Hence,

$$W(1-\bar{\alpha}) = \frac{D_R c_0 S\bar{\alpha}}{4} \qquad (11.136a)$$

Eq. (11.139a) may be rearranged to give

$$D_R = \frac{4W(1-\bar{\alpha})}{c_0 S\bar{\alpha}} = \frac{4W}{c_0 R} \qquad (11.136b)$$

where R is the room constant defined as

$$R = \frac{S\bar{\alpha}}{1-\bar{\alpha}} \qquad (11.137)$$

The total energy density D in an enclosure is the sum of the energy densities from the direct and reverberant fields, or

$$D = D_0 + D_R = \frac{W}{c_0}\left(\frac{4}{R} + \frac{Q}{4\pi r^2}\right) \qquad (11.138)$$

The total energy density in the enclosure can be expressed as

$$D = \frac{p_{rms}^2}{\rho_0 c_0^2} \qquad (11.139)$$

Substituting Eq. (11.139) into Eq. (11.138), we get

$$p_{rms}^2 = W\rho_0 c_0\left(\frac{4}{R} + \frac{Q}{4\pi r^2}\right) \qquad (11.140)$$

Both sides of the above equation may be divided by p_{ref}^2 and the numerator and denominator of the right side may be multiplied by $(\rho_0 c_0)_{ref}$ and S_{ref} to give

$$\frac{p_{rms}^2}{p_{ref}^2} = \frac{W}{(p_{ref}^2 S_{ref})/(\rho_0 c_0)_{ref}}\frac{\rho_0 c_0}{(\rho_0 c_0)_{ref}}S_{ref}\left(\frac{4}{R} + \frac{Q}{4\pi r^2}\right) \qquad (11.141)$$

For normal temperature and atmospheric conditions

$$\frac{\rho_0 c_0}{(\rho_0 c_0)_{ref}} = 1 \text{ and } W_{ref} = \frac{p_{ref}^2 S_{ref}}{(\rho_0 c_0)_{ref}} \qquad (11.142)$$

Hence, Eq. (11.141) can be written as

$$\frac{p_{\text{rms}}^2}{p_{\text{ref}}^2} = \frac{W}{W_{\text{ref}}} S_{\text{ref}} \left(\frac{4}{R} + \frac{Q}{4\pi r^2} \right) \qquad (11.143)$$

Taking logarithm of both sides of the above equation we get

$$L_p = L_W + 10 \log_{10} \left(\frac{4}{R} + \frac{Q}{4\pi r^2} \right) + 10 \log_{10} S_{\text{ref}} \qquad (11.144)$$

If R is in m^2 and r is in m, $S_{\text{ref}} = 1$ m^2 and Eq. (11.144) becomes

$$L_p = L_W + 10 \log_{10} \left(\frac{4}{R} + \frac{Q}{4\pi r^2} \right) \qquad (11.145)$$

It is to be noted here that in the direct field $\frac{4}{R} \ll \frac{Q}{4\pi r^2}$ and in the reverberant field, $\frac{4}{R} \gg \frac{Q}{4\pi r^2}$. Accordingly, the appropriate terms alone may be used in Eq. (11.145).

11.6.3 Decay of a Sound Field in an Irregularly Shaped Enclosure

The theory explaining reverberation time and the acoustical behaviour of rooms was developed by Sabine. A quantity of great concern to both architects and musicians in room acoustics is reverberation time which is defined as the time taken for SPL to fall by 60 dB from the steady state value, the moment the source is switched off. If an impulsive sound is generated in a room with reflecting boundaries, there are multiple reflections at the boundaries, after which there is more or less a uniform sound field. This field then decays due to absorption of sound energy by the materials of the room surfaces. The rate of decay depends on the absorptive properties of the reflecting surfaces and the distances between them. Figure 11.13 shows the build-up of sound in an enclosure after a sound source has been turned on at time $t = -0.6$ s. Also shown is the decay of sound for three different cases: (a) very little absorption, (b) medium absorption and (c) large absorption. In this section an expression is derived for the time taken for the sound field in an enclosure to decay by 60 dB after a sound source is turned off.
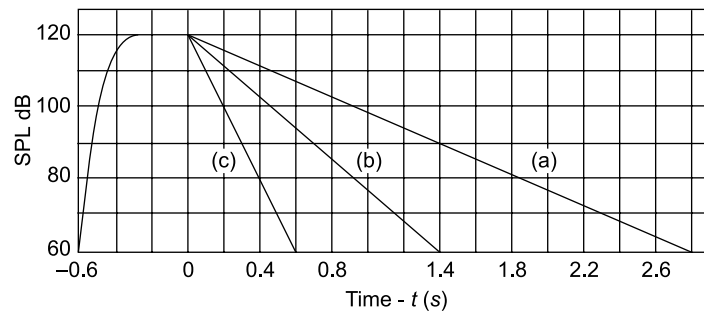
FIGURE 11.13   Build up and decay of sound in an enclosure

Let the energy density in the room initially be D. After each reflection, the energy density D decreases by a factor $(1 - \overline{\alpha})$ as shown in Fig. 11.14. Let d be the average distance or the mean free path travelled

by a sound wave between reflections and let t' be the average time between reflections or the mean free time calculated as

$$t' = \frac{d}{c_0} \tag{11.146}$$

Hence, the expressions for the energy density become

$$D(t') = D(1-\bar{\alpha}) \text{ after the first reflection}$$

$$D(2t') = D(1-\bar{\alpha})^2 \text{ after the second reflection}$$

$$D(nt') = D(1-\bar{\alpha})^n \text{ after the } n\text{th reflection} \tag{11.147}$$

Studies conducted, both analytically and experimentally, on several enclosures of different shapes, have shown that d is given by the expression

$$d = \frac{4V}{S} \tag{11.148}$$

where V is the volume of the room and S is the total surface area of the room. Hence, Eq. (11.146) can be rewritten as

$$t' = \frac{4V}{c_0 S} \tag{11.149}$$
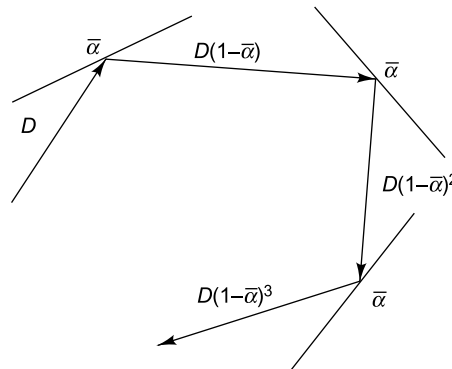
FIGURE 11.14 Effects of several reflections of a sound wave

The time t taken by the wave to undergo n reflections is

$$t = nt' \tag{11.150}$$

Substituting the above equation in the expression for t' in Eq. (11.149), we get

$$n = \frac{c_0 S t}{4V} \tag{11.151}$$

Substituting for t from Eq. (11.150) and for n from Eq. (11.151) into Eq. (11.147), we get

$$D(t) = D(1-\bar{\alpha})^{(c_0 S t / 4V)} \tag{11.152}$$

Plugging in the value of $D = \dfrac{p_{rms}^2}{\rho_0 c_0^2}$ into the above equation, we get

$$\frac{p_{\text{rms}}^2(t)}{p_{\text{rms}}^2(t=0)} = (1-\bar{\alpha})^{(c_0 St/4V)} \tag{11.153}$$

Taking logarithm on both sides, we get

$$L_{p(t)} - L_{p(t=0)} = 10 \log_{10}\left\{(1-\bar{\alpha})^{(c_0 St/4V)}\right\} \tag{11.154}$$

Reverberation time T is defined as the time it takes for the SPL to decrease 60 dB after the source is switched off and is obtained as

$$-60 = \frac{c_0 ST}{V} 2.5 \log_{10}(1-\bar{\alpha}) \tag{11.155}$$

The above equation may be written in terms of natural logarithm as

$$60 = 1.086 \frac{c_0 T}{V}\left\{-S\ln(1-\bar{\alpha})\right\} \tag{11.156}$$

Room absorption a is defined as

$$a = S\ln(1-\bar{\alpha})^{-1} \tag{11.157}$$

Substituting for a in Eq. (11.159), the reverberation time can be written as

$$T = \frac{60V}{1.086 c_0 a} \tag{11.158}$$

This is called the Eyring reverberation time formula. If c_0 is in terms of m/s, V is in m^3 and a is in m^2, T in s can be written as

$$T = 0.161 \frac{V}{a} \tag{11.159}$$

If the average sound absorption coefficient $\bar{\alpha}$ is less than 0.1,

$$\text{In }(1-\bar{\alpha})^{-1} \approx \bar{\alpha} \tag{11.160}$$

and room absorption a can be written as

$$a \approx S\bar{\alpha} \tag{11.161}$$

The reverberation time can then be written as

$$T = 0.161 \frac{V}{S\bar{\alpha}} \tag{11.162}$$

The procedures for measuring many of the acoustic quantities described in this chapter are described in Chapter 13.

FURTHER READINGS

1. Ambrose, J.E. and Ollswang, A.J., *Simplified Design for Building Sound Control (Parker Ambrose Series of Simplified Design Guides)*, John Wiley & Sons, Inc., New York, 1995.
2. Barron, R.F., Industrial *Noise Control and Acoustics*, CRC Press, Boca Raton, Florida, 2003.
3. Beranek, L.L. and Ver, I.L., *Noise and Vibration Control Engineering: Principles and Applications*, Wiley-Interscience, New York, 1992.
4. Beranek, L.L., *Acoustics,* Acoustical Society of America, Woodbridge, New York, 1993.
5. Beranek, L.L., *Concert Halls and Opera Houses: Music, Acoustics, and Architecture*, Springer-Verlag, New York, 2003.
6. Beranek, L.L., *Noise and Vibration Control,* Institute of Noise Control Engineering, Poughkeepsie, New York, 1988.
7. Bies, D.A. and Hansen, C.H., *Engineering Noise Control: Theory and Practice*, Taylor & Francis, New York, USA, 2003.
8. Blackstock, D.T., *Fundamentals of Physical Acoustics*, Wiley-Interscience, New York, 2000.
9. Cremer, L. and Müller, H.A. (translated by T. J. Schultz), *Principles and Applications of Room Acoustics*, Volumes 1 and 2, Applied Science Publishers, Barking, Essex, 1982.
10. Crocker, M.J., *Encyclopaedia of Acoustics*, Wiley-Interscience, New York, 1997.
11. Crocker, M.J., *Handbook of Acoustics*, Wiley, New York, 1998.
12. Crocker, M.J., *Handbook of Noise and Vibration Control*, John Wiley and Sons, Hoboken, New Jersey, 2007.
13. Day, B.F., Ford, R.D. and Lord, P., *Building Acoustics,* Elsevier Applied Science, Amsterdam, New York, 1969.
14. Everest, F.A. and Pohlman, K.C., *Master Handbook of Acoustics*, McGraw-Hill, New York, 2000.
15. Fahy, F.J. and Walker, J.G., *Fundamentals of Noise and Vibration*, E & FN Spon, London, 1998.
16. Fahy, F.J., *Foundations of Engineering Acoustics*, Academic Press, London, 2001.
17. Filippi, P., Habault, D., Lefebvre, JP and Bergassoli, A., *Acoustics: Basic Physics, Theory and Methods*, Academic Press, London, 1998.
18. Filippi, P., *Theoretical Acoustics and Numerical Techniques (Cism International Centre for Mechanical Sciences Courses and Lectures)*, Springer-Verlag, New York, 1984.
19. Hall, D. E., *Basic Acoustics*, Krieger Publishing, Malabar, FL, 1993.
20. Harris, C., *Handbook of Acoustical Measurements and Noise Control*, The McGraw-Hill, New York, 1991.
21. Hemond, C.J., *Engineering Acoustics and Noise Control*, Prentice Hall, Englewood Cliffs, New Jersey, 1983.
22. Kinsler, L.E. and Frey, A.R., *Fundamentals of Acoustics*, John Wiley & Sons, New York, 1982.
23. Kinsler, L.E. Frey, A.R., Coppens, A.B. and Sanders, J.V., *Fundamentals of Acoustics*, John Wiley and Sons, New York, 1999.
24. Kleppe, J.A., *Engineering Applications of Acoustics*, Artech House Publishers, Boston, Massachusetts, 1989.
25. Lindsay, R.B., *Physical Acoustics*, Van Nostrand Reinhold Co., New York, 1974.
26. Maekawa, Z. and Lord, P., *Environmental and Architectural Acoustics*, Taylor & Francis eBookstore, 2007.

27. Malecki, I., *Physical Foundations of Technical Acoustics*, Pergamon, New York, 1969.

28. Mechel, F.P. and Munjal, M.L., *Formulas of Acoustics*, Springer, Berlin, New York, 2004.

29. Morse, P.M. and Ingard, K.U., *Theoretical Acoustics*, Princeton University Press, Princeton, New Jersey, 1986.

30. Möser, M. and Zimmermann, S., *Engineering Acoustics: An Introduction to Noise Control*, Springer-Verlag, Berlin, New York, 2004.

31. Peters, R.J. and Reeves C.W., *The Noise and Acoustics Handbook*, Coxmoor Publishing Co., UK, 1998.

32. Pierce, A.D., *Acoustics: An Introduction to its Physical Principles and Applications*, Acoustical Society of America, Woodbury, New York, 1989.

33. Rayleigh, J. W. S. and Lindsay, R.B., *The Theory of Sound*, Vols. 1 and 2, Dover Publications, New York, 1976.

34. Reynolds, D.D., *Engineering Principles of Acoustics: Noise and Vibration*, Allyn & Bacon Inc., Boston, 1981.

35. Rossing, T.D. and Fletcher, N.H., *Principles of Vibration and Sound*, Springer-Verlag, New York, 2004.

36. Seto, W.W., Acoustics, *Schaum's Outline Series*, McGraw-Hill, New York, 1971.

37. Skudrzyk, E.J., *Foundations of Acoustics, Basic Mathematics & Basic Acoustics*, Springer-Verlag, New York, 1971.

38. Smith, B.J., Peters, R.J. and Owen, S., *Acoustics and Noise Control*, Longman, London, New York, 1996.

39. Subrahmanyam, N. and Lal, B.V., *A Textbook of Sound*, Vikas Publishing House Pvt. Ltd., New Delhi, 1986.

40. Temkin, S., *Elements of Acoustics*, John Wiley & Sons Inc., New York, 1981.

Chapter

12

Acoustic Transducers

At the front end of all acoustic measurement systems is a transducer called the microphone, which converts sound pressure variations into electrical signals. This chapter gives a general introduction to microphone theory and provides information on the various types of microphones available in the market, their characteristics and the selection criteria commonly used to identify the microphone most suitable for a particular application. It also gives the background information regarding the characteristics, specifications and applications of various acoustic exciters. Finally, it describes various calibrators and calibration techniques.

12.1 Parameters to be Considered in the Choice of Microphones

Microphones have a wide range of applications such as acoustic measurements, telecommunications, broadcasting, recording of music, consumer electronics, etc. For any specific microphone application, the acoustic engineer can choose from a number of different transduction principles. A microphone may sense the pressure, the pressure gradient or the particle velocity, which is then converted to an electrical signal. All precision measurement microphones are mostly pressure sensing condenser microphones for the reason that they detect what the human ear detects, namely pressure. They use a constant electrical charge for conversion of the diaphragm displacement into an analogue electrical signal. If particle velocity or sound intensity is the parameter to be determined, pressure sensing microphones can be used. The measuring microphones influence the sound field, especially at higher frequencies, where the wavelength and microphone dimensions are of the same order of magnitude. However, the influence depends on the type of sound field (pressure-field, free-field or diffuse-field). When selecting a microphone, it is important to first consider the measurement requirements and understand how they impose demands on the performance of the microphone.

12.1.1 Technical Considerations

Technical considerations for good acoustic and electric performance are discussed below.

12.1.1.1 Microphone type depending on sound field When making acoustic measurements, it is to be ensured that the microphone itself does not influence the sound field being measured and change the sound pressure so that it will not lead to a measurement error. The influence varies depending on the type of field, i.e. free-field or pressure-field or diffuse field and on the type of microphone, its dimensions as well as the diaphragm impedance (to a small extent). This influence may be small enough

to be negligible in some cases, whereas in other cases it may be unacceptable amounting to several dB, and is often a function of frequency of sound. The pressure-sensing microphone can be designed for use in different types of sound fields, i.e. pressure-field, free-field and diffuse-field (equivalent to random incidence response). Microphone types which have been designed for specific applications and which therefore correct for the influence of the field are commercially available. These microphones are optimized to have a flat frequency response called the optimized response in one of these sound fields. The differences between the responses are seen only at higher frequencies, differing by less than 0.1 dB below 1000 Hz. If such specifically designed microphones are not used, the user must evaluate, and if possible, correct for the influence.

Pressure-field microphone: The sound pressure in a pressure-field has the same magnitude and phase at any position within the field. Pressure fields may be found in enclosures or cavities which are small in dimensions compared to the wavelength. Typical examples are couplers applied for testing of earphones or calibration of microphones. The frequency response characteristic of a pressure-field microphone is optimized to be as flat as possible, for uniform pressure on the outer surface of its diaphragm. For this, the diaphragm should be as stiff as the walls of the coupler within which it is applied. Since this is not practically possible, appropriate corrections may be made.

Free-field microphone: Free-field microphone types have a flat free-field frequency response characteristic to sound waves that are perpendicular to the diaphragm (at zero degree incidence). When making outdoor measurements away from reflecting surfaces or indoors in an office offering a lot of natural acoustic damping, a free-field microphone offers the best choice. The basic difference between a free-field microphone and a normal pressure microphone is essentially in the design compensation of frequency response of the former such that it can be directed at the noise source with minimal errors resulting from the introduction of the microphone in the sound field. The correction for the influence of the microphone body on the sound field can be made by post-processing of measured sound spectra, but the compensation might also be built into the microphone itself. The compensation is made by introducing heavy damping for the diaphragm resonance, leading to a frequency response characteristic which is essentially flat in an undisturbed free-field.

Diffuse-field microphone: Dedicated microphones for diffuse-field measurements are rare, the reason being that the diffuse-field correction or the pressure change created by the microphone itself is so small that many pressure-field microphones also have good diffuse-field characteristics. For measurements in enclosed areas with hard reflective surfaces where reverberations are likely, pressure-field microphones adapted for random incidence measurements offer the best choice. This is because the random incidence response of a pressure-field microphone is much 'flatter' or more constant across the frequency range than that of a free-field microphone.

12.1.1.2   *Technical specifications of a microphone*

Frequency response: The microphone should have a wide frequency range and flat frequency response. Since microphones generally have a wide operational frequency range, frequency response should be considered in relation to other requirements such as the type of sound field.

Dimensions: The size (diameter) of the microphone is to be chosen according to the highest frequency range of interest and the dynamic range of sound pressure levels (SPLs), and usually reduces with higher frequency and SPL to be measured. It is also true that the larger microphones have higher sensitivity than the smaller ones.

Equivalent noise dB$_N$ (A): This is the level of sound pressure required to produce output voltage corresponding to the inherent A weighted noise voltage, measurable with 3% non-linear distortion and is dictated by the inherent noise of the microphone and pre-amplifier combination.

Dynamic range: This is the range from dB$_N$ (A) to the maximum SPL. For a wide linear dynamic range, low inherent noise and low distortion are required. The lower limit of the dynamic range is dictated by the inherent noise of the microphone and pre-amplifier combination. This is of concern for use in office environments where levels as low as 10–12 dBA may have to be measured. The upper limit of the dynamic range is dictated by the maximum SPL; e.g. in measurements of exhaust systems where high SPLs (140–150 dB) may be encountered. It is normally either the lower or the upper limit of dynamic range that is of interest depending on the application.

Sensitivity: This is described in terms of output mV per unit input Pa pressure fluctuation. Pressure sensitivity implies sensitivity to the actual pressure acting on the microphone, whereas free field sensitivity implies sensitivity to pressure that existed in the sound field before insertion of the microphone, these two being equal at low frequencies.

Open circuit sensitivity (S_o): This is defined as the pressure-field sensitivity, valid with an idealized pre-amplifier which does not load the microphone. It is the ratio of open circuit voltage to pressure and is determined at a frequency of 250 or 1000 Hz and is used for calibration and monitoring of the microphone.

Loaded sensitivity (S_c): When the microphone cartridge is connected to the pre-amplifier, its input voltage is attenuated by the pre-amplifier input capacitance. This effect is valid over a wide frequency range. The loaded sensitivity is described in terms of the open circuit sensitivity as $S_c = S_o + G$ (dB) where G is the gain of the microphone and pre-amplifier combination.

Phase response: This should be considered when choosing microphones for sound intensity measurements. It is not normally the absolute phase response that is important, but the relative phase response between a pair of microphones, necessitating the phase response characteristics to be closely matched. Special pairs of microphones, with matched phase responses, are available.

Type of polarization: There are two different types of microphone constructions, one that employs an external voltage supply called the polarization voltage to establish the charge through a large resistor to polarize the air gap between the back plate and diaphragm (externally polarized) and one where the polarization charge is stored in an electret layer on the back plate of the microphone (pre-polarized). Externally polarized microphones are ideal for general field and laboratory use and for high temperature measurements. Pre-polarized microphones are suitable for portable sound level meters where their lightweight and lack of a requirement for a polarization voltage is desirable. Such microphones also offer slightly better performance in very humid environments.

Directivity: The directional characteristic is the variation in relative sensitivity of the microphone as a function of the angle of incidence. This can be represented graphically by polar plots given at fixed frequencies. The directional characteristic is relevant if the microphone is used for free-field measurements.

Stability of the transducer: The stability of a measurement microphone is a very important feature. It implies high stability of sensitivity and frequency response during short term and long term fluctuations of temperature, humidity, etc.

Calibration chart: The calibration chart is an important piece of documentation providing essential information about the performance characteristics of the microphone. It should have comprehensive

specifications and performance documentation in the form of individual calibration charts. The open-circuit sensitivity, frequency response, dynamic range, etc. are examples of the information given on the calibration chart. It allows the user to verify the performance of the microphone and pre-amplifier system and to calibrate the measurement system. The microphone should have high suitability for measurement and for calibration using practical and accurate methods. Emphasis is also to be given to periodic calibration with standard noise sources so that an allowable accuracy of ±1 dB is available throughout the guaranteed frequency range of the microphone and the sound level meter.

12.1.1.3 *Environmental considerations*

When selecting a microphone, the environment where measurements are to be made should be considered. The microphone should have minor influence from environment and should operate satisfactorily over a range of environmental conditions with very little influence from ambient pressure, temperature, humidity, wind effects, vibration, magnetic and electromagnetic fields, dust, pollution, etc. It should have good mechanical robustness, bump and shock resistance, chemical resistance, corrosion resistance, etc. Some parameters which should be considered are as follows:

Robustness: In a protected laboratory environment all measurement microphones can be used, but more robust general purpose microphones are required for field use.

Temperature: At normal temperatures (–30 to +125 °C) all microphones may be used. At high temperatures (up to 300 °C) special microphones (e.g. Falcon Range™) should be used and at very high temperatures, above 300 °C, a probe microphone should be employed. The tip of the probe on the probe microphone can withstand up to 700 °C.

Atmosphere: All microphones may be used in normal air, but for measurements where corrosive industrial gases exist, for example, when making measurements in industrial chimneys, corrosion resistant (Falcon Range™) microphones should be used.

Wind: The effect of wind is to produce a turbulent air stream around a microphone and this is wrongly sensed as sound. Wind noise in outdoor measurements can greatly degrade measured acoustic data and therefore a windscreen should be used. It is generally made of a specially prepared moisture and corrosion resistant porous polyurethane foam. Where wind speeds are greater than 24 kmph, wind noise may make measurements invalid even with a wind screen (which typically reduces noise by approximately 15 dB for wind speeds up to 120 kmph). This is because the spectra of measured sound and wind often overlap, especially for low frequency acoustic sources making it impossible to separate them by band selective filtering.

Humidity: In general, humidity does not influence the sensitivity and frequency response of a microphone. However some microphones have a layer of quartz on the diaphragm which absorbs moisture, leading to a decrease in tension of the diaphragm and a corresponding decrease in microphone sensitivity. Such a situation can arise when sudden changes in temperature and humidity occur, for example, when going from a warm, humid environment to a cool air-conditioned building. The moisture will attenuate the sensitivity of the microphone and as a side effect, increase the inherent noise level. For high humidity, pre-polarized microphones offer greater reliability because they are more resistant to the attenuation of the polarization voltage which can occur at high humidity levels.

Effect of magnetic field: Microphones are typically designed with materials that provide a very low sensitivity to magnetic fields. Special microphones (Falcon Range™ from Brüel & Kjær, Denmark) are available which have significantly lower sensitivity to magnetic fields than earlier microphones.

Effect of vibration: Vibration sensitivity is defined as the ratio of output voltage of the microphone to casing acceleration. The vibration sensitivity of the microphone normal to the diaphragm is well defined as it is determined by the mass of the diaphragm; the sensitivity is much smaller in all other directions. Pre-amplifiers and electrical adaptors may also contribute to the vibration sensitivity of the measurement channel.

12.2   VARIOUS TYPES OF MICROPHONES

In this section, various types of microphones such as carbon granule, condenser, electret capacitor, electrodynamic/moving coil, piezoelectric and ribbon microphone are discussed. Their principles of operation and typical specifications are also described.

12.2.1   The Carbon Granule Microphone

The carbon button microphone (Fig. 12.1) is one of the earliest types of microphones, having been used extensively in all early telephones and in recording and broadcasting systems.
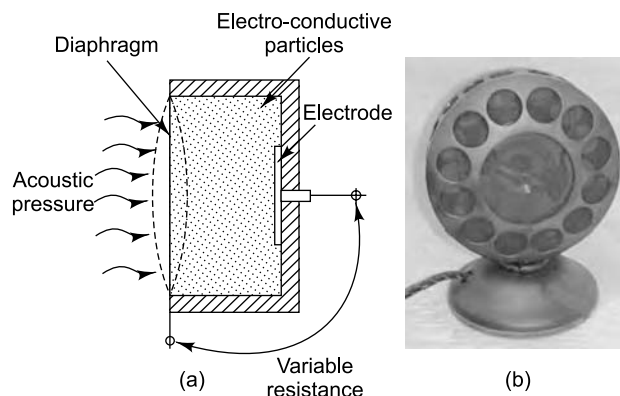
FIGURE 12.1   Carbon button microphone: (a) principle of operation
(b) photograph (http://www.stokowski.org)

In this device, the sound field acts upon an electroconductive diaphragm that develops pressure on a capsule of compressed carbon powder. This capsule or carbon button, as it is called, contains carbon granules between two carbon discs. The sound waves incident on the diaphragm cause it to vibrate, transmitting the displacement to the carbon particles and changing their pressure dependent contact resistance. The front and rear contacts are insulated and brought out to terminals. When a DC voltage is applied across the capsule, the alternating resistance produces an AC voltage drop, which is proportional to the displacement of the diaphragm. The mouthpiece acts as a horn to increase the acoustic pressure on the diaphragm.

Typical specifications

    Sensitivity   : 0.1–0.2 V/μbar
    Resistance   : 30–100 Ω
    Frequency range : 250–3200 Hz

Advantages

 Carbon microphones are cheap and durable.

Disadvantages

 (i) The main disadvantage of the carbon microphone is the so-called 'carbon hiss'.
 (ii) If the hermetic seal of the button is damaged, moisture can cause the granules to pack, decreasing the resistance of the microphone and making it less sensitive.
 (iii) It does not have a very good frequency response.

12.2.2   Condenser Microphone

The condenser microphone was perfected as a much-needed substitute for the carbon microphone which was very noisy. It was subsequently used in broadcasting, until it was replaced in the recent past by the dynamic microphone. Today it is almost exclusively used for all technical acoustical measurements. This microphone is so called because it makes use of the principle of a condenser (capacitor) to convert acoustic signals to an output voltage. The classic design of a pressure-sensing microphone is shown in Fig. 12.2(a). It consists of a metal housing with a thin metallic membrane or diaphragm, which is adequately protected by a grid and which constitutes one element of a parallel plate condenser. A back-plate is mounted in parallel behind the delicate and highly tensioned diaphragm (which is designed to have a high resonant frequency), and forms the other plate of the condenser. This fixed plate is called the polarization electrode and is maintained at a potential of about 200 V. A suitable electrical insulator supports this polarization electrode and insulates it from the housing at ground potential. The diaphragm and the front of the back-plate form the plates of the active capacitor which generates the output signal of the condenser microphone.

 When sound is incident upon the diaphragm, the alternating pressure of the sound wave causes small displacements of the diaphragm, resulting in a small change in condenser capacitance. There is a small static pressure equalization vent (which is nothing but a narrow air channel) in the housing of the microphone, allowing air leak to the interior, so that the static pressure inside the microphone assembly is always equal to the ambient atmospheric pressure. The distance between the microphone diaphragm and back plate must be constructed to very stringent mechanical tolerances and is typically set to 20 μm. Similarly, the tension of the diaphragm, which is stretched like the skin of a drum, must be carefully controlled, since the sensitivity of the microphone is inversely proportional to the stiffness of the diaphragm system. The higher the tension, the higher is the stiffness and the lower the microphone sensitivity. The thickness of the diaphragm may vary from about 1.5 to 8 μm. The stray capacitance or the passive capacitance between the back plate and the housing is kept as small as possible.

 The stiffness and mass of the diaphragm system determine the diaphragm resonance frequency, which in turn determines the upper frequency limit of the transducer. Thus the microphone size, the sensitivity and the frequency range are interdependent and cannot be selected in isolation, requiring the user to make a compromise when selecting a microphone. Generally, small microphones work well at higher frequencies and create smaller disturbances in the sound field, but they also have lower sensitivity. The microphone has to be located very close to the pre-amplifier to minimize stray capacitances. Besides, the bias supply must be extremely well regulated and ripple-free, since every variation will result in distortion of the acoustic signal. Figure 12.2(b) shows a condenser microphone. Table 12.1 shows the typical specifications of condenser microphones of various sizes.

The instantaneous output voltage may be derived from the formula which follows:

$$E \times C = Q_0$$

$$(E_0 + e) \times \frac{\varepsilon \times A}{D_0 + d} = E_0 \times \frac{\varepsilon \times A}{D_0} \qquad (12.1)$$

$$\therefore \qquad e = E_0 \frac{d}{D_0}$$

where A is the area of capacitor plates, C is the instantaneous capacitance between plates, D_0 is the distance between plates at rest position, d is the displacement of the moveable plate (diaphragm) from rest position, E is the instantaneous voltage between plates, E_0 is the polarization voltage, e is the voltage change caused by plate displacement, Q_0 is the constant charge on the plate and ε is the dielectric constant of air.

Thus it is seen that the output voltage of the system and hence the sensitivity is proportional to the bias voltage and the diaphragm area, and inversely proportional to the separation and the diaphragm stiffness. There is a linear relationship between output voltage and displacement, even if the corresponding capacitance changes are non-linear.

Surface microphones (Fig. 12.2c) currently available in the market are designed for measurement of the true surface pressure and are highly useful for automotive and aerospace applications. In these microphones the diaphragm is kept flush with the microphone housing in order to minimize the wind-generated noise of the microphone and the vent for static pressure equalization is placed just next to the diaphragm on the front of the microphone. This is highly appropriate in applications where the microphone is exposed to turbulent pressure fluctuations and where the static pressure can vary rapidly with microphone position. These transducers are meant to be mounted directly on a vehicle during wind-tunnel tests; they may be mounted even on curved surfaces with the help of mounting pads. They are generally rugged, stable and corrosion-resistant. They also have built-in pre-amplifiers. The sensitivities of these microphones are optimized in order to achieve a good noise floor and still allow measurements at higher levels without clipping. The devices also support IEEE 1451.4 (TEDS) for remote identification and reading of their calibrated sensitivity.
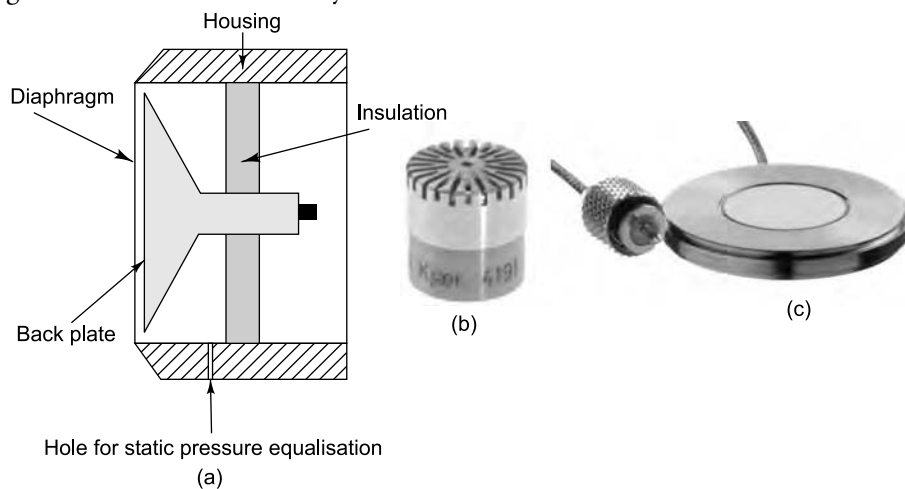
FIGURE 12.2 Condenser microphone: (a) schematic drawing, (b) photograph of condenser microphone, (c) photograph of surface microphone (Courtesy of Brüel & Kjær, Denmark)

TABLE **12.1**   Typical specifications of condenser microphones

Diaphragm size (in.)	Lower frequency (Hz)	Upper frequency (Hz)	Dynamic range (dB)	Sensitivity (mV/Pa)
1	2–6	8–18	–80 to145	50
1/2	3–7	9–40 K	–10 to 150	50
1/4	4	70–100 K	30–170	3.5
1/8	7	140	30–180	1.0

Advantages

  (i) Flat frequency response over the entire audio frequency range
 (ii) Low self-noise
(iii) High stability
 (iv) Reasonably high sensitivity
  (v) Fairly good stability with respect to time, temperature and humidity
 (vi) Small dimensions of the microphone making diffraction of sound negligible
(vii) Very low sensitivity to magnetic fields.

Disadvantages

  (i) Water droplets form on the diaphragm and backplate (electrode) when the microphone is taken from humid environment to a cool room. Therefore, the capacitance of the microphone changes.
 (ii) Polarization voltage of 200 V is required for the back plate.

12.2.3   Electret Capacitor Microphone

An electret capacitor microphone (ECM) also known as pre-polarized microphone is a type of condenser microphone which has overcome all the inconveniences of the traditional capacitor microphone, like the need for a polarization power supply. It is now used almost universally, in applications such as telephones and consumer electronics and has replaced the carbon microphone for general purpose applications. It is an excellent, easy to use and inexpensive device which can be operated on a 5 V supply voltage, but it normally requires an integrated pre-amplifier. The development of the polymer electret film and the field-effect transistor has spurred its growth. The externally applied charge described in Section 12.2.2 is replaced by a fixed charge in an electret material in this microphone. An electret is a stable high temperature polymer material that has been permanently electrically charged or polarized. The name comes from electrostatic and magnet. The electret contains trapped or 'frozen' electrical charges which produce the necessary electrical field in the air gap. The frozen charge remains inside the electret and stays stable for thousands of years. The static charges in the material get aligned and permanently embedded in an electret in the same way that a magnet is made by aligning the magnetic domains in a piece of iron. Electrets are made by first melting a suitable dielectric material such as a plastic or wax and then allowing it to re-solidify in a powerful electrostatic field, causing the static charges to align themselves to the direction of the electrostatic field to produce a permanent electrostatic 'bias'. The sensitivity of the microphone depends on the field strength in the air gap. Typically, the field strength and equivalent polarization voltage correspond to those used for microphones with external polarization, namely 200 V. These pre-polarized microphones, which are generally more complex than ordinary condenser microphones are mainly intended for use in portable battery operated and hand-held instruments.

There are three major types of electret microphones:

Foil-type or diaphragm-type: This is the cheapest type, where a film of electret material is used as the membrane or diaphragm itself. Obviously the quality of this device is poor, since the electret material does not make a particularly good diaphragm.

Back electret: In this type, the diaphragm is made of an uncharged material and an electret film is applied to the back plate of the microphone capsule. This design yields a transducer of better quality.

Front electret: This is the newest type, where the condenser is formed by the diaphragm and the inside surface of the capsule, the back plate being totally eliminated. Any conductive film may be used for the diaphragm in this design and the electret film is applied to the inside front cover.

Figure 12.3(a) shows the principle of operation of the foil type electret microphone and Fig. 12.3(b) shows photographs of electret microphones.
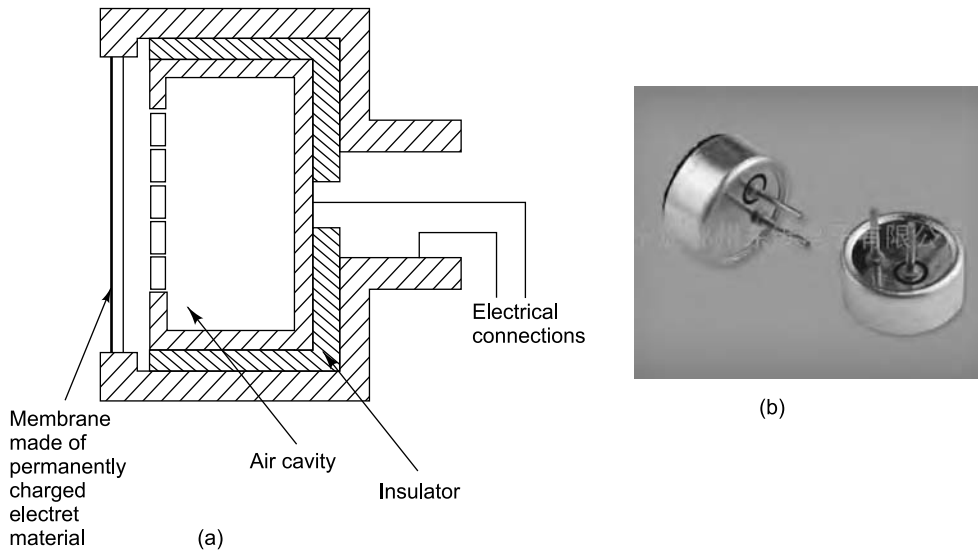
FIGURE 12.3   Electret microphone: (a) principle of operation, (b) photographs (Courtesy of www.ecvv.com)

Electret microphones were traditionally considered low-cost and low quality transducers. However, the best ones today match the capacitor microphones in almost all characteristics except low noise performance and can even offer the long-term stability and ultra-flat frequency response of a condenser microphone. The capacitor in the electret is carefully sealed away from floating charges so that the electret does not become neutralized. They are used in many applications, from high-quality recording to built-in microphones in small sound recording devices and telephones.

Advantages
  (i) This microphone is remarkably small.
 (ii) Polarization voltage is not required.
 (iii) It has omni-directional characteristics.
 (iv) It is very inexpensive.

Typical specifications

    Size: 9 mm diameter and 6 mm height
    Frequency bandwidth : 15 kHz
    Sensitivity   : 0.6 mV/µbar
    Power supply   : 2–10 V at <1 mA

12.2.4 Electrodynamic/Moving Coil Microphone

This microphone is often found in recording studios, broadcast and motion picture production, home hi-fi and video recording systems and on stages for live sound reinforcement. The principle of the dynamic microphone was known in 1877 when Bell developed the telephone, but lack of electronic amplification made it impossible to use. Dynamic microphones are similar to conventional loudspeakers in most respects. Both have a diaphragm (or cone) with a voice coil attached near the apex. Both have a magnetic system with the coil in its gap. The difference is in how they are used. A dynamic microphone operates like a speaker 'in reverse'. In fact, many intercom systems use small speakers with lightweight cones as both speaker and microphone, by simply switching the same transducer from one end of the amplifier to the other! The diaphragm, carrying a coil of fine wire placed in a strong annular magnetic field, is exposed to sound waves and is moved by changing sound waves. This moves the coil, which causes current to flow as lines of flux from the magnet are cut. The voltage e induced in the coil is proportional to its amplitude of vibration, which in turn, depends on the sound pressure and is expressed as

$$e = BLv \tag{12.2}$$

where B is the flux density in Tesla (Webers/m^2), $L = \pi dN$ is the total length of wire in the coil, d is the diameter of the coil and N the number of turns; v is the velocity of the diaphragm in m/s.

    Figure 12.4(a) shows the construction of a moving coil microphone and Fig. 12.4(b) shows a photograph. The domed diaphragm acts like a rigid piston and carries the coil of wire, which moves in the annular gap of the high magnetic field. The pole pieces of the microphone are of soft iron, with permanent magnets providing the field. The diaphragm is acted upon by the acoustic pressure, so that the microphone is a pressure microphone, with omni-directional characteristics.
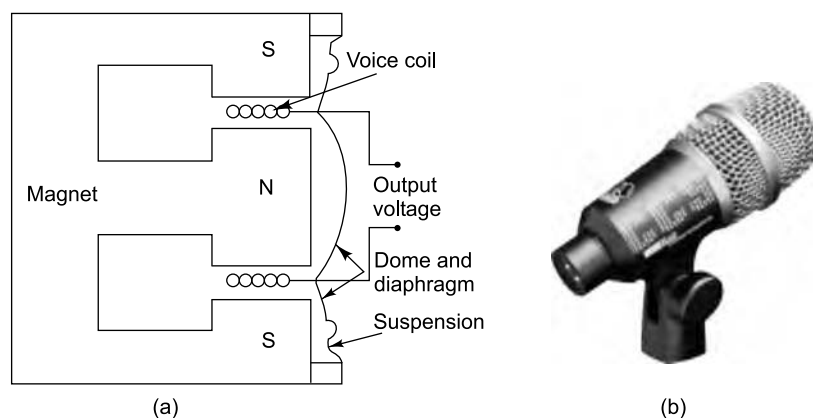
FIGURE 12.4   Electrodynamic microphone: (a) schematic drawing, (b) photograph
(Courtesy of www.spectrumaudio.com).

Advantages

  (i) Dynamic microphones are rugged and reliable.

  (ii) They do not need bias voltages or batteries or external power supplies.

  (iii) They are capable of smooth, extended response, or are available with tailored response for special applications.

  (iv) They have a fairly good sensitivity with an excellent signal-to-noise ratio.

  (v) Dynamic microphones are of the low-noise type and are relatively rugged.

  (vi) Most high-quality microphones are still dynamic microphones.

  (vii) They are not susceptible to moisture.

  (viii) They operate at fairly high temperatures.

  (ix) They need little or no regular maintenance and with reasonable care will maintain their performance for many years.

Disadvantages

  (i) Their sensitivity is quite low, no more than –90 or –100 dB, so amplification is essential. At least 40 dB can be gained with transformers, bringing the output up to –50 dB when applied to the amplifier, which is not too bad.

  (ii) The other drawback of these microphones is that they are susceptible to external magnetic fields.

  (iii) The high-quality microphones are relatively expensive.

Typical specifications

Internal impedance	: $<10\Omega$
Size of dynamic element	: 1″ diameter
Frequency response	: 40–18,000 Hz
Output level	: –53 dBV/Pa @ 1000 Hz

12.2.5 Piezoelectric Microphone

Piezoelectric microphones employ crystals that can become electrically polarized and produce voltages proportional to the strain. Since the piezoelectric effect is reversible, a piezoelectric microphone will function as a source when an alternating voltage is applied to its terminals. The description of the piezoelectric effect is made complicated by the many directional quantities and the crystal symmetries that enter. Rochelle salt is one of the first materials to be used in acoustic transducers. Single crystals of Rochelle salt have been employed widely in microphones. Unfortunately, the crystals deteriorate in the presence of moisture and get permanently damaged if the internal temperature exceeds 46 °C. Historically, quartz crystals have been of great significance as transducers. They are impervious to water as well as most corrosive materials, can be subjected to extreme temperatures and are easily manufactured. They can generate longitudinal waves, shear waves or combinations. Other useful materials are sintered ceramics including barium titanate, lead zirconate, lead titanate and mixtures of these and associated compounds. If a microphone is described as 'crystal', it usually contains Ammonium dihydrogen phosphate (ADP); if it is called 'ceramic', barium titanate is the active element.

    Piezoelectric microphones are often constructed in a manner similar to that shown in Fig. 12.5. Here, the sound waves act on a light diaphragm whose centre is linked to an end or corner of the piezoelectric element by means of a driving pin. Although a single element could be used, two elements are usually sandwiched together to form a bimorph. Stresses in the crystals, produced by a sound field, generate an

output proportional to the acoustic pressure. The voltage output of a bimorph element is proportional to its strain. Many designs incorporate a built-in pre-amplifier next to the crystal. This arrangement reduces the electrical noise and output impedance.

Advantages
They have satisfactory frequency response and are relatively high in sensitivity, low in cost and small in size. Piezoelectric microphones are therefore widely used in public-address systems, sound level meters and hearing aids.
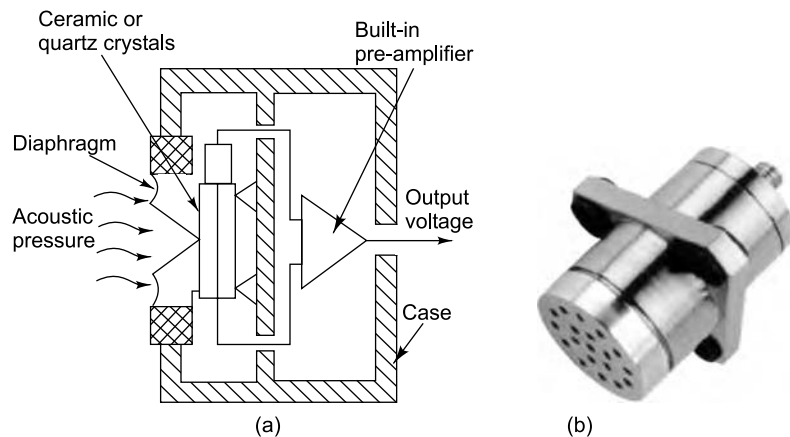
FIGURE 12.5   Piezoelectric microphone: (a) schematic drawing, (b) photograph
(Courtesy of Brüel & Kjær, Denmark)

Disadvantages
Piezoelectric microphones give low output at moderate internal impedance, and must always be used with amplifiers.

Specifications

Charge sensitivity	:	0.16 pC/Pa
Frequency response (±1 dB)	:	2–4 kHz
(±3 dB)	:	1–10 kHz
Dynamic range	:	100–180 dB

12.2.6   THE RIBBON MICROPHONE

The microphones discussed in Sections 12.2–12.6 are called pressure microphones because they are activated by the acoustic pressure acting on one side of a diaphragm. These microphones are essentially omni-directional because pressure is a scalar quantity. The ribbon microphone is a pressure gradient sensitive microphone and therefore has directional sensitivity. The pressure gradient is a vector quantity parallel to the direction of propagation and in phase with the displacement.

Let us consider a harmonic pressure wave $p(x,t)$ impinging on a 'surface' not 'diaphragm'. 'Surface' implies exposure to pressure equally on front and back, while 'diaphragm' implies exposure on the front side only. The net driving force will be equal to the product of the difference in pressures between the

two sides and the common surface area *A*. This pressure difference can be computed as the space rate of change of acoustic pressure (pressure gradient) $\partial p/\partial x$ multiplied by the effective acoustical distance *L* between the two sides of the surface. Let the pressure wave have an angle of incidence θ. When θ is zero, the sound wave is incident normally on the surface. The expression for the desired angle-dependent force *f* is as below:

$$f = -(\partial p/\partial x)\ AL \cos \theta = ikpAL \cos \theta = i\omega\rho_o uAL \cos \theta \qquad (12.3)$$

This force is at right angles to the pressure and is proportional to the propagation constant $k = \omega/c_0 = 2\pi/\lambda$ where c_0 is the speed of sound. This equation holds when $kL \ll 1$, that is, when the size of the microphone is small compared to the wavelength of sound. Equation (12.3) states that *u* is proportional to *p* and independent of frequency; this is true irrespective of whether the pressure wave is plane or spherical as long as the distance from the source is large and wavelength is small. If the latter conditions are not satisfied, the velocity will be inversely proportional to frequency, giving rise to what is called the proximity effect.

The most common pressure-gradient microphone is the ribbon microphone. Figure 12.6 shows a ribbon microphone. The pressure gradient sensitive surface or the ribbon is a corrugated aluminium foil and is suspended in a strong magnetic field. Moving the ribbon will change the emf generated which is proportional to *u*. The distance *L* is determined by the size of the baffle in which the ribbon is suspended. The generated emf is typically small; the velocity can be made higher by reducing the mass of the surface. The sensitivity is proportional to cos θ. The ribbon microphone is thus equally sensitive to sound coming from front and back and is completely insensitive to sounds coming from 90° or 270°. By combining the cosine sensitivity of the ribbon microphone with the isotropic sensitivity of a pressure microphone, we get a cardioid microphone where the response is proportional to (1 + cos θ). Such a microphone favours sounds coming from 0°, and discriminates against sounds coming from 180°.

Advantages

 (i) Ribbon microphones are very sensitive and accurate.
 (ii) They can also be oriented so that noise sources can be put in a zone of low sensitivity. These features have made the ribbon microphone the standard for broadcasting, and the lozenge-shaped shiny microphone a familiar sight.

(a)

(b)
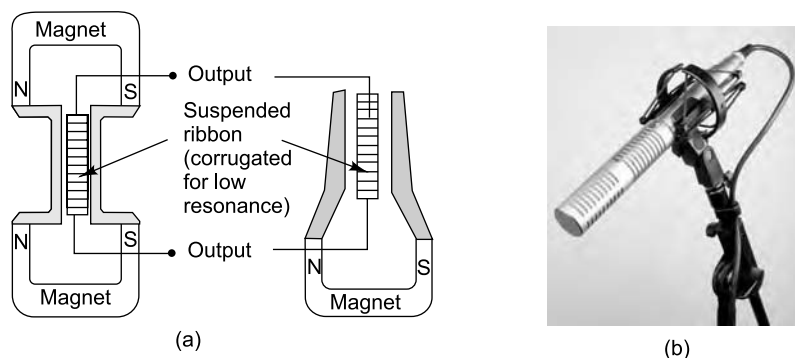
FIGURE 12.6   Ribbon microphone: (a) schematic drawings for two different designs, (b) photograph (Courtesy of www.cascademicrophones.com)

Disadvantage

They are very delicate and expensive to make and repair.

Specifications

Frequency response	:	20 Hz–20 kHz
Maximum SPL	:	132 dB
Output sensitivity	:	30 mV/Pa
Equivalent noise level	:	6 dB SPL, A-weighted as per IEC 651
Dynamic range	:	126 dB
Directional pattern	:	Bi-directional

12.3 ACOUSTIC EXCITERS

They convert electrical energy to acoustic energy. The commonly used loudspeakers are:

(a) electrodynamic loudspeaker (b) electrostatic loudspeaker (ESL). There is also a very special type of acoustic exciter called the electropneumatic transducer (EPT).

12.3.1 Technical Specifications of a Loudspeaker

Type of speaker: Depending on the frequency range of operation, speakers may be classified as woofers (low frequency), squawkers (mid-frequency), tweeters (high frequency) or full-range speakers.

Rated power: This is the maximum input power that can be given to a speaker before unacceptable distortion occurs or alternately before the loudspeaker gets thermally destroyed. Specifications are in terms of the nominal (or continuous) power and peak (or maximum short-term) power that a loudspeaker can handle. A woofer may be damaged at much less than its rated power if it is driven beyond its mechanical limits at lower frequencies (e.g. by heavy bass music). Tweeters can also be damaged by music or by a sine wave input at high frequencies due to amplifier clipping.

Impedance: This value is required for appropriate impedance matching with the power amplifier and is typically 4 or 8 Ω.

Frequency response: This is the range of frequencies over which the speaker produces a constant output and is the measured output or SPL when the speaker is powered over a specified range of frequencies for a constant input level. It is often quoted as the maximum deviation in SPL such as within ±3 dB. Generally a frequency response plot is provided by the manufacturer.

Sensitivity: This is the SPL produced by a loudspeaker and is measured along its axis in anechoic conditions (in a non-reverberant environment) typically at one or more specified frequencies when driven with an input of 2.83 V corresponding to the voltage across a standard 8 Ω speaker driven at 1 Watt. It is usually specified in dB SPL, and is measured at 1 m.

Maximum SPL (dB): This is the highest output the loudspeaker can produce, without getting damaged or without exceeding a particular distortion level. The rating given by manufacturers is often inflated and is misleading since it is given without reference to frequency range or distortion level.

Directivity: This can be represented by polar plots given at fixed frequencies as for the case of microphones. Curves of equal SPL are plotted representing the propagating wave front. At low frequencies sound propagation is omni-directional (spreads out in all directions); for high frequencies it becomes much more directional, separating into lobes.

Baffle or enclosure type: It has also been found that the production of a high-fidelity loudspeaker requires that the speaker be put in a closed box. Hence, the type of enclosure is mentioned for enclosed systems only: typically sealed, bass reflex, etc.

12.3.2 Electrodynamic/Moving Coil Loudspeaker

Figure 12.7(a) is a simplified drawing of a moving-coil loudspeaker which is nothing but a dynamic microphone in reverse. Of course, the designs are quite different, because in practice a loudspeaker with a large power rating cannot be used as a microphone. In the loudspeaker a light voice coil is mounted such that it can move freely inside the magnetic field of a strong permanent magnet. The speaker cone is attached to the voice coil and is connected by a flexible mounting to the outer ring of the speaker support. The magnetic field created by current flowing through the voice coil interacts with the magnetic field of the speaker's magnet, forcing the coil and its attached cone to move back and forth. This in turn moves the air to produce a sound which is a replica of the original electrical wave. The vibrating diaphragm (cone) is usually appreciably larger than the voice coil, to enhance the efficiency of radiation at the lower frequencies. Figure 12.7(b) shows the photograph of a dynamic speaker.
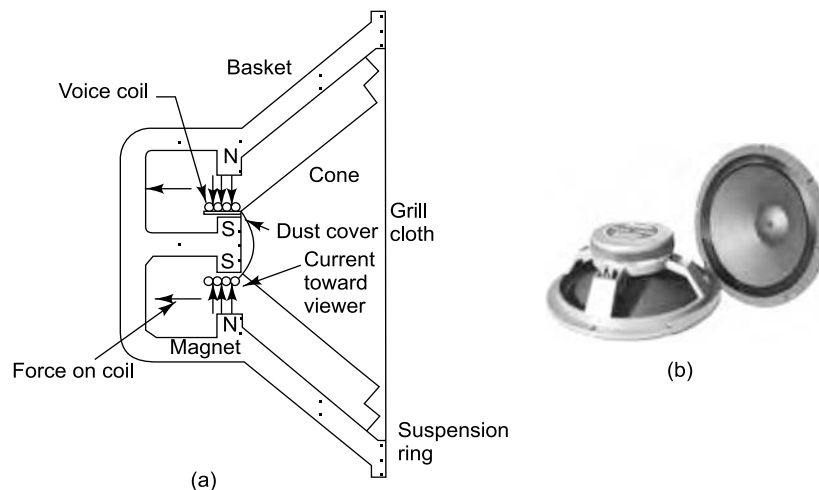
Figure 12.7   Moving coil loudspeaker: (a) schematic drawing, (b) photograph (www.iconicspkrs.com)

Loudspeakers are almost always the limiting element in the fidelity of a reproduced sound at home or in a theatre. The free cone speaker is very inefficient at reproducing sound wavelengths larger than the diameter of the speaker and becomes directive at high frequencies. For this reason, a number of speakers with different diaphragm diameters are often used in high-fidelity applications to cater to various frequency ranges. There are relatively massive speakers i.e. speakers of 25 to 30 cm diameter (woofers) radiating the lower frequencies (subwoofers and woofers) up to 500 Hz, smaller speakers for the mid-range frequencies (squawkers) i.e. speakers of 10 to 12.5 cm diameter and still smaller ones of around 7.5 cm diameter to radiate the highest frequencies (tweeters and super tweeters) for the range 300–20,000 Hz. More power is needed in the bass range making multiple drivers a practical necessity. These various speakers can be driven through electrical filtering networks that deliver to each its appropriate range of frequencies or can be band-limited naturally through the characteristics of their own electrical, mechanical, and radiation impedances.

Loudspeakers are generally enclosed for the reason that an un-enclosed speaker behaves as an acoustic dipole, giving rise to a poor low frequency response (because sound from the back of the diaphragm cancels that from the front) and highly directional radiation. By mounting the loudspeaker in an infinite baffle, it is forced to behave as a monopole and radiate into the half space in front of the baffle. To create such an infinite baffle, the baffle is folded around the back of the loudspeaker, or in other words the loudspeaker is put in an enclosure.

Advantage
They are small and inexpensive.

Disadvantages
   (i) The electrodynamic cone loudspeaker is a low impedance device with high current requirements.
  (ii) The total weight of the moving parts of a dynamic speaker: its voice coil, suspension system, and cone is much more than that of the air that the speaker drives, thus preventing response to high frequency signals.
 (iii) A magnetic speaker resonates at many frequencies, behaving somewhat like a bell. The transient response is poor since the 'ringing' continues long after the original note has stopped.
 (iv) It is not possible to achieve high efficiency (especially at low frequencies) along with compact enclosure size and adequate low frequency response. Therefore, one has to make a trade-off when designing a speaker system.
  (v) It has a low efficiency of around 5%, resulting in say 10 W of sound power for 200 W of electrical power from an amplifier.

Typical specifications

Impedance	:	4–8 Ω
Music Power rating	:	200 W
Power rating for sine wave	:	100 W
Frequency response ($\pm$ 3 dB)	:	40 Hz–20 kHz

12.3.3 Electrostatic Loudspeaker

The ESL uses a thin flat diaphragm sandwiched between two electrically conductive grids forming a pair of capacitor plates, with a small air gap between the diaphragm and grids. The diaphragm is usually made from a polyethylene terephthalate (PET) film of thickness 2–20 μm, with exceptional mechanical properties and with a conductive coating material such as graphite. The membrane is electrostatically charged with a high DC polarizing voltage, while the electrodes are fed with ground potential. Typical polarizing voltages are of the order of 2–3 kV. One plate is held stationary, while the other, the diaphragm, moves in response to mechanical or electrical excitation. If a transient voltage is applied across the plates, the diaphragm moves in response to the changing charge. For low distortion operation, the diaphragm must operate with a constant charge on its surface, rather than with a constant voltage. The diaphragm's conductive coating is chosen and applied in such a manner as to give it a very high surface resistivity. The grids are driven by the audio signal, front and rear grids being driven in anti-phase. Using grids on both sides cancels out non-linearity. As a result, a uniform electrostatic field proportional to the audio signal is produced between the grids. This causes a force to be exerted on the charged diaphragm, and its resulting movement drives the air on either side of it. The grids must be able to generate as uniform an electric field as possible, while still allowing sound to pass through, and should be perfectly flat. Suitable

grid constructions are therefore perforated metal sheets, a frame with tensioned wire or wire rods. Figure 12.8 shows an electrostatic speaker.

The principle of operation of an ESL is as follows: A charge, Q, placed in a field of strength $V_{sig}/2d$ experiences the following force:

$$F = QV_{sig}/2d \tag{12.4}$$

where F is the force (N), V_{sig} is the signal voltage (V), Q is the strength of the charge (C), and $2d$ is the distance between the electrodes (m), d being the distance between the diaphragm and one electrode. Since $Q = CV_{pol}$ = constant, the total capacitance between the plates will vary as the diaphragm moves and the voltage falls. Here C is the total capacitance (F) of the system, and V_{pol} is the polarizing voltage (V).
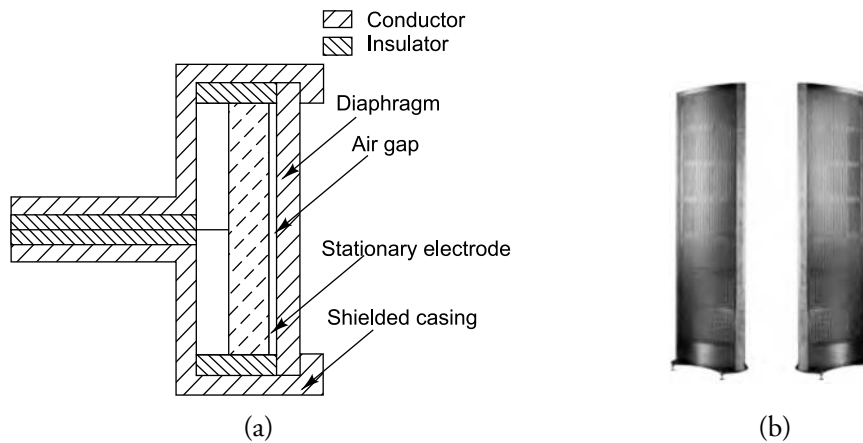
FIGURE 12.8   Electrostatic loudspeaker: (a) schematic drawing, (b) photograph
(Courtesy of http://www.kingsaudio.com).

Advantages

  (i)   The main advantage of the speaker is the extremely lightweight of the diaphragm, due to which it offers a faithful reproduction (both in amplitude and phase) of high frequency content, making it superior to the heavy moving coil speaker. This is highly desirable for music, since overtones and higher harmonics contribute significantly to the overall richness of the sound.

 (ii)   Most electrostatic speakers are tall, flat and thin in design and generally do not have enclosures, acting as vertical dipole line sources with highly directional characteristics. This brings about different acoustic behaviour in comparison with conventional electrodynamic loudspeakers. Curved panels are also used, making the positioning of the speakers in an enclosure a little less stringent.

(iii)   Since electrostatic speakers do not need an enclosure, problems associated with cabinet design can be got rid of.

(iv)   With rapid progress in the research and development of materials and cheaper polymers, these speakers are bound to become more popular.

Disadvantages

  (i)   The chief disadvantage is poor bass response due to phase cancellation arising from the lack of an enclosure, and the difficulty in reproducing low frequencies with a diaphragm with small allowable

displacement. This is often remedied with a hybrid design using a dynamic loudspeaker to handle lower frequencies with the electrostatic diaphragm handling middle and high frequencies.

(ii) The directionality of these speakers can also be a disadvantage since the area where the effect of proper stereo imaging can be heard in an auditorium becomes small.

(iii) These speakers are sensitive to ambient humidity levels.

(iv) They have a tendency to attract dust, insects, conductive particles and moisture, leading to deterioration in diaphragms and necessitating periodic replacement.

(v) They are large in size and expensive, putting them out of the reach of most consumers.

Typical specifications

Frequency response : 50 Hz–25 kHz
Sensitivity   : Equivalent to 84 dB/W @1 m
Impedance   : 8 Ω
Maximum power   : 260 W (music)

12.3.4   Electropneumatic Transducer

This is a highly specialized transducer which is capable of producing an enormous amount of acoustic power, of the order of a few 1000s of Watts resulting in SPLs of the order of 150 dB or more and is used for conducting high level acoustic tests. It is an electrodynamically controlled air modulator which converts the pressure energy of air into acoustic energy quite efficiently and can reproduce almost any type of random signal fed into its coils. The principle of operation of this transducer is similar to that of a siren. The transducer consists of a magnetic structure which houses and supports a pneumatic valve. The magnetic structure is designed to receive and direct air flow through an air filter to the pneumatic valve. This electrodynamically operated pneumatic valve responds to any applied electrical waveform and controls or modulates the velocity of the air stream supplied from a compressor. The resulting velocity fluctuations superimposed upon the direct flow velocity generate acoustic energy. This sound output is generally sent out through a horn which concentrates the sound into a directed beam so that most of the energy is radiated in one direction. The basic elements of the EPT are shown in Fig. 12.9. Air is delivered at constant pressure from a pneumatic compressor to the plenum chamber from where it escapes through one or more orifices. The orifice areas, which are controlled by the action of the electrodynamic driver, modulate the air stream velocity and produce acoustic energy.
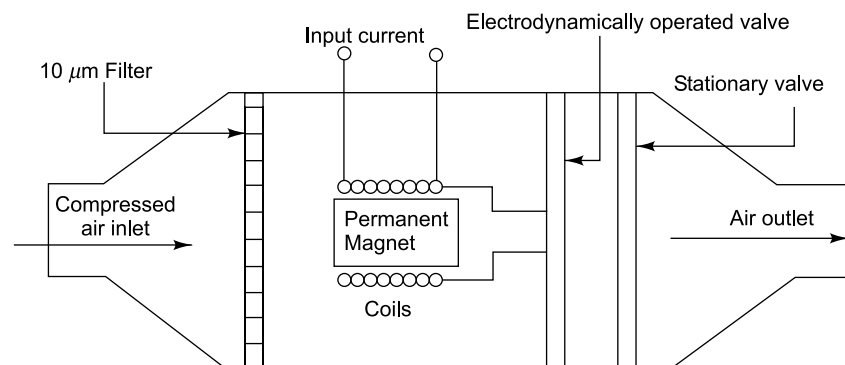
FIGURE 12.9   Electropneumatic transducer

Figure12.10(a) shows the EPT and horn assembly. Figure 12.10(b) shows a test plate (mounted on a test frame) to be subjected to acoustic excitation.

(a)   (b)
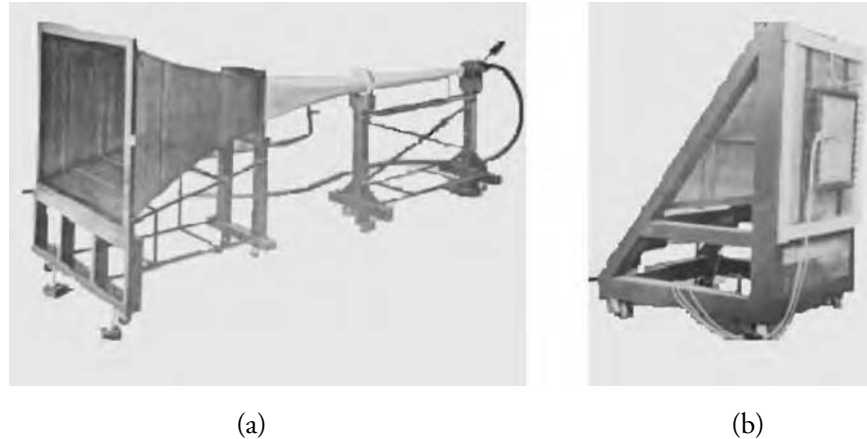
FIGURE 12.10   Acoustic excitation using EPT: (a) EPT with horn, (b) test frame with test plate
(Courtesy of IIT Madras, Chennai)

Advantages
A very large amount of acoustic power is obtained from one source; no dynamic speaker can deliver this amount of power.

Disadvantage
The transducer needs a steady supply of compressed air.
Specifications

Rated acoustic power	:	4000 W
Maximum air flow rate	:	12.7 m^3/min
Maximum required air pressure	:	2.76×10^5 Pa
Maximum current	:	6 A (RMS)
Nominal impedance	:	2 Ω at 100 Hz
Nominal diameter	:	21.5 cm

12.4   ACOUSTIC CALIBRATORS

This section explains techniques for the calibration of microphones. The calibration can either be performed in the field or in a calibration laboratory. Calibration essentially means determination of the sensitivity of the device which is generally expressed in terms of mV/Pa. The method used to calibrate a microphone must have a known amount of uncertainty. Calibration is important for the measurement to be recognized by legal authorities or if compliance with international standards is to be claimed. For absolute confidence in the results, it is advisable to perform a field calibration before and after a measurement, irrespective of whether the measurements are absolute or relative.

Most sound level calibrators are portable, easy-to-use and characterized by the production of a well-defined sound pressure at a single frequency, usually in the range of 200 Hz–1 kHz, at which

the calibration is performed (IEC 942 'sound calibrators'). To perform calibrations across the entire frequency range, a multitone calibrator may be used to check the performance of the measurement system. Such a calibrator provides a number of pure tones at single frequencies in steps of one octave. When using calibrators that produce a single frequency, the calibration is strictly valid only at that reference frequency. However, considering the fact that microphones generally have a flat frequency response over a very wide range of frequencies, it can be expected that they will give the same electrical output at all frequencies in the specified frequency range for sound pressures of equal magnitude, making calibration at a single frequency sufficient.

12.4.1 Piston Phone

A piston phone is a calibrated reference sound source which provides a quick and simple means of calibration in the field or laboratory conditions for calibration of precision condenser microphones and high sensitivity piezoelectric pressure sensors. It is a self-contained, easy to operate, battery-powered instrument which generates a very stable and accurate reference SPL for calibration by the to and fro movement of a piston in a coupler cavity. The reference SPL is generated by varying the volume of the cavity at a controlled frequency by a known displacement, and is proportional to the variation of the coupler volume, i.e. to the stroke and area of the piston. The cavity volume is that which is enclosed at the head of the piston phone by the transducer and its dedicated coupling adaptor which ensures that an equal cavity volume is maintained irrespective of the sensor used. Two opposed reciprocating pistons which are driven by a precision-machined, rotating cam disc alter the cavity volume by a known displacement at a tachometer-controlled frequency. A pressure change accompanies the change in volume of the cavity, resulting in the accurate reference SPL. The ambient barometric pressure, temperature, and humidity influence this SPL. A barometer is normally included to provide a correction for the ambient pressure. The error in SPL is a function of the error in the measurement of the stroke. The piston phone allows correct calibration to be made even in very noisy surroundings.

When such a calibrator is used for microphone calibration, the well-defined sound pressure output from the calibrator is applied to the front of the microphone and the sensitivity may be determined by dividing the output voltage by the sound pressure. When the calibrator is used to calibrate the entire measurement channel or sound level meter, the sound pressure is applied to the front of the microphone as before and a gain adjustment is made in the measurement system or sound level meter to give the correct reading. The sound calibrator always establishes a pressure field and hence a suitable correction has to be applied for a

FIGURE 12.11 Photographs of calibrators (Courtesy of Brüel & Kjær, Denmark)

free field calibration. It is to be noted that when calibrating a sound level meter, the weighting filter should not be switched on. When using a piston phone with a calibration frequency of 250 Hz, a signal reading which is 8.6 dB less than the actual value is got if the A weighting filter is switched on. (Please see Section 13.2.1 for information on weighting filters.) Figure 12.11 shows photographs of some calibrators.

Typical specifications

Generated SPL	:	134 dB SPL (±0.09 dB) (100 Pa)
Reference conditions	:	ambient pressure = 101.3 kPa, ambient temperature-
		20 °C, ambient Humidity = 65 % RH
Frequency	:	250 Hz (±0.5 %)
Distortion	:	<1.5 %
Coupler volume	:	20 cm^3

12.4.2 Electrostatic Actuator Calibration

The electrostatic actuator is a reliable method for determination of a microphone's frequency response under laboratory conditions. It is applicable to microphones with a metallic diaphragm. The actuator comprises a stiff and electrically conducting metal plate which forms an electrical capacitor together with the microphone diaphragm. The metallic grid is positioned close to (at approximately 0.5 mm of) the diaphragm of the microphone. By applying high voltages to the actuator, electrostatic forces equivalent to a pressure-field or sound pressure of approximately 100 dB are produced on the microphone diaphragm. The actuator method has an edge over sound based methods in that it provides a simpler means to produce a well-defined calibration pressure over a wide frequency range, but its accuracy is not high enough for determination of the microphone sensitivity, which has to be done with an acoustic calibrator/ piston phone at a reference frequency. This method cannot simulate a free-field or a diffuse-field condition and appropriate corrections must be made if such calibrations are required. The electrostatic force or pressure produced by the actuator is practically independent of environmental factors.

The actuator is not suited for absolute calibration since the sound pressure produced is heavily dependent on the distance between the actuator and the diaphragm. Besides, this method requires that the system use high voltages and that the grid be removable.

FURTHER READINGS

1. American National Standards Institute, ANSI S1.40—*Specifications for Acoustic Calibrators*, 1984, R1997.
2. Barron, R.F., *Industrial Noise Control and Acoustics*, CRC Press, Boca Raton, Florida, 2003.
3. Beranek, L. L., *Acoustic Measurements*, John Wiley & Sons Inc., New York, 1988.
4. Beranek, L.L., *Acoustics, Acoustical Society of America*, Woodbridge, New York, 1993.
5. Bies, D.A. and Hansen, C.H., *Engineering Noise Control: Theory and Practice*, Taylor & Francis, New York, USA, 2003.
6. Brüel & Kjær, *Microphone Handbook*, Vol. 1, Theory, BE 1447-11, Denmark, 1996.
7. Crocker, M.J., *Encyclopedia of Acoustics*, Wiley-Interscience, New York, 1997.
8. Crocker, M.J., *Handbook of Acoustics*, Wiley, New York, 1998.
9. Crocker, M.J., *Handbook of Noise and Vibration Control*, John Wiley and Sons, Hoboken, New Jersey, 2007.
10. Davis, D. and Davis, C., *Sound System Engineering*, Focal Press, Boston, USA, 1997.
11. Everest, F.A. and Pohlman, K.C., *Master Handbook of Acoustics*, McGraw-Hill, New York, 2000..

12. Groves, I.D., Jr., *Acoustic Transducers* (*Benchmark Papers in Acoustics*), Hutchinson Ross Publishing Company, Stroudsberg, Pennsylvania, USA,1981.
13. Harris, C., *Handbook of Acoustical Measurements and Noise Control*, Publisher: Springer-Verlag, New York, LLC, 1998.
14. Hemond, C.J., *Engineering Acoustics and Noise Control*, Prentice Hall, Englewood Cliffs, New Jersey, 1983.
15. Huber, D.M. and Williams, P., *Professional Microphone Techniques*, Mix Books, Emeryville, CA, 1998.
16. *International Electrotechnical Commission, IEC 942*, Sound Calibrators, Class 1, 1988.
17. Kamichik, S., *Practical Acoustics*, Prompt Press, Edinburgh, 1999.
18. Kinsler, L.E. and Frey, A.R., *Fundamentals of Acoustics*, John Wiley & Sons, New York, 1982.
19. Kinsler, L.E. Frey, A.R., Coppens, A.B. and Sanders, J.V., *Fundamentals of Acoustics*, Wiley-VCH, 1999.
20. Kleppe, J.A., *Engineering Applications of Acoustics*, Artech House Publishers, Boston, Massachusetts, May 1989.
21. Peters, R.J. and Reeves C.W., *The Noise and Acoustics Handbook*, Coxmoor Publishing Co., UK, 1998.
22. Ristic, V.M., *Principles of Acoustic Devices*, Wiley, New York, 1983.
23. Rossi, M. and Roe, P.R.W., *Acoustics and Electroacoustics*, Artech House Publishers, Boston, Massachusetts,1988.
24. Talbot-Smith, M., *Audio Engineer's Reference Book*, Focal Press, Boston, SA, 2001.
25. http://www.bksv.com
26. http://www.hyperphysics.phy-astr.gsu.edu

Common Measurements in Acoustics

13.1 INTRODUCTION

Noise and vibration go hand-in-hand with noise being a part of the energy radiated by vibrating objects, but the methods of measurement of noise and vibration differ significantly. Noise measurements may involve simple sound level measurements or detailed statistical and frequency analysis of the signal; or they may involve more complicated sound intensity or power measurements. The transducers and techniques used for vibration and sound measurement and analysis are different. This chapter describes the measurement of sound pressure level, sound power, acoustic intensity, reverberation time (RT), sound power absorption coefficient and transmission loss (TL). The instrumentation required and assumptions made in measurement have been outlined.

13.2 SOUND LEVEL MEASUREMENT

Sound pressure has always been the acoustic parameter of measurement interest since it is this quantity that the human ear detects. Microphones have and are being widely used for the measurement of sound pressure in spite of the fact that modern technology has made it possible for sound intensity and particle velocity measurements to be made quite easily. The sound level meter (SLM) is the basic instrument designed for the measurement of sound, both objectively as well as subjectively, in approximately the same way as the human ear responds. It has already been mentioned in Chapter 11 that the decibel (dB) is chosen as the scale for sound pressure level measurement since this logarithmic scale compresses the entire range of sound pressures of practical interest in audio-acoustics from 20 µPa, which is the threshold of audibility, to 20-63 Pa, which is the threshold of pain (when the source is radiating at a frequency of 1 kHz). Sound pressure level depends on the distance and direction with respect to the source, as well as the nature of acoustical environment, such as free field or reverberant or diffuse field.

The chief features of the SLM are:
- Adjustable preamplifier
- Frequency weighting networks: Linear, *A*, *B*, *C* and *D*
- Internal 1/1 and 1/3 octave filters or an output to external filters
- Main amplifier
- Detector circuit and
- Indicating meter having 'fast', 'slow', 'peak' and 'impulse' responses.

It is required that the acoustical pressure fluctuations be measured with a suitable electro-acoustic transducer which most faithfully converts the sound pressure into voltage. Hence, the microphone constitutes the front end of any SLM. The most suitable microphone for the SLM is the condenser microphone. The electrical output of the microphone being quite small, it is amplified by a preamplifier before being processed. In most SLMs, the microphone is attached directly to the instrument. In others, the microphone and its preamplifier constitute a separate unit with a shielded cable connecting it to the instrument proper. The SLM itself is usually shaped to avoid disturbances due to its own reflection of sound on the microphone. The effect of operator holding the meter should also be carefully considered.

The indicating meter (with a logarithmic scale in dB) usually meets the need for subjective and objective measurements of noise. The former measurement simulates human binaural perception of sound quantitatively from a source, for which use of internationally standardized *A*, *B*, *C* and *D* frequency response weighting networks is made; these impart the characteristics of the human ear to the measuring instrument. The objective measures are indicative of the true sound from sources, correlating well with their vibrations; for this the Linear weighting network is made use of. The SLM is designed to cater to both these needs in the field with portability and reliability. Some SLMs come with built-in 1/1 and 1/3 octave analysers for a simple form of spectrum analysis.

Other optional features include a calibrated logarithmic attenuator, as well as peak and impulse reading capabilities using a detector circuit to capture the properties of the fluctuating sounds. The details regarding the detector circuit for time averaging are given in Section 13.2.4. The last stage of an SLM is the read-out unit which displays the SPL in dB (LIN) or dB (*A*). Most modern SLMs have digital displays, indicating the maximum RMS value measured using a rectifier. The signal may also be available at output sockets, in either AC or DC form, for connection to external instruments such as tape recorders or analysers, to allow recording and/or further processing. The piston phone or acoustic calibrator is also a standard accessory for an SLM. The essential components of the SLM are indicated in the block diagram in Fig. 13.1. Figure 13.2 shows a photograph of the meter.
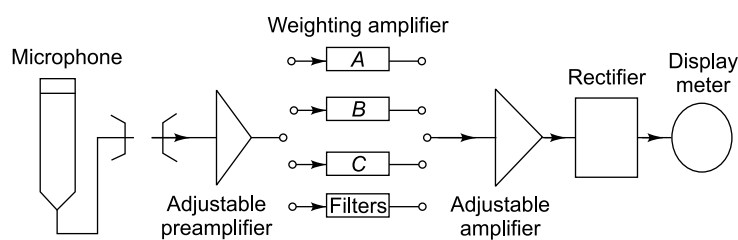
FIGURE 13.1   Sound level meter

FIGURE 13.2   Photograph of sound level meter (Courtesy of Brüel & Kjær, Denmark)

13.2.1   A, B, C, D and LIN Frequency Weighting Networks

There are generally four types of weighting networks, which adjust the response of the instrument to simulate the response of the human ear, which is logarithmic in its response to both loudness and frequency. These networks fall within the limits specified by American National Standards Institute (ANSI) or International Electrotechnical Commission (IEC). The function of the *A*-weighting network

is to mimic the response of the human ear. Its characteristic is based on the historical equal-loudness contours and is matched reasonably well to human hearing sensitivity for a wide range of sounds. This weighting is stipulated for virtually all governmental and industrial regulations, as well as noise exposure studies and is also frequently used for environmental noise measurements. It has good correlation with hearing damage. Originally, the *A*-weighting was meant only for quiet sounds with SPLs in the vicinity of 40 dB, but has now been made mandatory for all levels.

The linear weighting or LIN scale gives an overall value as would be sensed with a microphone having a flat frequency response in the audio-acoustic frequency range. This weighting network which provides a nominally uniform response up to 20 kHz is used for the objective technical evaluation of noise from sources. The *B* and *C* scales are meant to be used when the noise is louder than that which would be measured with an *A* scale. *C*-frequency weighting is meant for describing the loudness of industrial noise and is used for the measurement of the peak value of a noise in some legislations; it is fairly 'flat', with only a small attenuation at both low and high frequencies. Some of the earlier instruments also had the *D* network for evaluation of air traffic noise. This network had a gain about 10 dB higher than that of the *A* scale in the 3-4 kHz region. It may be said that the *B* and *D* scales are obsolete now. The frequency responses of the *A*, *B*, *C*, *D* and LIN weighting networks are shown in Fig. 13.3.
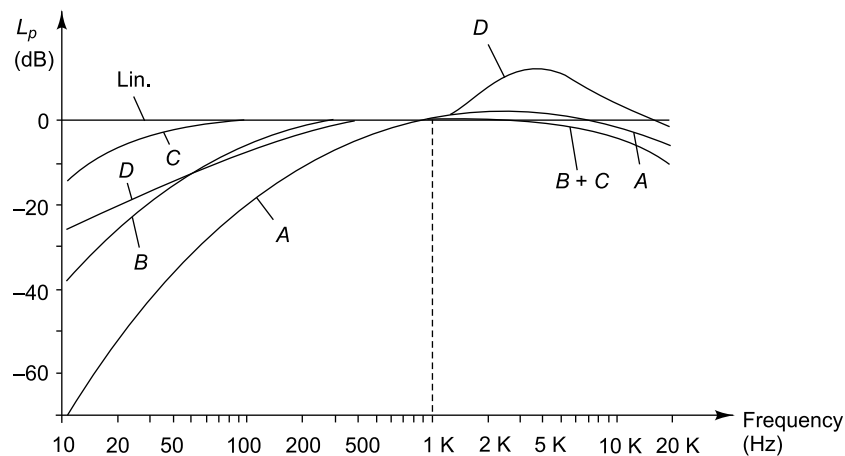
FIGURE 13.3 Response of weighting networks

The weighted SLM gives only a single number reading for sound pressure level. It does not give any information regarding the frequency content of the signal. However it can be used to find out whether a noise signal contains frequencies primarily above 1 kHz. To do this, measurements should be made using both the *A* and *C* weighted networks. The *A*-weighted sound pressure level in dB (*A*) should be subtracted from the *C*-weighted sound pressure level in dB (*C*); if the difference is large, the signal is primarily composed of frequencies less than 1000 Hz, if it is small, it consists of frequencies above 1000 Hz.

13.2.2   1/1 and 1/3 Octave Filters

For vibration measurements, constant bandwidth narrow band analyzers may be used. For acoustic measurements on the other hand, the entire frequency range from 20 to 20,000 Hz, i.e. the audio

frequency range is divided into octave or one-third octave bands. The SLM is coupled to 1/1 or 1/3 octave analysers to obtain the spectral distribution of the sound energy. Such analysers are constant percentage bandwidth analysers and are used for analysis of acoustic signals, the reason being that the ear responds to frequencies in a logarithmic fashion. The advantage of constant percentage bandwidth filters is that two neighbouring filters combine to one filter with a flat top, but with double the width (on a logarithmic scale). The octave bands cover the range of frequencies from 20 Hz to 20 kHz with a number of 1/1 octave bandpass filters. The term 'octaving' means doubling and each filter has a centre frequency and bandwidth twice that of the previous one. The 1/1 filter is the widest amongst octave filters used, with a bandwidth of 1 octave. But generally a higher frequency resolution is desired for spectrum analysis of a noise signal than that afforded by this weighting network. Such a refinement can be achieved by using subdivisions into smaller bandwidths. For example in a 1/3 octave band analyser, each 1/1 octave filter is divided into three filters with the centre frequency of the central filter being the same as that of the original 1/1 octave bandpass filter. The 1/1 and 1/3 octave filters are called constant percentage bandwidth filters, with the 1/1 octave filter having a bandwidth close to 70% of its centre frequency and the 1/3 octave filter, 23% of its centre frequency. The 1/3 octave filters are the more popular of the two, their advantage being that their bandwidth at frequencies above 500 Hz corresponds well to the frequency selectivity of the human auditory system.

When a set of 1/1 or 1/3 octave bandpass filters are used to analyse a noise signal, only the sound energy with frequency components contained in the filter passband of each filter is allowed to pass through the respective filters. Such filters are sufficient if the energy in the sound signal is fairly evenly distributed and relatively flat over a rather broad range of frequencies. However, often a signal may contain significant amounts of energy at one or more discrete frequencies. Clues to this can be obtained by noting whether the SPL in a particular 1/1 or 1/3 octave band is much higher than in the others, or whether a signal sounds as if it contains some pure tones. When this is the case, it may be necessary to conduct a more refined analysis of the signal to determine the frequencies at which these high energy levels exist. This can be accomplished by obtaining the spectrum levels of the sound signal using constant bandwidth filters with narrow bandwidths.

If f_c is the centre frequency of a bandpass filter, f_u the upper frequency limit and f_1 the lower frequency limit, then the relationships as given in Table 13.1 exist for 1/1 and 1/3 octave band filters. Table 13.2 shows the lower, upper and centre frequencies of 1/1 and 1/3 octave band filters. It is seen from these tables that the bandwidths steadily widen as the centre frequencies increase. Figure 13.4 shows examples of 1/1 and 1/3 octave pass bands.

TABLE **13.1**   1/1 and 1/3 octave band-filter frequencies

	1/1 octave bandpass filter	*1/3 octave bandpass filter*
Upper cutoff frequency f_u	$f_u = f_c \times 2^{1/2} = 2f_1$	$f_u = f_c \times 2^{1/6} = 2^{1/3}f_1$
Lower cutoff frequency f_1	$f_1 = f_c \times 2^{-1/2}$	$f_1 = f_c \times 2^{-1/6}$
Centre frequency of next band	Present $f_c \times 2$	Present $f_c \times 2^{1/3}$

For more detailed analysis, very narrow bandpass filters are available today. Their centre frequencies are obtained from the lowest centre frequency by multiplying them successively by $2^{1/n}$ where n may be equal to 3, 4, 5, etc. Filter bandwidths down to 1/96 octave have been realized. Table 13.3 shows the frequency ratios for typical narrow bands commencing from 31.5 Hz.

TABLE 13.2   Preferred band limits and centre frequencies (Hz) for 1/1 and 1/3 octave bands

Band	1/1 Octave bands			1/3 Octave bands		
	f_l	f_c	f_u	f_l	f_c	f_u
12	11	16	22	14.1	16	17.8
13				17.8	20	22.4
14				22.4	25	28.2
15	22	31.5	44	28.2	31.5	35.5
16				35.5	40	44.7
17				44.7	50	56.2
18	44	63	88	56.2	63	70.8
19				70.8	80	89.1
20				89.1	100	112
21	88	125	177	112	125	141
22				141	160	178
23				178	200	224
24	177	250	355	224	250	282
25				282	315	355
26				355	400	447
27	355	500	710	447	500	562
28				562	630	708
29				708	800	891
30	710	1000	1420	891	1000	1122
31				1122	1250	1413
32				1413	1600	1778
33	1420	2000	2840	1778	2000	2239
34				2239	2500	2818
35				2818	3150	3548
36	2840	4000	5680	3548	4000	4467
37				4467	5000	5623
38				5623	6300	7079
39	5680	8000	11,360	7079	8000	8913
40				8913	10,000	11,220
41				11,220	12,500	14,130
42	11,360	16,000	22,720	14,130	16,000	17,780
43				17,780	20,000	22,390
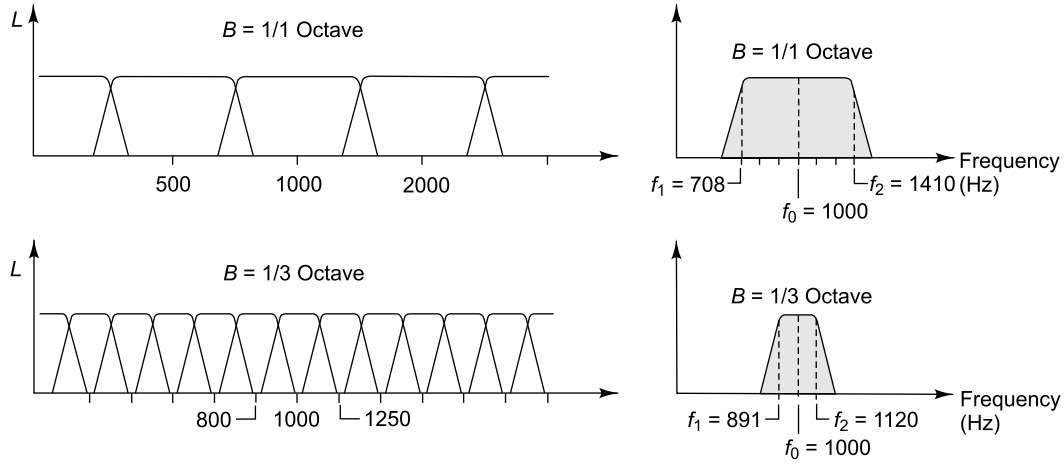
FIGURE 13.4   Octave and third-octave filters (Courtesy of http://www.bksv.com)

TABLE 13.3   Frequency ratios for narrow band octave filters

Centre frequency	Frequency ratio
Octave	2
One third octave (1/3)	1.25
One quarter octave (1/4)	1.1892
One fifth octave (1/5)	1.1486
One sixth octave (1/6)	1.1225
One eighth octave (1/8)	1.0905
One tenth octave (1/10)	1.0717

13.2.3   Spectrum Level

The spectrum level of a sound signal is defined as the sound level contained in a filter with a frequency bandpass of 1 Hz centred at a frequency f. The spectrum level is usually measured with constant bandpass, narrow band frequency analyzers. With respect to the sound pressure level, if the spectrum level $L_{P(SL)}$ of the signal is known, the band sound pressure level $L_{p(BL)}$ over a specified band (a measure of the total energy contained in the band) is obtained as

$$L_{P(BL)} = 10 \log_{10} \left[\int_{f_1}^{f_u} \frac{p_{rms}^2(f)}{p_{ref}^2} df \right]$$ (13.1)

If the sound energy is fairly flat and evenly distributed over the frequency band of interest, the above equation can be simplified as

$$L_{P(BL)} = 10 \log_{10} \left(\frac{p_{rms(avg)}^2 \Delta f}{p_{ref}^2} \right)$$ (13.2)

where $p_{rms(avg)}^2$ is the average mean square value of sound pressure in the bandwidth Δf. The band sound pressure level can be further simplified as

$$L_{P(BL)} = L_{P(SL)} + 10 \log_{10} \Delta f$$ (13.3)

13.2.4   Time Averaging

The sound level value is displayed as a reading on an indicating meter or as a digital value, after exponential time averaging of the signal. If the sound level fluctuates too fast, it is not possible to get a meaningful reading. For this reason, two detector response characteristics were standardized: the *F* (for Fast) and *S* (for Slow) characteristics. *F* typically has a time constant of 125 ms and provides a fast reacting display response. *S* has a time constant of 1 s, giving a slower response which helps average out the display fluctuations, which would otherwise be impossible to read using *F* mode. Besides there are two more forms of averaging: *I* for impulse and *P* for Peak hold. Figure 13.5 shows the detector characteristics in these various modes of operation.
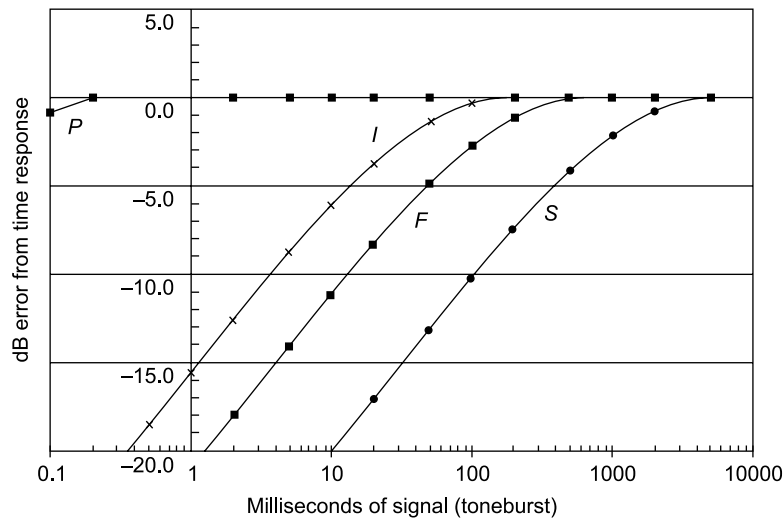
FIGURE 13.5   Characteristics of detector circuit

Impulsive noise consists of one or more bursts of sound energy, each of duration less than about 1 s. Impulses are characterized by the peak pressure, the rise time and the duration of the peak. The *I* or impulse mode is to be used if the sound to be measured consists of isolated impulses or a large amount of impact noise; under such circumstances, the normal *F* and *S* detector responses are not short enough to detect and display transient noise in a manner similar to the human perception of impulsive sounds. The *I* characteristic has a time constant of 35 ms. The perceived loudness of short duration sounds is typically lower than that of steady continuous sounds, but the risk of damage to hearing is not necessarily lower. For this reason, some SLMs include a circuit for capturing either the peak value of a sound or its maximum RMS value, irrespective of its duration during the monitoring period. This is done by using a hold circuit. Time constant for the peak detector is 50 μs.

13.2.5   Integrating Sound Level Meters

The hearing damage potential of a given sound environment depends not only on its level, but also on its duration. Therefore both these quantities must be measured to indicate the energy received. For constant sound levels, this is easy, but if the sound level keeps varying, a single integrated value computed

over a well defined sampling period should be used. For the measurement of fluctuating noise, such as traffic noise or for community noise-annoyance assessments, integrating SLMs are available to measure energy equivalent level L_{eq}, which is nothing but the equivalent continuous sound level over a prescribed time. In the absence of any uniform method for determining community reaction to intruding noise, equivalent continuous noise level L_{eq} has become the basic measure. For an *A*-weighted L_{eq}, the symbol L_{Aeq} is used and may be computed according to the equation shown below.

$$L_{Aeq} = 10 \log_{10} \frac{1}{T} \int_0^T \left[\frac{p_A(t)}{p_{ref}(t)} \right]^2 dt \qquad (13.4a)$$

where T is the total measurement time, $p_A(t)$ is the *A*-weighted continuous acoustic pressure and p_{ref} is the reference acoustic pressure (20 µPa).

This L_{eq} has the same energy content and consequently the same hearing damage potential as the fluctuating sound. It is customary to calculate a noise exposure level normalized to a nominal 8-hr working day. When the *A*-weighted SPLs over a period of time are presented in the form of a histogram,

$$L_{eq} = 10 \log_{10} \left[\frac{1}{100} \left\{ f_1 10^{L_1/10} + f_2 10^{L_2/10} + \ldots + f_n 10^{L_n/10} \right\} \right] \qquad (13.4b)$$

Here f_i denotes the percentage of the total time for which the *A*-weighted noise levels lie within the band interval with the corresponding centre frequency and L_i is the *A*-weighted level. L_{eq} is the basis for calculating L_{DN}, the day-night average sound level and L_{NP}, the noise pollution level. Day-night average (DNL) or L_{DN} is defined as the *A*-weighted equivalent sound pressure level expressed in dB averaged over a period of 1 day or 24 hours, with an additional level of 10 dB imposed on the equivalent sound levels at night, i.e. between 10 p.m. and 7 a.m, to account for increased annoyance due to noise during the night hours. This average is used to define the level of noise a community is exposed to and for aviation noise analysis.

The noise pollution level (NPL) or L_{NP} is related to L_{eq} and was developed by the U.S. Department of Housing and Urban development (HUD) as a guideline for all types of community and environmental noise sources in USA. L_{NP} is based on the evidence that human annoyance for a specific noise exposure is a function of the average level, as well as the variability of the sound source (assuming that the less steady it is, the more distracting and annoying it becomes). L_{NP} combines the ambient noise level with the fluctuations of the noise over time leading to the definition:

$$L_{NP} = L_{eq} + k\sigma \qquad (13.5)$$

where L_{eq} is measured in dB (*A*), σ is the standard deviation of the instantaneous *A*-weighted SPLs with time, k is a constant which is assigned a value 2.56, this providing the best fit to data from a number of studies of subjective response to variable noise levels. This measurement system applies to any environment, unlike those specifically concerned with aircraft/road traffic. Table 13.4 gives some guidelines regarding acceptable values of LNP. The variance may be calculated as

$$\sigma^2 = \sum t_j L_j^2 - \left(\sum t_j L_j \right)^2 \qquad (13.6)$$

where t_j are the fractions of time L_j occur during 24 hours.

TABLE 13.4   LNP criteria proposed by HUD

Category	L_{NP}, dB (A)
Clearly acceptable	<62
Normally acceptable	62–74
Normally unacceptable	74–88
Clearly unacceptable	>88

13.3   SOUND POWER FROM SOUND PRESSURE LEVEL MEASUREMENT

For describing noise emission characteristics of a machine, sound power level is preferred to sound pressure level measurements, since the latter is dependent on the distance between the source and the observer, as well as the environment in which measurements are made. Sound power level may be measured in accordance with international standards ISO 3740 to 3746 (ISO 3742 has been withdrawn). Sound power levels have the following advantages:

 (i) They can be used to calculate sound pressure level at a given distance from the source in a specified environment.
 (ii) They can be used to compare noise radiated by different sources of the same type.
(iii) They can be used in planning to determine the amount of TL.
(iv) They can be used in developing quieter machinery and equipment.

13.3.1   Sound Power Measurement: ISO 3745 and 3746

The strength of a noise source is completely described by two quantities: the sound power level and the directivity. The sound power level is an indication of the total sound power radiated by a source in all directions. It is measured as an *A*-weighted quantity or in 1/1 or 1/3 octave bands like sound pressure level. Directivity gives an indication of the radiation with direction around the source and is also measured in each frequency band. For machines with pronounced directivity patterns such as a jet engine, considerable amount of data is required. Some sources are however, non-directional, radiating uniformly in all directions (omnidirectional). These sources are small compared to the wavelength of the sound they radiate. The International Standards Organization has published six standards 3741–3746 outlining different methods of sound power determination. The first three deal with measurements in a diffuse field, while the last three deal with measurements in a free field environment. The most accurate sound power level values can be obtained by making measurements in anechoic chambers with the source suspended from the ceiling at the centre of the room. However, these chambers are expensive and are not always available to a practising engineer; besides some sources are too heavy to be suspended, or are associated with a reflecting plane. In such cases, measurements can be carried out in a semi-anechoic chamber, where the equipment is placed on a hard reflecting surface with absorption coefficient less than 0.06. Alternately, sound power levels can be determined accurately in a flat outdoor area or in a laboratory room which provides a free field over the reflecting plane.

Sound power in the free field can be found by adding the products of the areas times the acoustic intensities for the areas on any hypothetical surface containing the source. Obtaining sound power level from sound pressure level is described in ISO 3746 and requires measurements of *A*-weighted sound pressure levels at four or more positions located on a hypothetical measurement surface of an area

completely enveloping the source or, if there is a reflecting plane, these positions should lie on the surface enveloping the source and terminating on the reflecting plane. The location of the source under test, the measurement surfaces and the microphone positions are defined by a coordinate system. In order to obtain the average value of the mean square pressure on the measurement surface, an array of fixed microphone positions should be used, the locations of which may be defined by a hypothetical reference box. One of the following two shapes can be used for the measurement surfaces:

- A hemispherical surface of radius r : For sources usually mounted in and/or to be measured in large open areas under satisfactory acoustical conditions, a large measurement distance is selected and in this case the hemispherical measurement surface is preferred.
- A rectangular parallelepiped, the sides of which are parallel to those of the reference box: For sources usually mounted and/or to be measured in rooms or spaces under unfavourable acoustical conditions (for example, when there are many reflecting objects or a high level of background noise), small measurement distances are selected and usually this dictates the selection of a parallelepiped measurement surface.

ISO 3745 suggests three different methods of positioning the microphone on equal areas on the surface of a hemisphere centred at the acoustic centre (in practice the geometrical centre) of the source. The radius of the hemisphere should be more than twice the major source dimension or four times the average distance of the source from the reflecting plane, whichever is larger, and in any case not less than 1 m.

In the first method, a single microphone is moved from positions 1 to 10, as shown in Fig. 13.6(a) or an array of fixed microphones is used and their outputs are sampled sequentially.

In the second method, a single microphone is scanned at constant speed along horizontal circular paths successively. A minimum of five circular paths should be used, as shown in Fig. 13.6(b), where the annular area of the hemisphere associated with each path is equal. Alternatively, the microphone could be placed at the respective heights, while the sound source is rotated at a constant speed using a turntable.

The third method requires scanning a single microphone along a meridional arc, i.e. the microphone is traversed along a quarter of a circular arc about a horizontal axis to the centre of the source. At least eight microphone traverses at equal azimuthal angles around the source should be carried out, and may be achieved by rotating the source through 45° before each traverse. Sound power level of the source can be calculated for each frequency band from Eq. (13.7).

$$L_W = L_p + 10 \, \log_{10} \left(\frac{2\pi r^2}{S_0} \right) - K - B \tag{13.7}$$

where L_W is the band sound power level in dB, L_p is the mean surface sound pressure level over the test hemisphere in dB, r is the radius of the hemisphere, S_0 is the reference area ($1\mathrm{m}^2$) K is the band environmental correction factor in dB and B is the pressure and temperature correction term used when the atmospheric temperature and pressure differ significantly from 20 °C and 10^5 Pa, respectively. The factor K may be taken as 0.5 dB.

The pressure and temperature correction terms may be calculated as per Eq. (13.8) for air.

$$B = 10 \, \log_{10} \left\{ \frac{p_x}{10^5} \right\} + 5 \, \log_{10} \left\{ \frac{306}{273 + t_x} \right\} \tag{13.8}$$
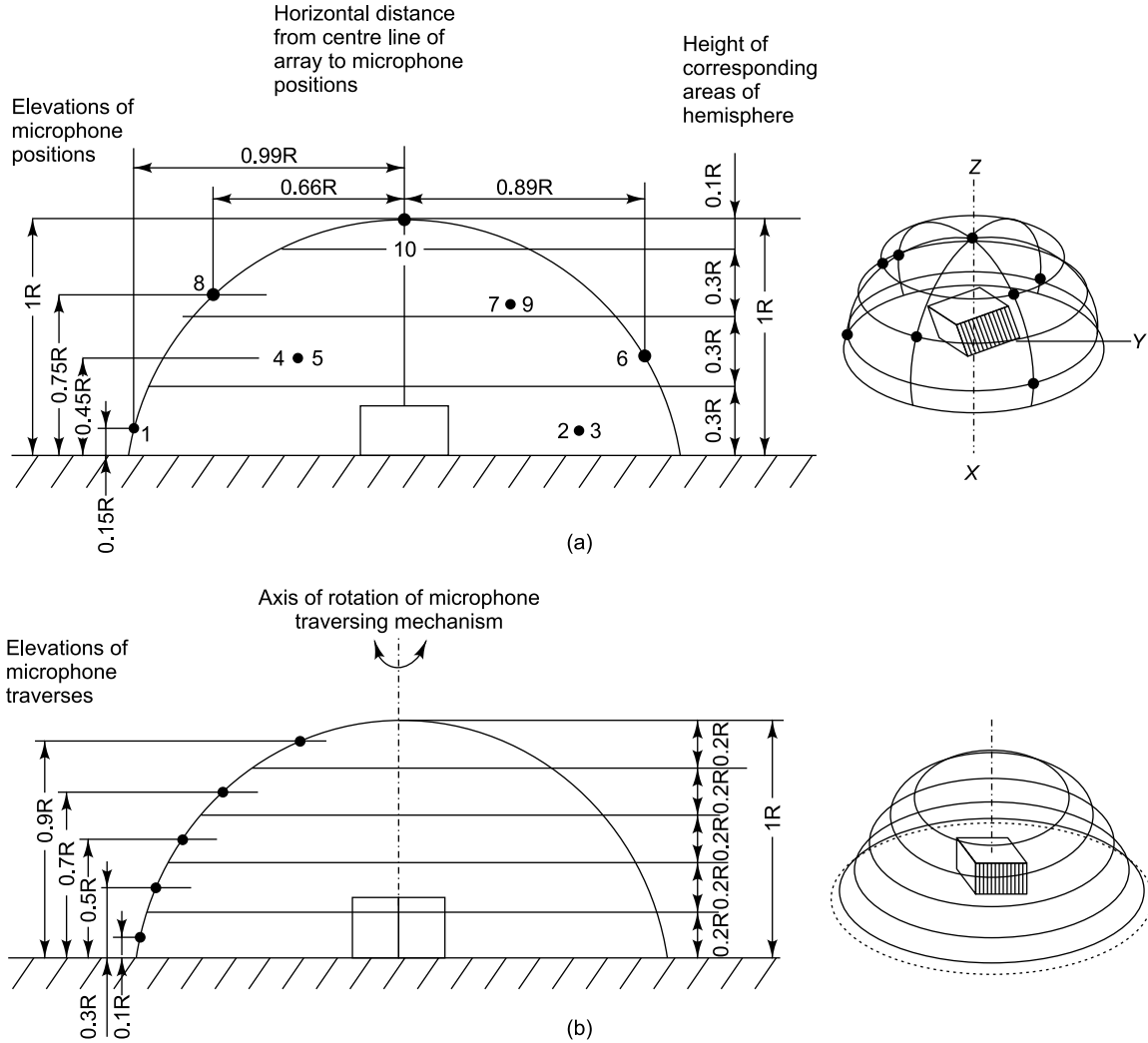
FIGURE 13.6 (a) Microphone positions on equal areas on the surface of a hemisphere, (b) positions of scanning microphone

Here p_x and t_x are the pressure and temperature at the location in Pa and °C, respectively. *B* works out to 0.06. For normal conditions, *B* may be taken as zero for air.

Associated with each microphone position is a surface area, S_i, over which the average sound pressure level is assumed to be equal to that measured at the microphone. The sound power level in each frequency band of interest can be calculated from the following equation.

$$L_W = 10 \log_{10}\left[\sum_{i=1}^{N}\left(10^{(0.1L_{pi})} \times S_i\right)\right] - K - B \tag{13.9}$$

where L_{pi} is the band sound pressure level at the *i*th measuring position in dB and S_i is the surface area associated with the *i*th measuring position in m².

The overall sound power level can be calculated from the 1/1 or 1/3 octave band sound power levels for each band as per the expression

$$L_{W,\text{overall}} = 10 \log_{10} \sum_j 10^{0.1 L_{W,f_j}}$$

(13.10)

where L_{W,f_j} is the sound power level in the jth 1/1 or 1/3 octave band. It is possible to determine the directivity index (DI) when the mean surface sound pressure level has been determined. For measurements over a hemisphere, the directivity index in dB in the ith direction can be calculated from

$$DI = L_{pi} - L_p + 3$$

(13.11)

13.3.2 Sound Power Measurement: ISO 3744 Engineering Method

Sound power levels can be determined with engineering accuracy according to ISO 3744 in a flat outdoor area or in a laboratory room which provides a free field over the reflecting floor. Measurements on larger machines (having a largest dimension of 15 m) can be carried out at site where contributions of the reverberant field to the sound pressures on the measurement surface are small compared with those of the direct field of the source. Since the test objects may be of various sizes and shapes, the measurement surface may be a hemisphere, a rectangular parallelepiped or a conformal surface (Fig. 13.7), which is the same as a rectangular parallelepiped, except that the corners are round and formed by portions of cylinders and spheres. Though the conformal surface is expected to give more accurate results, more effort is required for positioning the microphone and therefore use of the hemispherical surface is generally preferred. The background sound pressure levels should ideally be at least 10 dB below the noise level of the source. Also, the ratio of the room absorption a (defined in Eqs 11.160 and 11.164) to the measurement surface area should be greater than or equal to 6; the higher the value of this ratio, the better. The total absorption of the room (m^3) can be determined in each octave band using the formula

$$a = 0.16(V/T)$$

(13.12)

where V is the volume of the room in m^3 and T is RT of the room in seconds.
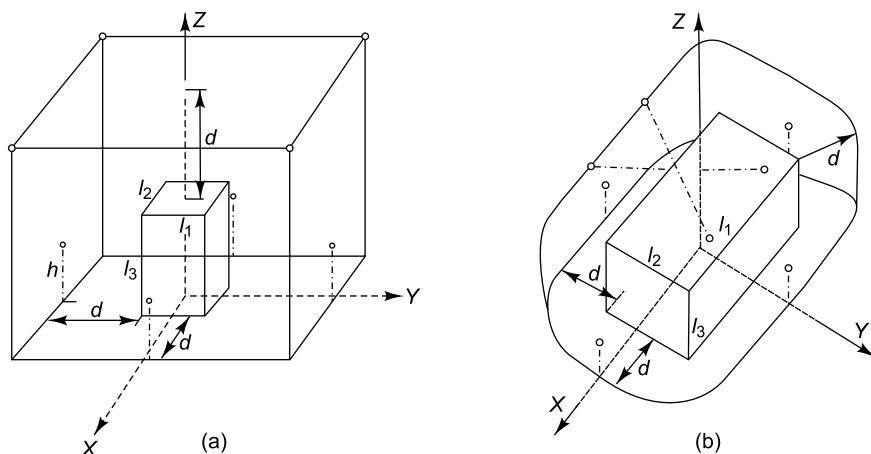
FIGURE 13.7 Microphone locations: (a) on parallelepiped, (b) on conformal surface

13.3.3   Sound Power Measurement: ISO 3741–3743 Diffuse Field Environment

The diffuse field is another type of well-defined environment for evaluation of sound power and can be obtained in a highly reverberant enclosure, which is relatively cheaper to construct than an anechoic chamber. The accuracy of this method depends on the diffuseness of the field and this is not always easy to achieve, especially at very low frequencies and when the source radiates pure tones. The lowest frequency of interest decides the minimum volume of the room. A volume of 200 m^3 is required for the 125 Hz octave band. A volume larger than that required may bring about non-uniformity of the reverberant fields in the higher frequency bands due to air absorption. For very precise measurements, the guidelines given in ISO 3741 may be followed. When very high precision is not required and when cost and labour involved in measurements are of prime concern, the method described in ISO 3743 may be followed. This method can be used for testing small machines in a special test room with a minimum room volume of 70 m^3, the maximum size of the source being 0.7 m^3.

A room with hard reflecting surfaces such as concrete walls may be used. The prescribed value of RT for such rooms is met at higher frequencies, though it is usually high in the low and mid-frequency range. For reduction of RT at low frequencies, membrane absorbers may be used and for middle and high frequencies, perforated panels with mineral wool interiors may be used. Initially the RT of the room should be measured in 1/3 octave bands and should be normalized with respect to the RT at 1000 Hz. If the normalized curve falls within the prescribed limiting curves, the enclosure can be used; else, it should be adapted to get the required changes.

The sound source should be placed near a corner of the room. A microphone should be made to scan a path of at least 3 m at constant speed. Alternatively, discrete averaging with a fixed number of microphones may be done to yield more accurate results at the expense of cost. In a diffuse field, the steady state sound energy is equal to the sound energy transmitted by the source minus that absorbed by the room boundaries. Therefore, L_W can be calculated as

$$L_W = L_p - 10\log_{10}\left(\frac{T}{T_0}\right) + 10\log_{10}\left(\frac{V}{V_0}\right) + 10\log_{10}\left(1 + \frac{S\lambda}{8V}\right) + 10\log_{10}\left(\frac{B}{1000}\right) - 14 \text{ dB} \qquad (13.13)$$

where L_p is the mean band sound pressure level, T is RT in seconds, V is the volume of room in m^3, $V_0 = 1$ m^3, S is the total surface area of the room in m^2, λ is the wavelength at the centre frequency of the band in m and B is the barometric pressure in millibar.

13.4   Sound Intensity Measurement

Intensity measurement is a relatively new technique and helps locate and identify noise sources or sinks in a given complex environment. Generally, a good correlation exists between the sound radiated by a source and its vibrational velocity. Therefore the various mechanisms of sound generated by the source are accounted for in intensity measurements. Intensity measurements can be used for the determination of sound power of noise sources, determination of sound absorption coefficient and TL of structures in addition to noise source location. Sound intensity is defined as the time averaged rate of flow of energy through unit area of the medium as has been discussed in Chapter 11. Being a vector, it gives the direction of acoustical energy flow from the source to the sinks (or absorbers) and can be measured using a specially made probe, which incorporates a pair of highly phase sensitive, phase-balanced condenser

microphones, such that the gradient in sound pressure (driving force of acoustical energy flow) and mean sound pressure level at any space point can be simultaneously taken into account to compute the intensity. The advantages of using intensity measurement are the following.

(i) This technique excludes any influence from steady background noise, provided there is no absorption within the enclosing surface.
(ii) No anechoic or reverberation test chamber is needed.
(iii) Sound measurements can be carried in the near field or far field.
(iv) There is no restriction on the shape of the surface on which measurements are made.
(v) Sound power can be directly calculated from the measured intensity.

13.4.1 Sound Intensity Measurement Using the Two Microphone Method

Sound intensity is the time averaged product of the instantaneous pressure $p(t)$ and the corresponding particle velocity $u(t)$ in a given direction at a point in the sound field.

$$I = \frac{1}{T}\int_T p(t)u(t)dt \tag{13.14}$$

The pressure may be computed as the average value of pressures p_1 and p_2 obtained from two identical closely placed microphones A and B.

$$p = \frac{p_1 + p_2}{2} \tag{13.15}$$

The particle velocity is related to pressure gradient according to the linearized Euler equation (Eq. 11.4) as

$$\rho_0 \partial u / \partial t = -\text{grad } p \tag{13.16}$$

where ρ_0 is the mean mass density of air, p is the instantaneous pressure and u is the corresponding instantaneous particle velocity in the direction of propagation.

The particle velocity can be computed by integrating the pressure gradient signal, which is measured by obtaining the pressures p_1 and p_2 at locations A and B and dividing the pressure difference by the microphone separation distance Δr. This computation is valid as long as the distance between the microphones is very small as compared to the wavelength of the sound being measured. The particle velocity can be written as

$$\frac{\partial u}{\partial t} = -\frac{1}{\rho_0}\frac{p_2 - p_1}{\Delta r} \tag{13.17}$$

$$u \approx -\frac{1}{\rho_0 \Delta r}\int (p_2 - p_1)dt \tag{13.18}$$

where u is the particle velocity in the r direction. The sound intensity in the r direction is, then

$$I_r = \langle pu \rangle \approx \left\langle -\frac{(p_1 + p_2)}{2\rho_0 \Delta r}\int (p_2 - p_1)dt \right\rangle \tag{13.19}$$

where $\langle \ \rangle$ indicates expectation or time averaging. The principle of sound intensity measurement is as shown in Fig. 13.8.
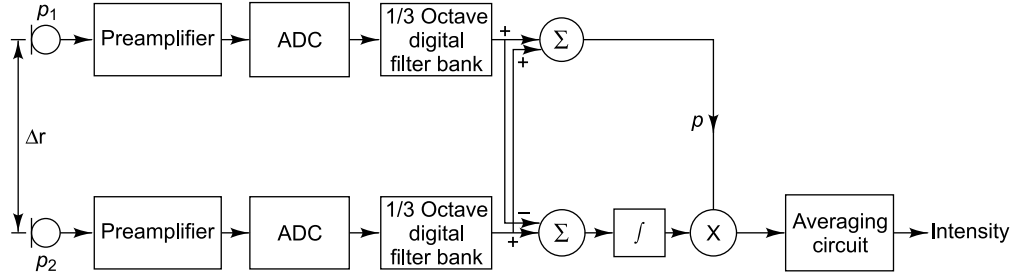
FIGURE 13.8   Block diagram showing sound intensity measurement

13.4.2   Sound Intensity Measurement using FFT Analyzer

Another approach to measurement of sound intensity is through the use of an FFT analyzer. In this method, the cross-spectrum between the two closely spaced microphones is obtained with the help of a dual-channel FFT analyzer and the intensity calculated from it. The theory behind this is as follows.

In Eq. (13.14), p and u are in general complex quantities and are not in phase. The Fourier transform of Eq. (13.15) is

$$P = \frac{P_1 + P_2}{2} \tag{13.20}$$

The Fourier transform of Eq. (13.18) is

$$U = \frac{i}{\omega \rho_0 \Delta r}\left(P_2 - P_1\right) \tag{13.21}$$

The expression for intensity in the frequency domain is

$$I(\omega) = \frac{1}{2}\operatorname{Re}\left\{P(\omega)U^*(\omega)\right\}$$

$$= \frac{1}{2}\operatorname{Re}\left\{\left(\frac{P_1 + P_2}{2}\right)\left[\frac{i}{\omega \rho_0 \Delta r}\left(P_2 - P_1\right)\right]^*\right\} \tag{13.22}$$

This can be rearranged as

$$I(\omega) = \frac{1}{2}\operatorname{Re}\left\{\frac{-i}{2\omega \rho_0 \Delta r}\left(P_2 P_2^* + P_1 P_2^* - P_2 P_1^* - P_1 P_1^*\right)\right\}\mathfrak{M} \tag{13.23}$$

$P_1 P_1^*$ and $P_2 P_2^*$ are auto-spectra which are real quantities and which do not contribute to the real part of the total expression or the active sound energy. Therefore,

$$I(\omega) = \frac{1}{2}\operatorname{Re}\left\{\frac{-i}{2\omega \rho_0 \Delta r}\left(P_1 P_2^* - P_2 P_1^*\right)\right\} \tag{13.24}$$

But

$$P_1 P_2^* = (P_2 P_1^*)^* \tag{13.25}$$

Hence

$$P_1 P_2^* - P_2 P_1^* = 2i \operatorname{Im}\{P_1 P_2^*\} \tag{13.26}$$

Plugging Eq. (13.26) into Eq. (13.24),

$$I(\omega) = \frac{1}{2} \operatorname{Re}\left\{ \frac{-i}{2\omega\rho_0\Delta r} \cdot 2i \operatorname{Im}\{P_1 P_2^*\} \right\} = \frac{\operatorname{Im}\{P_1 P_2^*\}}{2\omega\rho_0\Delta r} = \frac{\operatorname{Im}\{P_{12}\}}{\omega\rho_0\Delta r} \tag{13.27}$$

Here $I(\omega)$ is the acoustic intensity component in the direction r. $\operatorname{Im}\{P_{12}\}$ is the imaginary part of the cross-spectrum between the two microphone signals and can be directly measured using a dual-channel FFT analyzer. Since cross-spectral density P_{12} is needed for the solution, a phase calibration is necessary. For this, the microphones are flush-mounted on a rigid circular plate attached to the end of the tube. Ideally, in this configuration both microphones would measure the same pressure amplitude with zero phase shift.

13.4.3   Instrumentation for Measurement of Sound Intensity

The sound intensity measurement system consists of an intensity probe, a real-time analyser and a post-processing system. The sound intensity probe has two microphones with a solid cylindrical spacer separating them as shown in Figs 13.9(a) and (b). The microphones should have identical phase response and a flat amplitude response as a function of frequency. Microphones of different sizes and different spacer lengths are used in these probes for different frequency ranges. Table 13.5 gives the commercially available sizes of microphones, spacers and the frequency ranges of operation.
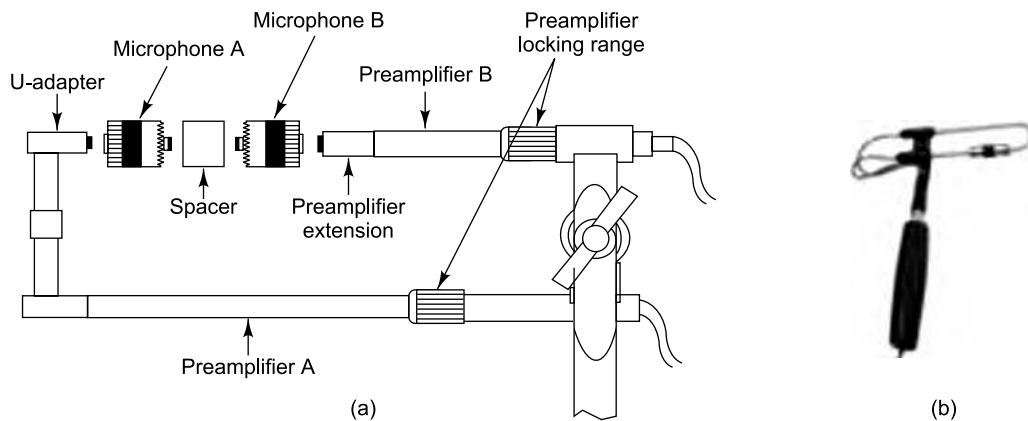
FIGURE 13.9   Sound intensity probe: (a) schematic, (b) photograph (Courtesy of Brüel & Kjær, Denmark)

TABLE 13.5   Microphone sizes, spacer sizes and frequency range of operation

S. No.	Microphones	Spacer size	Frequency range (Hz)
1.	1/2 inch pair	12 mm	100–500
2.	1/2 inch pair	50 mm	50–1250
3.	1/4 inch pair	6 mm	200–10,000
4.	1/4 inch pair	12 mm	200–5000

13.4.4   Methods of Sound Intensity Measurements

The two methods of measurements commonly used are the sweep or scan method and the point method.

13.4.4.1   Sweep or scan method   In this method, the intensity probe is slowly swept over the hypothetical surface enclosing the noise source, while the analyzer is made to continuously average the measurements. Figure 13.10 shows the scanning patterns which can be employed. These different patterns give different errors in the sound power obtained. This error can be quantified as ΔL_W and is the difference between the reference power source and the power measured by the intensity probe. This method is much faster and more accurate than the point method described subsequently. The errors obtained with the various scanning patterns employed and depicted in Fig. 13.10 are given by ΔL_W: for (a) 0.9, (b) 0.2, (c) 1.3, (d) 0.9 and (e) 0.7 dB.
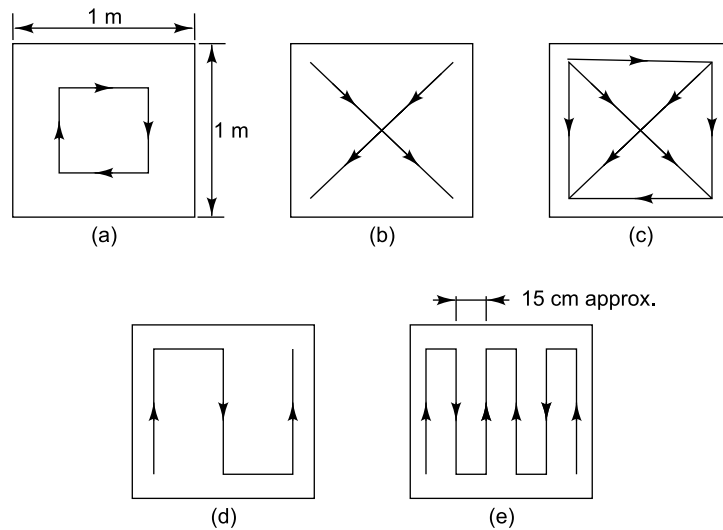
FIGURE 13.10   Commonly used scanning patterns with errors: (a) 0.9 dB, (b) 0.2 dB, (c) 1.3 dB, (d) 0.9, (e) 0.7 dB

13.4.4.2   Point method   In this method, the intensity is computed based on measurements made at fixed points (usually forming a grid). These measurements can be used for obtaining intensity maps for noise source identification. For ΔL_W less than 1 dB, approximately 0.09 m^2 of area per point is required. For good results 0.04 m^2 of area per point should be used.

13.4.4.3   Errors   There are some errors associated with the sound intensity measurement technique and they are due to phase mismatch, finite difference approximation, near field and statistical errors. These errors impose restrictions on the measurement frequency range, both the high frequency and low frequency limits. The low frequency limit is decided by the amount of phase mismatch between the two microphones and the channels of the FFT analyser and can be minimized by the following precautions:
* Precise matching of the p and u channels, i.e. by choosing a pair of microphones with a low phase mismatch (below 0.5° from 100 to 10,000 Hz).
* Adjusting Δr accurately.

The high frequency limit in intensity measurements is due to the approximation of the pressure gradient by a finite pressure difference and this is called bias error. The orientation of the probe θ

with respect to the source direction also gives rise to errors in measurement. Figure 13.11 shows the probe orientation θ.

The sum and difference of p_1 and p_2 are good approximations for p and grad p, respectively. if $\Delta r \ll \lambda$, where λ is the wavelength. If Δr remains constant (for a selected spacer), the error decreases with decrease in θ. Therefore, by choosing a spacer and probe orientation of 0°, error in intensity level for a given frequency can be determined.
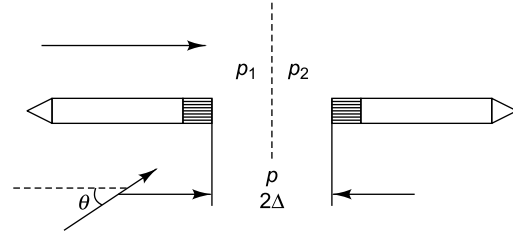
FIGURE 13.11   Probe orientation θ

13.5   REVERBERATION TIME MEASUREMENTS

RT is a quantity of great significance in architectural acoustics and is obtained from spatial acoustic analysis for characterising the acoustic properties of a room or enclosure. RT is used to describe room acoustics by analysing the decay of sound intensity and is the time it takes for the sound pressure level in an enclosure to decay by 60 dB or to one millionth of the original energy after a sound source has been switched off. This time depends on the absorbing properties of the room surfaces; larger the amount of absorbing material, smaller is the RT. In practice, all surfaces absorb some amount of sound energy, leading to a finite RT. Values of RT may range from fractions of a second to a few seconds, depending upon the size of the room and the nature of the materials used in its construction. Highly reflective materials, such as a concrete or tile floor, brick walls, and windows, will increase the RT as they are very rigid. Absorptive materials, such as curtains, heavy carpet and people, reduce the RT. In fact, people tend to absorb quite a bit of energy, reducing the RT. Bigger rooms have longer RTs since the sound waves have to travel a longer distance between reflections. The air in the room itself will also attenuate the sound waves, reducing the RT. Table 13.6 gives an indication of the RTs of some commonly used spaces.

While designing an enclosure, it is imperative to identify appropriate values of RT depending on its intended use and then to decide on the required building materials to be used in the construction. In the case of rooms with acoustic problems in existing rooms, the measurement of RT is usually the first step in identifying a suitable remedy and depending on the application, RT may need to be corrected. A small RT is desirable to minimize masking effects of sound, while a large RT is desirable if a weak source of sound is to be audible everywhere; hence a compromise between the two extremes is desirable.

TABLE 13.6   Reverberation times of some common enclosed spaces

Enclosure	Reverberation time (s)
Recording studio	0.4–0.5
Living room	0.5
Lecture hall	0.5–1
Opera house	1.1–1.3
Small music hall	1.5
Concert hall	1.5–2.1

To measure the RT, a sound source and instrumentation for capturing the sound decay are required. A diffuse field environment is to be created for the measurement. The experimental procedure involves generating a sound using a source and capturing the decay after the source is switched off. One primary

difficulty encountered with a pure tone sound source is the formation of a pattern of standing waves, especially pronounced at low frequencies. Hence, a method commonly employed to average out the anomalies is to use a warble oscillator, in which by slowly warbling (adopting frequency modulation of) the frequency of the sound emitted by a loudspeaker, continuous changes in the standing wave patterns are produced. There are two other options for the sound source: impulse excitation or noise excitation (ISO140 series of standards). It is important for the noise to have sufficient bandwidth to cover the entire frequency range of interest. The source is positioned in a corner of the room to ensure that it undergoes as many reflections as possible. It is suddenly switched off after steady-state conditions have been reached. RT is measured by collecting response data throughout the space using calibrated microphones. These signals are then digitized using an analogue-to-digital converter and then sent to a personal computer with software for mathematical data analysis. Fast sampling is necessary to ensure that the signal reconstructed from the samples is an accurate representation of the analogue input signal. The record length should be long enough to ensure that the decay event is fully captured. The earlier practice was to amplify, rectify and record the level of the signal by a stylus on a moving strip of paper. The rate of decay in dB/s could be obtained from the slope of the curve and from this RT was calculated.

Although the RT is defined as the time it takes for the sound to decay 60 dB, this is often difficult to measure in practice, the reason being the high existing background noise. To be able to measure a 60 dB decay with a background noise level of 55 dB, the acoustical output level of the noise source should typically be more than 120 dB (55 + 60 + another 5 dB), which is a requirement very hard to meet with many practical sources. Therefore it has become a practice to measure 15, 20, 30 or 40 dB decay instead of the complete decay and subsequently to extrapolate it to 60 dB, assuming that the measured decay is representative of the entire one. Hence the decay used, i.e. T20 or T30 or T40, is specified and a check for consistency is made by comparing T20 and T40 records to ascertain if the decay is linear (when plotted as a graph with a logarithmic level scale). Most instruments currently available for building acoustics have at least two ways of calculating the decay simultaneously. The frequency range of interest is 100–3150 Hz, but many measurements are made in 50–5000 Hz range, this being the requirement for laboratory measurements. RT measurements are made in 1/3 octave bands in this frequency range since measurements made at a single frequency are insufficient to completely characterize the reverberation characteristics of an enclosure due to the frequency dependence of the absorption coefficients of materials.

An ideal sound decay will form a straight line when drawn with logarithmic axes. However, actual sound decays always contain fluctuations, the influence of which may be reduced by a triangular weighting function applied to the middle part of the decay curve and ignoring the fluctuations at the beginning and end. In order to account for any possible influence from the background noise level, a minimum distance to the noise floor may be specified (ISO 354 requires a minimum of 15 dB distance). Figure 13.12 shows a typical decay curve.
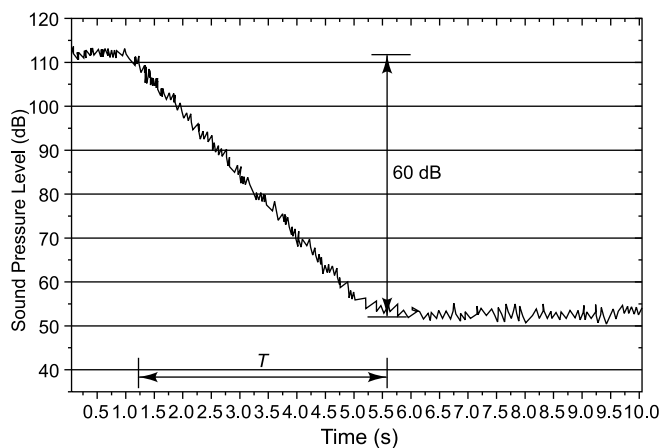
FIGURE 13.12   Typical decay curve

According to ASTM C423, the decay rate is to be determined as follows:

1. Once the noise source is turned off, sound pressure levels in all one-third octave bands of interest are to be sampled every 50 ms or less. An arithmetically averaged decay curve is to be obtained over N or more decay curves, where N should be at least 50.

$$\langle L_m \rangle = \frac{1}{N} \sum_{n=1}^{N} L_{mn} \tag{13.28}$$

where $\langle L_m \rangle$ is the average of the sound pressure levels measured at the mth data point across the N decays and L_{mn} is the sound pressure level measured at the mth data point during the nth decay.

2. In each band, the following are to be found: the first data point that occurs 100 ms or more after the noise source is turned off and the last data point that is within 25 dB of the earlier value.

3. In each band, the unadjusted decay rate d' is to be calculated; this is the negative of the slope of the linear, first-order regression on the averaged decay curve computed as:

$$d' = \frac{6}{M(M^2 - 1)\Delta t} \left[(M+1) \sum_{m=1}^{M} \langle L_m \rangle - 2 \sum_{m=1}^{M} \langle L_m \rangle \right] \tag{13.29a}$$

where M is the number of data points in the jth band and Δt is the time interval with which the sound pressure level spectra were sampled.

The unadjusted decay rate may be optionally adjusted for air absorption as

$$d = d' - d_{\text{air}} \tag{13.29b}$$

where the correction is based on coefficients taken from ANSI S1.26 for the appropriate temperature and relative humidity.

13.6 Sound Absorption Measurement

Sound absorbing materials play a very important role in architectural acoustics, the design of sound recording and broadcast studios, automobile interiors, cinema theatres, factories, workshops, firing ranges, etc. They are used to reduce reflected sound or reverberations and hence the steady-state noise levels in an enclosure, reduce or control reverberation and improve listening environments for speech and music. Materials with high absorption coefficients are usually soft and porous and sound absorption of a material is related to its physical properties, such as air flow resistance, porosity, elasticity and density. Examples of common sound absorbing materials are acoustic ceiling tiles and soft furnishings and curtains. With today's growing focus on noise control issues, acoustic material testing is becoming increasingly important. The absorbing properties of the materials covering the surfaces of a room are required in the computation of its RT.

When sound waves strike such surfaces, air flows in and out of the holes. Frictional forces convert the sound energy into heat, though the actual amount of this energy is small. When a sound wave impinges on a surface, part of the energy is absorbed and part of it is reflected. The sound absorbing characteristic of a material is defined in terms of its absorption coefficient, which is the ratio of the energy absorbed by the surface to the energy incident on it. When all the incident sound energy is reflected, the value is zero, and it is one when all the energy is absorbed. A material like foam insulation has an absorption coefficient close to 1, while marble which is a very good reflector has an absorption coefficient very close to 0.

Absorption coefficient also varies with the frequency and angle of incidence of the sound wave. Normal incidence sound absorption coefficients can be quite useful in certain situations where the material is placed within a small acoustical cavity close to a sound source, for example, a closely fitted machine enclosure. Generally, higher frequencies are more easily absorbed than the low frequencies found in human speech. Therefore, when specifying the absorbing qualities of a material, manufacturers depict it as a function of frequency in 1/1 or 1/3 octave bands. The selection of a sound absorbing material has to be based on its absorbing characteristics for a particular noise source. Materials that are good absorbers permit sound to pass through them relatively easily. When sound waves are reflected from the surfaces of a room, they interact with the sound absorbing materials and lose some energy. However, they require large thickness or many paths for the sound energy to be significantly reduced. This is why sound absorbers are generally not good sound barriers; the latter are essentially materials that prevent the passage of sound and are usually solid, fairly heavy and non-porous. The absorption coefficients of different materials used for the construction of buildings is shown in Table 13.7. Absorption coefficients are used to help predict or design a space with a desired reverberation response.

TABLE 13.7 Absorption coefficient of construction materials

Material	125 Hz	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Bricks, unglazed	0.01	0.01	0.01	0.01	0.02	0.02
Marble	0.03	0.03	0.03	0.03	0.03	0.03
Heavy carpet on felt	0.08	0.27	0.39	0.34	0.48	0.63
Asphalt tile	0.02	0.03	0.03	0.03	0.03	0.02
Bricks, unglazed, painted	0.01	0.01	0.02	0.02	0.02	0.03
Concrete block	0.36	0.44	0.31	0.29	0.39	0.25
Concrete block, painted	0.10	0.05	0.06	0.07	0.09	0.08
Concrete	0.01	0.05	0.06	0.07	0.09	0.08
Wood	0.15	0.11	0.10	0.07	0.06	0.07
Glass, Ordinary Window	0.35	0.25	0.18	0.12	0.07	0.07
Plaster	0.013	0.015	0.02	0.03	0.04	0.05
Plywood	0.28	0.22	0.17	0.09	0.10	0.11
Fibre glass	0.30	0.34	0.50	0.80	0.93	0.96
Plate glass	0.18	0.06	0.04	0.03	0.02	0.02

There are two methods of measuring sound absorption: one involves the use of an impedance tube and is suitable for small samples, while the other involves the use of a reverberation chamber and is suitable for large objects, furniture, panels, etc. The simplest way to measure the absorption coefficients and acoustic impedances of samples is by using a Standing Wave Apparatus. This apparatus was developed essentially to measure the sound absorbing properties of building materials and absorptive coatings for submarines (to protect them from sonar detection) during World War II. The standing wave tube method allows one to make quick and easy, yet perfectly reproducible, measurements of absorption coefficients and is based on measurement of the Standing Wave Ratio (SWR). It also allows for accurate measurement of the normally incident acoustic impedance, and requires only small samples of the absorbing material. The small size and durable aluminum construction make it easy to transport and it can be used for estimating the properties of walls, ceilings, installed building materials, road surfaces, different ground surfaces or interiors of vehicles.

There are two standards for the measurement of acoustical properties in an impedance tube: ASTM E1050-98 and ISO 10534-2. Both describe well-defined acoustical conditions and special instrumentation to ensure a high and known degree of repeatability and reproducibility. Both describe what has come to be known as the 'two-microphone' or 'transfer-function' method of measuring absorption and impedance of acoustical materials. This test method allows one to compare relative values of sound absorption when it is impractical to procure large samples for accurate random-incidence measurements in a reverberation room.

An alternate method for determining the absorbing properties of a material, also in wide use today, involves the use of a special reverberation chamber. A large piece of the sample (say 10 m²) is placed in the chamber. The difference in the RT with and without the material yields the absorbing properties of the material. This method is generally more expensive, since it requires a reverberation chamber and precisely calibrated sensors and is more cumbersome than the standing wave method. Besides it does not yield normally incident acoustic impedance data. However, it is superior for determination of absorbing properties that depend on the size of the material and for measuring absorption characteristics for randomly incident sound waves since the standing wave tube measures the absorption coefficient only for normal incidence and the measured coefficients are generally somewhat smaller than those measured using the reverberation room method.

13.6.1 Measurement of Reflection and Absorption Coefficients and Impedance Based on Standing Wave Ratio Method

The standing wave tube or Kundt tube is shown in Fig. 13.13. At one end of the tube is a loudspeaker which produces an acoustic wave that travels down the tube, impinges with normal incidence on the test sample placed at the other end and gets reflected back. The amplitude of the reflected wave depends on the incident wave and the characteristic impedance of the test sample. Since this method is based on plane wave propagation, there is a maximum frequency which a tube can support and this is $f_{max} = c_0/1.7D$, where D is in metres.
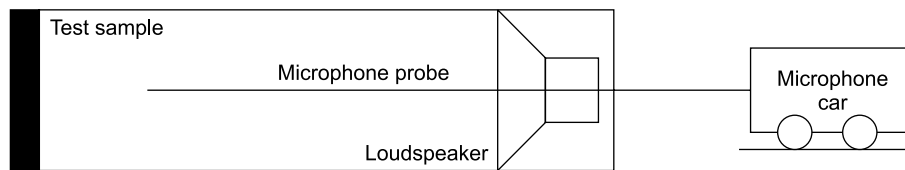
FIGURE 13.13   Standing wave tube

Both the forward and the backward travelling waves combine and there will be a resulting phase interference between the waves, leading to the formation of a standing wave pattern in the tube. At some particular points in space called points of destructive interference (nodes), the two waves will always be out of phase; the result is that they will cancel each other at these points. At points of constructive interference (antinodes), the two waves will always be in phase. If 100% of the incident wave is reflected, then the incident and reflected waves have the same amplitude, resulting in the nodes in the tube having zero pressure and the antinodes having double the pressure. If some of the incident sound energy is absorbed by the sample, then the incident and reflected waves have different amplitudes; the nodes in the tube no longer have zero pressure. When there is a standing wave in an impedance tube, if a microphone

is moved along a line between the sound source and the reflecting surface, a series of maxima and minima in sound level is found. The pressure amplitudes at nodes and antinodes are measured with a travelling microphone attached to a car which slides along a graduated ruler.

Figure 13.14 shows the pressure amplitude in a standing wave tube. Figure 13.14(a) shows the pressure amplitude in the tube with a rigid termination at $x = L$. All the sound energy incident on the termination is reflected with the same amplitude. There is a small reduction in amplitude due to absorption along the walls as the waves travel back and forth along the tube. Figure 13.14(b) shows the standing wave when the tube is terminated at $x = L$ with some sound absorbing material. Here it is seen that some of the incident acoustic energy is absorbed by the material, resulting in a decrease in the amplitude of the reflected waves. Besides the impedance mismatch at the absorbing material introduces a phase shift upon reflection.

From the plane wave assumption, the sound field inside the waveguide is described by

$$p = p_i + p_r \qquad (13.30)$$
$$= Ae^{i(\omega t - kx)} + Be^{i(\omega t + kx)}$$

where the first term represents the incident wave and the second term the reflected wave, $k = \omega/c_0$ is the wave number, A is the complex pressure amplitude of the incident wave, B that of the reflected wave and ω is the angular frequency.
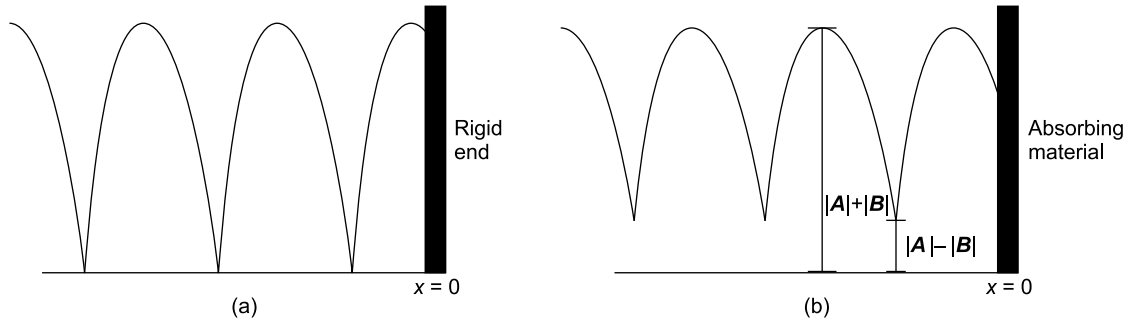
FIGURE 13.14 Standing waves in Kundt tube: (a) with rigid termination, (b) with sound absorbing material

The maximum pressure amplitude at an antinode is $|A| + |B|$, while the minimum amplitude at a node is $|A| - |B|$. It is not possible to measure A or B directly. However, we can measure $|A| + |B|$ and $|A| - |B|$ using the traversing microphone. The ratio of the pressure maximum (antinode) to the pressure minimum (node) is called the SWR S.

$$S = \frac{|A| + |B|}{|A| - |B|} \qquad (13.31)$$

This ratio, which always has a value greater than or equal to unity, is used to determine α_r, the sample's reflection coefficient, α_a, its absorption coefficient, and z, its impedance. Sound power reflection coefficient is defined as

$$\alpha_r = \left|\frac{B}{A}\right|^2 = \frac{(S-1)^2}{(S+1)^2} \qquad (13.32)$$

The sound power absorption coefficient for the test sample is

$$\alpha_a = 1 - \alpha_r = 1 - \frac{(S-1)^2}{(S+1)^2} \tag{13.33}$$

A pressure minimum occurs when

$$\cos\left(kx + \frac{\theta}{2}\right) = 0 \ \text{ and } \ \sin\left(kx + \frac{\theta}{2}\right) = 1 \tag{13.34}$$

which requires that

$$kx + \frac{\theta}{2} = -\frac{\pi}{2} \tag{13.35a}$$

or
$$\theta = -2kx - \pi \tag{13.35b}$$

Here the quantity x equals the distance from the test sample to the first pressure minimum and θ is the phase lag of B with respect to A. The complex mechanical impedance of the test sample is then obtained as

$$z_n = z_0 \frac{A+B}{A-B} \tag{13.36}$$

z_0 is the characteristic impedance of the medium and z_n is the specific acoustic impedance of the test specimen. Since the mechanical impedance of the test sample is generally a function of frequency, it may be required to repeat the above measurements over the frequency range of interest in various 1/1 and 1/3 octave bands. At low frequencies the thickness of the material is important since absorption increases with thickness. Low frequency absorption can be further increased by fixing the material at a distance of one-quarter wavelength from a wall, instead of directly on it.

13.6.2 Two Microphone Technique for Determination of Absorption Coefficient—The Transfer Function Method

The two-microphone transfer function (impedance tube) method is described in BS EN ISO 10534-2:2001 and in ASTM E1050-98. This method is designed for measurement of the following quantities: (i) absorption coefficient, (ii) reflection coefficient, (iii) complex acoustic impedance, (iv) complex acoustic admittance and (v) TL.

This technique can be used to evaluate acoustic properties very rapidly since no traversing microphone is necessary. It is also a much faster measurement technique compared to the measurement of sound absorption in a reverberation room according to the method specified in ISO 354. This method uses an impedance tube, two microphones and a digital frequency analysis system for the determination of the sound absorption coefficient of sound absorbers for normal sound incidence that is, at 0° in the frequency range 150–6000 Hz. The test specimens are small circularly cut samples of the sound absorbing materials. A loudspeaker or a high-output acoustic driver (sound source) is mounted at one end of an impedance tube and a small sample of the material to be tested is placed in a sample holder at the other end. A rigid plunger with an adjustable depth is placed behind the sample to provide a reflecting surface. The sound source is typically made to generate broadband, stationary random sound waves which propagate as

plane waves in the tube, impinge on the sample normally and are reflected back, resulting in a standing wave interference pattern inside the tube.

The sound pressure is measured by a pair of microphones mounted flush with the inner wall of the tube near the sample end and the complex transfer function between the two measurements is calculated using a multichannel spectrum analyser. This method separates the incident and reflected energy from the measured transfer function, and then estimates the required acoustic properties of the test sample. The usable frequency range depends on the diameter of the tube and the spacing between the microphone positions. A 100 mm diameter tube is used to cover measurements over the frequency range 150–1600 Hz, and a 30 mm diameter tube to cover measurements in the frequency range 1200–6000 Hz. Two different diameter tubes may be used to cover the full frequency range.

In the ASTM E1050-98 measurement technique, the microphone closer to the source is taken as the reference channel. Theory has been developed relating the auto and cross-spectral densities of the incident and reflected waves to the auto and cross-spectral densities of the two microphone signals. The origin of the coordinate system is at the termination of the tube (with the sample) and x_1 and x_2 are the distances to the two pressure measurement points. The two-microphone method is shown schematically in Fig. 13.15.
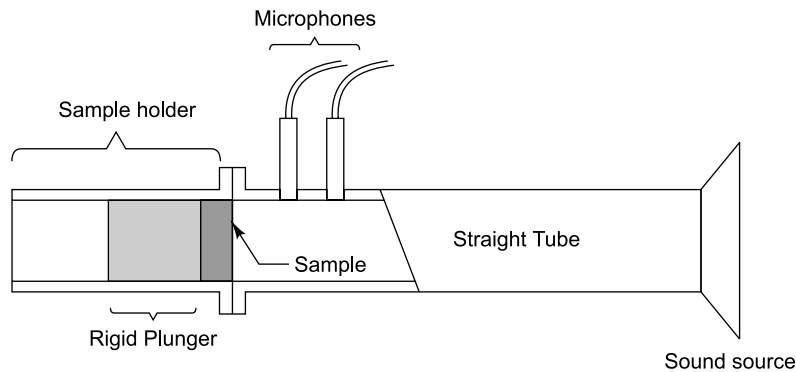
Figure 13.15   Schematic of two-microphone method of sound absorption

An expression for the reflection coefficient is obtained by measuring pressure at two different axial locations. From the transfer function H_{12}, the complex reflection coefficient R of the material is determined as

$$R = \frac{H_{12} - e^{-iks}}{e^{iks} - H_{12}} e^{i2k(L+s)} \qquad (13.37a)$$

In the above expressions H_{12} is the estimated frequency response function between the two microphones and is given by

$$H_{12} = E\left[\frac{S_{12}}{S_{11}}\right] \qquad (13.37b)$$

where $E[\ ]$ is the expectation operator, S_{12} is the estimated cross spectral density and S_{11} is the estimated auto spectral density. L is the distance from the sample face to the first microphone in ASTM E1050-98

and s is the distance between the microphones. From the complex reflection coefficient, the absorption coefficient α_a is determined using Eq. (13.38a) and normal specific acoustic impedance z_n of the sample may be determined from Eq. (13.38b).

$$\alpha_a = 1 - |R|^2 \qquad (13.38a)$$
$$z_n = \rho_0 c_0 (1+R)/(1-R) \qquad (13.38b)$$

A quick absolute test for calibrating this set-up is done by measuring the imaginary part of impedance of a closed tube of length L_0. The theoretical normalized impedance $z/\rho_0 c_0$ of a closed tube of length L_0 for calibration is given by:

$$\text{Im}(z/\rho_0 c) = -\frac{1}{\tan k L_0} \qquad (13.39)$$

Figure 13.16 shows a photograph of the impedance tube.

FIGURE 13.16   Photograph of impedance tube (Courtesy of http://www.bswa-tech.com)

Typical specifications of impedance tube
Frequency range: 125–3150 Hz   1600–6300 Hz
Inner diameter of test tube: 60 mm   30 mm
Loudspeaker: 20 W, 8 Ω, 100 mm diameter
Microphone: ¼ inch diameter

13.6.2.1   *Precautions to be taken during measurement*   *Construction of the tube*: The tube is made preferably of a material with high damping such as brass, which has damping several times that of materials such as aluminium and density also three times as much. It is required that the tube be sufficiently massive and rigid to avoid transmission of noise into the tube from outside and to prevent vibration excitation of the tube by the sound source or by background sources. The tube must be long enough to present a stable plane-wave sound field to the sample under test. The standards recommend a tube length of at least three diameters, but a length of not less than 10–15 diameters is preferred. Typically the upper working frequency f_u is limited by: $d < 0.58\,\lambda_u$ (for a circular tube) or $d < 0.50\,\lambda_u$ (for a rectangular tube) where d is the inside diameter of circular tube (m) or side of rectangular tube. A wall thickness which is 10% of the tube diameter is preferred.

Microphone mounting and frequency range: The microphones must be mounted flush with the inside wall of the tube and isolated from the tube (to minimize sensitivity to vibration). Both the microphones should be of laboratory grade and of the same type and size. Their diameter should be small in comparison with their spacing to reduce the influence of their acoustic centres; $d_{mic} < 0.2s$ where d_{mic} is the diameter of the microphones and s the spacing (m) between them. Their diaphragm diameter should be small to minimize high frequency spatial averaging across the diaphragm face: $d_{mic} << \lambda_{max}$ where $\lambda_{max} = c_0/f_{max}$ corresponds to the wavelength of the highest frequency sound. The upper frequency limit is chosen to ensure the occurrence of plane wave mode propagation and accurate phase detection; typically ½ or ¼ inch microphones are used. The spacing f_{max} between microphones is chosen such that $s < (0.40$ to $0.45)$ λ_{max} where λ_{max} is the maximum side length of the rectangular tube (m) and f_{max} is the highest frequency (Hz). The spacing between the sample and closest microphone must be large enough to avoid proximity distortions to the acoustic field; for a sample with strongly asymmetrical layer, $\Delta r > 2d_{mic}$ is preferred.

Signal-to-noise ratio: The standards recommend that the background noise inside the tube be lower than the sound field, by at least 10 dB, but preferably by 20–30 dB. For this, the TL of the tube should be greater than 45 dB. This can be achieved by ensuring that the tube is constructed of heavy materials and is properly sealed at the microphone ports, sample holder, sound source, etc. The minimum signal-to-noise ratio occurs at the minima of the standing waves which can be as much as 25 dB below the maxima. Besides, it is often required to measure material absorption at relatively low levels to match field conditions (e.g. approximately 65 dB to simulate conditions in offices or vehicles). This poses a requirement of a low background noise level in the tube.

Sample holder: The sample holder should also be massive and rigid and should have a massive back plate of length at least 20 mm behind the sample. It should also have a smooth, machined surface for mating with the tube to ensure that no sound leaks into or out of the tube. The back plate or plunger should be adjustable so that materials of different thickness may be tested and to allow for an air space behind the sample, if necessary.

Sound source: The sound source should have the required frequency range and a high power rating (e.g. 50–100 W) so that high intensity sound may be generated inside the tube for certain types of testing. For certain materials, the absorption depends on the intensity of the sound source and such materials should be tested at several levels above and below field conditions. The sound source should also have a frequency response which is fairly flat (within ±10 dB).

Mounting the sample: The impedance tube method requires samples of the test object which are of the same size as the cross-section of the impedance tube. According to standards, the sample must fit snugly in the specimen holder so that there is no space between its edge and the holder. The test specimen is generally sealed around the edges using Vaseline or layers of tape to minimize leakage of sound. The sample should also be in contact with the rigid plunger. In applications where the acoustic material is used with an air gap, the test should duplicate field conditions by maintaining the same air gap between the sample and the plunger.

Temperature correction: The temperature should be held constant during the test to within 1 °C.

13.6.3 Absorption Coefficient Measurement Using a Reverberation Chamber

The reverberation room method ideally determines the sound absorption coefficient for diffuse sound incidence. This method requires a large (not less than 125 m³, preferably >200 m³) and carefully qualified

reverberation chamber and test specimens which are rather large (more than 5.4 m^2, and with aspect ratio not too high). Hence this technique is not convenient for research and development work, where only a few samples of the absorber are available. Estimates of the random incidence absorption coefficients can be obtained from normal impedance data for both locally reacting and bulk-reacting materials using available formulae. The sound absorption measurement in a reverberation chamber may be carried out according to ASTM C-423 or BS EN ISO 354:2003 for testing plane absorbers, i.e. flat areas of sound absorptive material such as carpeting or acoustic tiling, or for testing discrete sound absorbers such as chairs, free-standing screens, persons, pads or baffles. The reverberant sound field within the chamber should be kept as diffuse as possible during the steady-state and decay conditions since computation is based on this assumption. Hence, effort should be made to qualify a chamber for sound absorption measurements according to international acoustic standards.

The RT is initially measured in the empty room, later with the test item/items placed in the room. A band of random noise is generally used as the test signal since this enables measurements over the entire frequency range of interest in one pass; a harmonic signal may also be used. The source is kept on for a sufficiently long time to ensure that the sound pressure level in the room has reached steady state. The source is then suddenly turned off, causing the sound pressure level to decrease. The RT is then measured as the time taken for the SPL in the room to decay by 60 dB after the sound source has been switched off. The equivalent sound absorption area of the test specimen, a_T is computed using Sabine's equation, as the difference in Sabine with and without the object under test present in the reverberation chamber. Because of the sound absorption, the RT is shorter with the test specimens present. At least one loudspeaker position and three (preferably five) microphone positions with two readings in each case should be used, making the measured RT the average value of six (or 10) decay measurements in each frequency band to get space and time averaged decay rates. This process is done for each third octave band from 400 to 10,000 Hz. For the computation, the absorption coefficient of the materials of the room surface should be considered if they are significant; however, values for smooth, hard, rigid surfaces, such a reverberation room floor may be neglected.

Typical difficulties involved in this method: There may be spatial variations in the decay rates in the reverberation chamber due to the formation of standing waves. Such variations are significant at low frequencies and for test specimens with high sound absorption coefficients. Variations may be reduced by adding stationary and/or rotating diffusers suspended from the ceiling. Multiple microphone positions may be used, as also multiple sample positions. Standardized sample locations are defined in ASTM E795 and ISO 354, Amendment 1. Changes in temperature and humidity during the test, when the sample is taken into or out of the chamber, give rise to errors, especially at high frequencies. Insufficient amplifier and loudspeaker power or intrusive ambient noise create problems. Non-linear decays, that is, the decay rate not being constant, but varying with time, cause errors, especially at low frequencies, in small rooms, and when the sound absorption coefficient of the test specimen is high. Specimens with too small an area or too high an aspect ratio, as well as reverberation chambers with improper dimensions are other pitfalls.

For some test specimens, sound absorption coefficients greater than 1.00 are obtained. This happens in cases where the absorption footprint is larger than the area of the specimen and wave diffraction happens at the edges of the specimen. The smaller the frequency and specimen size and the larger the aspect ratio and the sound absorption, the larger is the diffraction effect. The Sabine absorption equation is

$$a_{ij} = \frac{0.9210 V d_j}{c_0} \tag{13.40}$$

where a_{ij} is the Sabine absorption of the reverberation chamber under the ith measurement condition in the jth frequency band (in metric Sabines), V is the chamber volume in m^3, d_j is the decay rate in dB per second for the jth frequency band and c_0 is the speed of sound.

13.6.3.1 *Plane absorbers*   These are generally placed directly against a surface of the reverberation chamber, usually the floor. Due to the non-parallel walls of the chamber, a sample placed on the floor will affect the decay time of all of the modes. The edges of the test specimen are sealed with an acoustically reflective frame made of steel, gypsum board or wood to prevent the edges from absorbing sound. For a plane absorber or a specified array of test objects, the acoustic absorption coefficient is obtained by dividing absorption a_r by the treated surface area S. For plane absorbers, the sound absorption coefficient is reported as:

$$\alpha_j = a_j/S + \alpha_{1j} \text{ for ASTM C423}$$

$$\alpha_j = a_j/S \text{ for ISO 354} \tag{13.41}$$

where the quantity α_{1j} is the estimated sound absorption coefficient of the surface covered by the plane absorber during the test.

13.6.3.2 *Discrete sound absorbers*   Discrete objects such as chairs, screens or persons are installed in the same manner as they are installed in practice. At least three individual test specimens should be used to provide a measurable change in the equivalent sound absorption area of the room. These individual objects are arranged randomly with a spacing of at least 2 m from each other and at least 1 m from any room edge. The sound absorption of the specimen is

$$a_j = a_{2j} - a_{1j} \tag{13.42}$$

where a_{2j} is the Sabine absorption of the reverberation chamber with the specimen present in the jth frequency band (in metric Sabines) and a_{1j} is Sabine absorption of the chamber with the chamber 'empty' in the jth frequency band (in metric Sabines).

For arrays of discrete absorbers, the sound absorption a_j of the array may be reported, or the averaged sound absorption of each element may be reported as a_j/N, where N is the number of elements in the array.

13.6.4 Noise Reduction Coefficient

A single number rating called the Noise Reduction Coefficient (NRC) is used to specify the absorbing properties of materials. It is computed as the average of the sound absorption coefficients measured in the 1/1 octave bands centred at 250, 500, 1000 and 2000 Hz and is rounded to the nearest multiple of 0.05 Sabine/m^2. The frequency range of interest may be extended down to the 63 Hz octave band in large chambers of concrete construction and sometimes to 5000 Hz on the upper side. Outside this frequency range, measurements are typically performed in an impedance tube as described in Sections 13.6.1 or 13.6.2.

13.7   MEASUREMENT OF SOUND TRANSMISSION LOSS

Acoustic barriers are used as acoustic insulators to block or prevent the passage of sound. Common barrier materials are gypsum board, plywood, concrete, etc. They are generally non-porous, heavy and air-tight materials. Soft porous materials and lightweight materials like rock or glass fibre on the other hand do not prevent the passage of sound; they absorb sound. When acoustic energy is incident on a barrier surface, some of it will be reflected and some will be transmitted through the barrier. Sound *TL* measures how effectively a material performs as a sound barrier. The fraction of incident energy transmitted is called the sound power transmission coefficient. The sound *TL* is 10 times the logarithm of the reciprocal of the ratio of the energy transmitted through the specimen to the energy incident on the specimen and is expressed in dB. It represents the loss in sound power due to transmission through the specimen. The higher the *TL*, lower is the sound passing through the wall. The expression for *TL* in dB is

$$TL = 10 \log\left(\frac{1}{\tau}\right) \tag{13.43}$$

where τ = the sound power transmission co-efficient.

There are two approaches to the measurement of *TL*. Both involve the measurement of SPL. The difference between the two is that one additionally uses RT, while the other uses sound intensity measurements. Such facilities are used for evaluating windows, doors and other elements that are built into walls. Table 13.8 shows the *TL* of some common building materials.

TABLE 13.8   Transmission loss of building materials

Barrier	TL (dB)
60 mm thick heavy wood door	25–30
100 mm brick	45
100 mm brick, plastered	47
200 mm brick, plastered	50
50 mm thick hardwood panel	30

13.7.1   Method using Sound Pressure Level and Reverberation Time as per ASTM E-90

Test Environment

This method involves the measurement of the sound *TL* of a specimen based on the difference in the space and time averaged SPLs between a source chamber and a receiving chamber. The source chamber is a reverberant chamber, while the receive chamber may be reverberant or semi-anechoic, the former being more popular. Each chamber should have a minimum volume of 50 m^3 coupled through an opening equal to the size of the test specimen. The measurement is based on the assumption that the sound fields in the two chambers are diffuse. Noise is generated in the source room and a part of the sound energy is transmitted through the test specimen into the receiving room. For generation of acoustic energy, a pink/white noise generator is typically used in conjunction with a power amplifier and speakers of sufficiently high rating so as to generate SPLs in the receiving chamber that are at least 10 dB higher than the ambient SPLs. They are

measured in both the source and receiver rooms in 1/3 octave bands in the frequency range between 100 Hz and 5000 Hz using an integrating SLM with 1/3 octave band analysis and RT measurement features. The frequency range of interest may be extended down to the 63 Hz octave band in large chambers and up to the 8,000 Hz octave band. The microphones chosen should have a flat frequency response over the frequency range of interest. Figure 13.17 shows a facility for measuring sound *TL*.
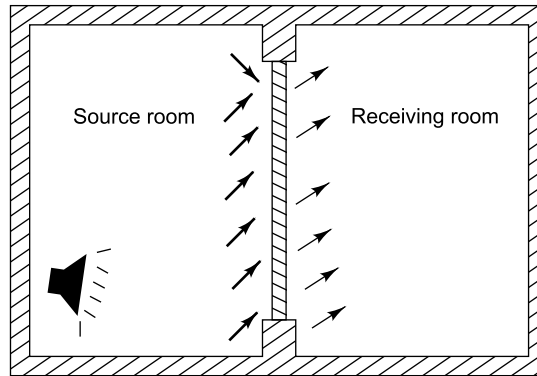
FIGURE 13.17   Facility for measuring sound transmission loss

It is advisable to have both the rooms supported on isolation springs to ensure efficient vibration and sound isolation between the two rooms. The support frame for the specimen is generally fixed to a concrete platform isolated from both chambers to reduce the transmission of structure-borne vibration from the structure to the two rooms and to ensure that the only path between rooms is through the specimen being tested. The *TL* limit of the test facility and any filler walls used to adapt the test specimen to the test opening should be known and the difference in levels should be corrected to account for their acoustical properties. The absorption co-efficient of the receiving chamber is to be determined from measurements of the RT. A facility for testing floors has one room above the other. There may be multiple paths other than through the test specimen in this case.

The major sources of measurement errors in the *TL* measurement which are highest at low frequencies are spatial/temporal variations in (i) the SPLs in both chambers and (ii) the decay rates in the receive chamber. These uncertainties may be reduced by spatial/temporal sampling and averaging of these parameters with a sufficiently large number of samples so as to ensure 95% confidence interval in the *TL* to the following values: 3 dB in the 125 and 160 Hz bands, 2 dB in the 200 and 250 Hz bands and 1 dB in the 315–4000 Hz bands.

The mathematical expression for *TL* is:

$$TL = L_{p(\text{source})} - L_{p(\text{receive})} + 10 \log_{10}(S/a) \tag{13.44}$$
$$TL = NR + 10 \log_{10}(S/a)$$

where

$L_{p(\text{source})}$ = space/time averaged sound pressure level in the source chamber in dB
$L_{p(\text{receive})}$ = space/time averaged sound pressure level in the receive chamber in dB
NR = Noise reduction
S = surface area of the test specimen in m^2
a = total absorption in receive chamber with the test specimen in place in Sabines (m^2)

13.7.2 Method using Sound Pressure Level and Sound Intensity

The measurement of the sound *TL* of a specimen by this method is based on the assumption that the sound field in the source chamber is diffuse, and that in the receiving chamber it is essentially a free field. The quantities to be determined are: sound pressure level in the source reverberation chamber and the average sound intensity radiated by the specimen into the receiving chamber. The mathematical expression is

$$TL = L_{p(\text{source})} - 6 - L_{i\,(\text{receive})} - 10\,\log_{10}(S) \tag{13.45}$$

where

$L_{p(\text{source})}$ = space/time averaged sound pressure level in the source chamber in dB

$L_{i(\text{recive})}$ = average sound intensity radiated by the specimen into the receive chamber in dB

S = surface area of the test specimen in m^2

The assumption behind this computation is that the angle of incidence of the energy in the source chamber is random in orientation and that the sound pressure level in the source chamber is proportional to the total energy incident on the specimen. The constant 6 dB term arises from the fact that source radiates into ¼ space, resulting in a 6 dB difference between the sound pressure level in the source room and the incident sound intensity. This technique for *TL* has not been standardized and the uncertainties in measurement are not well documented. Most of the instrumentation is similar to what is required for the earlier method. Additionally a sound intensity probe with two phase-matched microphones with uniform frequency response over the frequency range of interest, a dual channel analyzer with cross-spectrum analysis and 1/3 octave band filters are required.

13.7.3 Method using Impedance Tube

TL can be measured by the four microphone method in a plane wave tube. This method uses two sets of microphones, similar to that used in the measurement of absorption co-efficient using the two microphone method. The sound source (loudspeaker) is fixed at one end of the tube and the test sample is placed almost midway in a holder. A broadband stationary random noise is used to excite the loudspeaker in the source tube and the resulting plane waves propagate in the tube. The incident waves hit the sample, with part of the energy getting reflected back, part of it getting absorbed by the sample and part of it passing into receiving tube. The plane waves that exit the test specimen impinge on the termination at the end of the receiving tube and are reflected back. By measuring the sound pressures at the four fixed microphones, two in the source tube and two in the receiving tube, a complex transfer function is calculated using a four channel digital analyser. The contributions of the incident, reflected and transmitted waves are separated and TL is calculated. The impedance tube and microphone arrangement is shown in Fig. 13.18.
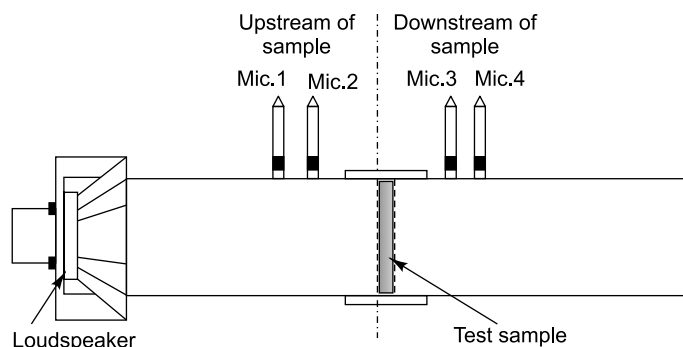
FIGURE 13.18 Schematic of transmission loss measurement setup using impedance tube

13.7.4 Single Number Ratings (STC)

Plots of sound TL data are complex and hence the sound TL of a specimen is often expressed in terms of a single number rating. The most common rating is the Sound Transmission Class (STC) of the specimen, which is determined in accordance with the calculation procedures of ASTM E413. The STC curve is a sliding contour that is fitted to the measured data plotted such that:

- the maximum deviation in any 1/3 octave band (i.e. for a TL value which falls below the contour) does not exceed 8 dB.
- the sum of the deviations falling below the reference contour in all bands does not exceed 32 dB.

Once the appropriate contour satisfying these conditions has been selected, the STC rating of a specimen is determined by comparing the measured TL values in the 16 contiguous 1/3 octave bands with centre frequencies between 125 and 4000 Hz, with the values of the STC reference contour shown in Fig. 13.19. The STC rating of a specimen corresponds to the TL value at 500 Hz for the highest reference contour which meets the above two conditions simultaneously and is expressed as a single number. As for TL, the higher the STC, the better the sound insulation. It may be noted that the STC contour is similar to the inverse of the equal loudness contour, in the sense that it discounts the lower frequency sounds to mimic the perception of the human ear.
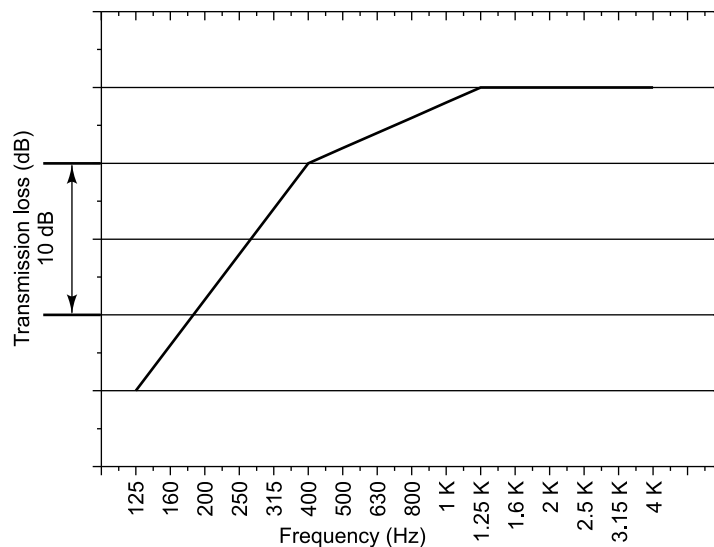
FIGURE 13.19   Reference contour for sound transmission class

Outside the USA, the Weighted Sound Reduction Index (SRI) or R_w defined as per ISO standard is used. This is defined in terms of the transmission coefficient, but uses a slightly different frequency range (100–3150 Hz). Measurement of SRI of a partition is also conducted in a pair of reverberation chambers as described for STC. In general, the two ratings give either the same number or only 1–2 dB difference, so they can generally be interchanged.

13.8 ACOUSTIC CHAMBERS

The most commonly used acoustic chambers are anechoic and reverberation chambers.

13.8.1 Anechoic Chamber

The word *anechoic* literally means echoless. Anechoic chambers are echo-free enclosures with sound absorption of 99% or more, or a reflection of 10% or less. In practice, anechoic chambers are designed, not just as echoless rooms, but also as rooms which are free from external noise. These rooms are used in the laboratory to create free-field condition, which implies that the sound level from a spherically radiating sound source in the chamber should decrease by 6 dB for every doubling of distance from the source. Also, there should be no external noise interference and no excessive variations in temperature, pressure and humidity.

To achieve these conditions, the rooms have to be large with massive walls and ceiling to isolate airborne and structure-borne sounds from outside. Anechoic chambers must have very low background noise and this consideration calls for superior sound insulation and isolation. The boundary surfaces of the chamber (walls, floor and ceiling) are covered with highly absorbing materials with normal incidence absorption co-efficient not less then 99% throughout the frequency range of interest. These rooms are often constructed as a vibration isolated room within another room. Since no sound can penetrate into such rooms, any sound produced inside the chamber can be studied for its intensity, frequency content, audibility, etc. Anechoic chambers are essentially used for research in acoustics and development work in fundamental research of speech and hearing. They are also used for studying sound patterns of loudspeakers, microphones and of consumer goods like washing machines, refrigerators, air-conditioners, automobile engines, etc. from the point of view of product comparison or noise control. Though the noise power of a machine can be calculated from sound intensity measurements made at site over a geometrical surface enclosing the machine, many quality standards are still based on the use of anechoic chambers.

The sound-absorbing materials typically used are porous materials like special acoustic foam, shaped in the form of wedges or perforated sheet metal wedges, filled with fibreglass insulation or acoustic tiles. The shape and dimensions of the acoustic foam wedges are dependent on the performance level required and cut-off frequency desired. The larger the lengths of the wedges, the lower are the frequencies they are capable of absorbing. The lowest frequency at which the sound energy absorption drops to below 99% or

FIGURE 13.20 Walls of anechoic chamber (Courtesy of IIT Madras)

reflected sound pressure level exceeds 10% is the cut-off frequency and the corresponding wavelength, the cut-off wavelength. Figure 13.20 is a photograph of an anechoic room, clearly showing the wedges.

Anechoic chambers should provide an ambient noise level of at least 20 dB, preferably 25 dB, below the lowest level of the signal to be measured. Their behaviour should adhere to the inverse square law variation of sound intensity under free field conditions. Consequently, the sound pressure level should decrease by 6 dB with every doubling of distance from the source. Under practical conditions, deviations of up to 1.5 dB are not uncommon in a fully anechoic room as per ISO 3745. Anechoic chambers are classified as:

(a) Fully Anechoic: all six surfaces are sound absorbent.

(b) Hemi Anechoic: the floor is a normal reflecting plane.

13.8.2   Reverberation Chambers

A reverberation chamber has hard surfaces which reflect sound and is used to simulate a diffuse field in the laboratory. It is designed such that no two walls are parallel, ensuring that multiple reflections take place at various angles of incidence to produce a non-directional or 'diffuse' sound field within the chamber. The sound power emitted from a machine can then be determined from the sound pressure level in the chamber and from the characteristics of the room such as its volume, surface area and RT. There are several design and construction factors to be considered in the construction of a reverberation chamber. The important ones are the cut-off frequency, inside and outside dimensions, exterior ambient and target interior noise levels, vibration isolation target and its means of achievement, shape, including non-parallel inside walls to avoid standing waves and non-porous walls to have high reflection. A very simple design could be in the form of twin parallelepiped pentagons (Fig. 13.21) such that no two walls are parallel, thus achieving multiple reflections, so as to achieve near diffuse field. The floor of the chambers is kept horizontal, while the ceiling of the chambers is slanting so that they are not parallel. For sound transmission loss measurements the chambers are designed in two different sizes, one being a large reverberation chamber and the other being a smaller one. The difference between the volumes of the chambers should be more than 10% as per the ISO140-1, ISO140-3 standards.
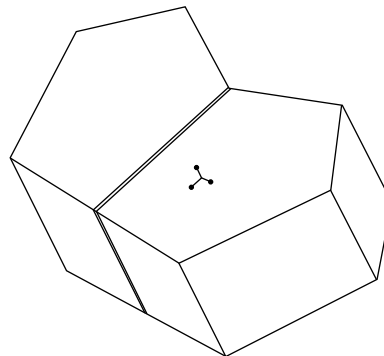
FIGURE 13.21   Isometric view of twin reverberation chambers

13.8.2.1   *Methods to achieve diffusivity*   Many methods have been devised in order to achieve a relatively high degree of diffusivity in reverberation chambers. Some of the more important ones are discussed further.

Rotating reflector: A motorized rotating reflector placed in the room effectively creates a variable boundary of the room, resulting in oscillation and the appearance and disappearance of modal patterns for a constant sound source. The circulation of the air in the room also breaks up the modal patterns to achieve a good diffuse sound field.

Reflecting surfaces: Highly reflective surfaces are added to the rooms by adding objects from the ceiling. This increases the sound reflecting area for a given room volume, without appreciably increasing the room absorption, thus enhancing the sound diffusion through multiple reflections.

Non-parallel walls: This enables the sound waves in the room to impinge on the walls randomly at various angles of incidence. For a constant volume, the total number of normal modes for an irregularly shaped room is very close to that for a room with parallel walls. However, the degeneracy of modes due to the overlap of modal patterns in an irregularly shaped room is much less and, consequently a more diffuse sound field is obtained.

Multiple frequency sound source: A source such as a warble (frequency modulated) tone or a third octave band noise has energy over a wider frequency range than a pure tone. Such a source is capable of generating a series of modal patterns, each catering to a narrowly grouped frequency band. The application of such a sound source is usually required, but is unrelated to the design and construction of the chambers.

13.9   COMMUNITY REACTION TO NOISE

Factories, residential areas, electrical substations, airports, highways, etc. are often near residential areas where the noise they radiate may be annoying. The reactions of individuals to noise from the above noise sources usually range from only mild remarks to legal action. At present there is no uniform method for determining community reaction to intruding noise. Different federal, state and local agencies have different criteria by which community noise levels are judged acceptable or unacceptable. The equivalent continuous noise level 'L_{eq}' has become the basic measure for describing community noise. It has been defined in Section 13.2.5.

13.9.1   Sound Standards and Noise Legislation

Legislation in countries such as Australia, Canada, Europe (EC), the United Kingdom and the United States, is fairly mature since the earlier laws which had been tried and tested for a number of years have now been remedied for their defects/limitations. By contrast, the legislation in countries like India is fairly new and/or quite limited in scope and is derived from international standards. The kind of legislation of concern to us is a noise abatement act, which is needed to control specific noise-making activities. This is usually based on different kinds of zones.

Industrial Zone: The land uses in this category would include, but would not be limited to, manufacturing activities, transportation facilities, warehousing, mining and other lands intended for such uses.

Commercial Zone: The land uses in this category would include, but would not be limited to, retail trade, personal, business and professional services, government services, amusements, agricultural activities and lands intended for such commercial or institutional uses.

Residential Zone: The land uses in this category would include, but not be limited to, single and multiple family homes, hotels, prisons, religious facilities, cultural activities, forest preserves, and land intended for residential or special uses requiring such protection.

Silence Zone: The land use in this category would include, but not be limited to, hospitals, educational institutions and courts. In order to ensure silence at such premises, the zone should extend to an area of 100 m around such institutions. Certain activities (e.g. the use of car horns and loudspeakers) are banned in a silence zone.

Table 13.9 shows a selected framework of standards for the above-mentioned zones from the Central Pollution Control Board of India. Both the day time and night time limits in terms of dB (A) are given. Table 13.10 shows the noise standards from other countries.

TABLE 13.9   Allowable noise limits (India)

Area zone	Daytime limits dB(A)	Night-time limits dB(A)
Industrial area	75	70
Commercial area	65	55
Residential area	55	45
Silence zone	50	40

TABLE 13.10   Allowable noise limits from other countries

Country	Industrial area Day/Night dB(A)	Commercial area Day/Night dB(A)	Residential area Day/Night dB(A)	Silence zone Day/Night dB(A)
Australia(1)	65/55	55/45	45/35	45/35
Australia(2)	65/65	60/60	50/40	45/35
Canada(1)	60/55	60/55	55/45	–
Canada(2)	65/60	65/60	55/45	–
Israel	70	55	50	45
Japan	60/50	60/50	50/40	45/35
Mauritius	60/55	60/50	60/50	–
U.S.(1)	75/75	65/65	60/60	–
U.S.(2)	65/65	65/65	65/55	–
U.S.(3)	70	62	55/50	–
U.S.(4)	80/75	65/60	55/50	–
U.S.(5)	60/55	60/55	55/50	–
U.S.(E.P.A.)	70	60	55	45
W.H.O. & E.C.	65	55	55/45	45/35

Notes: Australia(1): Australian Capital Territory; Australia(2): Australian Northern Territory; Canada(1): District of North Vancouver; Canada(2) : District of Burnaby; U.S.(1): State of Minnesota (from Pollution Control Agency); U.S.(2): State of Delaware (from Department of Natural Resources); U.S.(3): Huntsville, Alabama; U.S.(4): Denver; Colorado; U.S.(5): Davis, California; U.S.(EPA) : Environmental Protection Agency.

The last two rows give recommended (not legislated) values from the United States Environmental Protection Agency and from the World Health Organization respectively. The European Community endorses values approved by the World Health Organization.

13.9.2   Industrial Noise Regulations

One of the major reasons for industrial noise control programs is to prevent permanent hearing damage due to occupational noise exposure. A section of the 'Occupational Safety and Health Act' of 1970 was dedicated to the control and regulation of industrial noise. Basically, the OSHA noise regulations state that the noise level for a continuous exposure over an 8-hour period shall not exceed 90 dBA. Table 13.11 gives the allowable exposure levels for other exposure times.

For exposure levels and times not specified in Table 13.11, the permissible exposure *T* in hours to a continuous noise at a single level is given by

$$T = \frac{16}{2^{\{0.2(L-85)\}}}$$

(13.46)

TABLE 13.11   Allowable exposure levels for various exposure times

Sound level (dBA)	Time permitted (hr:min)	Sound level (dBA)	Time permitted (hr:min)
85	16	101	1:44
86	13:56	102	1:31
87	12:8	103	1:19
88	10:34	104	1:9
89	9:11	105	1:0
90	8:0	106	0:52
91	6:58	107	0:46
92	6:4	108	0-40
93	5:17	109	0:34
94	4:36	110	0:30
95	4:0	111	0:26
96	3:29	112	0:23
97	3-2	113	0:20
98	2:50	114	0:17
99	2:15	115	0:15
100	2:0		

where L is the A-weighted work place sound pressure level. In no case should the exposure level to continuous noise be greater than 115 dBA.

The exposure to continuous noise at two or more levels should not exceed a daily Dose D of unity, where D is given by

$$D = \frac{C_1}{T_1} + \frac{C_2}{T_2} + \frac{C_3}{T_3} + \ldots \frac{C_n}{T_n} \tag{13.47}$$

where $C_1, C_2, C_3, \ldots C_n$ are the actual exposure times in hours at given steady state A-weighted noise levels and $T_1, T_2, T_3, \ldots T_n$ are the exposure limits in hours, calculated from Eq. (13.46) for the steady state A-weighted noise level. The noise dose can be measured by an instrument called noise dosimeter.

FURTHER READINGS

1. American National Standards Institute, ANSI S1.40-Specifications for Acoustic Calibrators, 1984, R1997.
2. ASTM E1332, Standard Classification for Determination of Outdoor-Indoor Transmission Class.
3. ASTM E413-04, Classification for Rating Sound Insulation.
4. ASTM E90, Standard Test Method for Laboratory Measurement of Airborne Sound Transmission Loss of Building Partitions.
5. Barnard, A.R, and Rao, M.D, Measurement of Sound Transmission Loss Using a Modified Four Microphone Impedance Tube, in *Proceedings of the 2004 National Conference on Noise Control Engineering,* Baltimore, Maryland, 12–14 July, 2004.
6. Bies, D.A. and Hansen, C.H., *Engineering Noise Control: Theory and Practice*, Taylor & Francis, New York, USA, 2003.
7. Brüel & Kjær, *Acoustic Noise Measurements*, BT 0010-12, Denmark.
8. Brüel & Kjær, *Frequency Analysis of Sound*, BA 7660-06, Denmark.

9. Brüel & Kjær, *Noise Control–Principles and Practice*, Denmark,1986.

10. BS EN 1793-1:1998, Road Traffic Noise Reducing Devices-Test Method for Determining the Acoustic Performance, Part 1: Intrinsic Characteristics of Sound Absorption.

11. BS EN 1793-3:1998, Road Traffic Noise Reducing Devices-Test Method for Determining the Acoustic Performance, Part 3: Normalized Traffic Noise Spectrum.

12. BS EN ISO 140-1:1997, Acoustics-Measurement of Sound Insulation in Buildings and of Building Elements, Part 1: Requirements for Laboratory Test Facilities with Suppressed Flanking Transmission.

13. BS EN ISO 140-2:1991, Acoustics-Measurement of Sound Insulation in Buildings and of Building Elements, Part 2: Determination, Verification and Application of Precision Data.

14. BS EN ISO 140-3:1995, Acoustics-Measurement of Sound Insulation in Buildings and of Building Elements, Part 3: Laboratory Measurement of Airborne Sound Insulation of Building Elements.

15. BS EN ISO 354:2003, Acoustics-Measurement of Sound Absorption in a Reverberation Room.

16. BS EN ISO 717-1:1997, Acoustics-Rating of Sound Insulation in Buildings and of Building Elements, Part 1: Airborne Sound Insulation.

17. BSI BS EN 1793-2, Road Traffic Noise Reducing Devices-Test Method for Determining the Acoustic Performance, Part 2: Intrinsic Characteristics of Airborne Sound Insulation.

18. Burns, W. and Robinson, D.W., *Fundamentals of Acoustics Hearing and Noise in Industry*, Her Majesty's Stationery Office, London, 1970.

19. Chung J.Y., Cross-Spectral Method of Measuring Acoustic Intensity, Research Publication of General Motors Research Laboratory (GMR-2617), Warren (USA), 1977.

20. Chung, J.Y. and Pope, J., Practical measurement of acoustic intensity—The two-microphone cross-spectral method, Inter-noise 78: Designing for Noise Control, in Proceedings of the International Conference, San Francisco, CA, May 8–10, 1978 (A79-15551 04-71), Noise Control Foundation, Poughkeepsie, N.Y., 1978, pp. 893–900.

21. Diehl, G.M., *Machinery Acoustics*, Wiley, New York, 1973.

22. Fahy F.J., Measurement of acoustic intensity using the cross-spectral density of two microphones Signals, *Journal of the Acoustical Society of America*, 62 (L), 1057–1059, 1977.

23. Fahy F.J., *Sound Intensity*, Elsevier Applied Science, London, (UK), 1989.

24. Foreman, J., *Sound Analysis and Noise Control*, Van Nostrand Reinhold, New York, 1990.

25. Fry, A., Noise *Control in Building Services*, Pergamon Press, Oxford, New York, 1988.

26. Goelzer, B., Hansen, C.H. and Sehrndt, G.A., Occupational Exposure to Noise: Evaluation, Prevention, and Control, Special Report S 64, Dortmund and Berlin: Federal Institute for Occupational Safety and Health, 2001.

27. ISO 3740:2000, Acoustics-Determination of Sound Power Levels of Noise Sources—Guidelines for the Use of Basic Standards and for the Preparation of Noise Test Codes.

28. ISO 3741:1999, Acoustics-Determination of Sound Power Levels of Noise Sources Using Sound Pressure-Precision Methods for Reverberation Rooms.

29. ISO 3743-1:1994, Acoustics-Determination of Sound Power Levels of Noise Sources—Engineering Methods for Small, Movable Sources in Reverberant Fields, Part 1: Comparison Method for Hard-Walled Test Rooms.

30. ISO 3743-2:1994, Acoustics-Determination of Sound Power Levels of Noise Sources Using Sound Pressure—Engineering Methods for Small, Movable Sources in Reverberant Fields, Part 2: Methods for Special Reverberation Test Rooms.

31. ISO 3744:1994, Acoustics-Determination of Sound Power Levels of Noise Sources Using Sound Pressure—Engineering Method in an Essentially Free Field over a Reflecting Plane.

32. ISO 3745:2003, Acoustics-Determination of Sound Power Levels of Noise Sources Using Sound Pressure—Precision Methods for Anechoic and Hemi-Anechoic Rooms.

33. ISO 3746:1995, Acoustics-Determination of Sound Power Levels of Noise Sources Using Sound Pressure—Survey Method Using an Enveloping Measurement Surface over a Reflecting Plane.

34. Kinsler, L.E. Frey, A.R., Coppens, A.B. and Sanders, J.V., *Fundamentals of Acoustics*, John Wiley and Sons, New York, 2000.

35. Moore, B.C.J., *An Introduction to the Psychology of Hearing*, Academic Press, Orlando, Florida, 2003.

36. Pieper, R. M., Alexander, J., Bolton, J. and Tae-wook Yoo, Assessment of absorbers in normal incidence four-microphone transmission-loss systems to measure effectiveness of materials in *Proceedings of the Lateral-Flow Configurations of Filled or Partially Filled Cavities*, SAE 2007-01-2190, May 2007.

37. Reynolds, D.D., *Engineering Principles of Acoustics: Noise and Vibration*, Allyn & Bacon, Boston, USA, 1981.

38. Robinson, D.W., Towards a unified system of noise assessment, *Journal of Sound and Vibration*, 14, 279-98, 1971.

39. Suter, A.H., Noise, in ILO *Encyclopaedia of Occupational Health and Safety*, 4th ed., wholly rearranged and revised in print, on CD-ROM and online, International Labour Organization, Geneva, 1998.

40. Wilson, C.E., *Noise Control: Measurement, Analysis and Control of Sound and Vibration*, Krieger Pub. Co., Malabar, Florida, 1994.

Typical Vibration/Noise Measurements and Case Studies

In this chapter an effort has been made to relate the theory and practices described in the previous chapters to real life situations through case studies, with a view to bringing out the flavour of the measurements generally made. About fifteen simple case studies encompassing the areas of machine dynamics, vehicular vibration and machinery acoustics have been discussed. The aim is not to give complete details of the tests and the results, but information regarding the choice of transducers, equipment and measurement techniques used, as well as the parameters typically used for signal analyses. The subsequent processing and mode of display of data captured have also been dealt with.

14.1 EVALUATION OF LOSS FACTORS OF RUBBER SPECIMENS

This exercise was conducted to determine the loss factors of certain rubber materials which were being used as deadeners for automotive applications. The chemical compositions of these rubbers had been tweaked, so as to result in large values of loss factors. Typically, evaluation of loss factors is done very conveniently using what is called a complex modulus apparatus, which is nothing but a setup which facilitates forced vibration tests to be conducted on bar-like specimens for the measurement of dynamic (complex) modulus of elasticity and the measurement of internal damping/loss factors. Alternately, the experiment can be done using any forced vibration setup.

The complex modulus apparatus or Oberst apparatus (Type 3930, Brüel & Kjær, Denmark) was used for the purpose. This special apparatus allows the test specimen to be fixed either in a clamped-free or clamped–clamped condition and is so designed as to enable firm clamping of the samples with very low parasitic losses, and precise mounting of the transducers with respect to the sample. It has a guide pillar with adjustable supports for two identical non-contact electromagnetic transducers (Type MM0002, Brüel & Kjær, Denmark). One transducer, typically the one near the free end (if the test is being conducted in the cantilever mode), is used as the non-contact vibration exciter, since larger vibration amplitudes can be imparted here than near the fixed end. The other transducer is used as the vibration sensitive pickup.

The rubber specimens, the loss factors of which were to be evaluated were available in the form of sheets of thickness 2 mm. Some of them had a fibre glass cloth covering, while others had a very thin aluminium sheet covering. They were cut and bonded on to mild steel bars (having negligible damping) of size 185 mm × 12.2 mm × 0.9 mm. In this particular study, the beam was set into forced harmonic vibration; for this the exciter was fed with a variable frequency sinusoidal excitation signal from a function/arbitrary waveform generator, (33120A, Hewlett Packard, USA). Figure 14.1 shows a close-up of the Complex Modulus Apparatus along with the associated instrumentation.
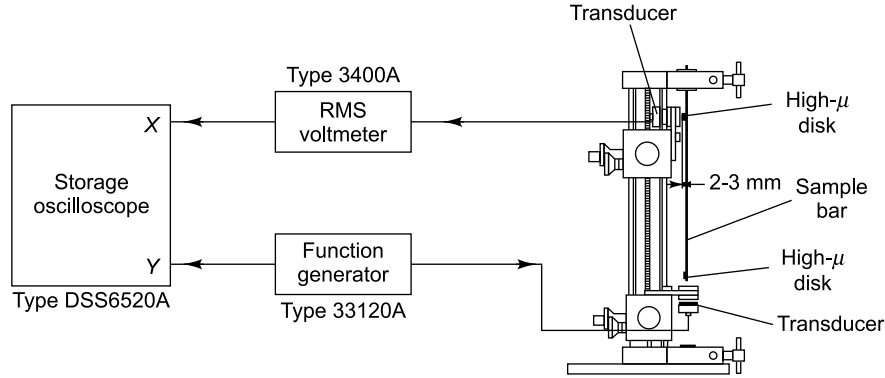
FIGURE 14.1   Setup for measurement of complex modulus

When the forcing frequency is changed, the vibration response changes and shoots up at the natural frequencies, the amplitude at resonance being limited only by the mechanical damping present in the bar. Natural frequencies of the first few modes and the corresponding half-power frequencies were determined using a digital storage oscilloscope (DSS 6520A, Kikusui Electronics Corporation, Japan) and an RMS voltmeter (3400A, Hewlett Packard, USA). The oscilloscope was used for monitoring the sinusoidal response of the strip and for ensuring that the readings were taken in the linear vibratory range of the specimen. The RMS voltmeter was used to read off the response voltage directly in dB to find out the half power frequencies. From the measured values, the loss factors were found by the half power method using the formula:

$$\eta = \frac{f_2 - f_1}{f_n} \tag{14.1}$$

where f_n is the natural frequency and f_1 and f_2 are the lower and upper half-power frequencies, respectively. The average loss factors measured for the first few modes of vibration were in the range 0.02–0.5 for rubber materials of various compositions.

14.2   EVALUATION OF VIBRATION DAMPING CHARACTERISTICS OF TUBES

In this experiment, the vibration damping characteristics of coaxially drawn electric resistance welded Carbon steel tubes were evaluated. Test specimens of length 445 mm, outer diameter 30 mm and thickness 6 mm were used. Previously published literature indicated that the most appropriate method of testing such tubes is by using free–free boundary conditions. Hence tests were initially conducted with free–free boundary conditions by imparting transient excitation to the tubes using an impact hammer (PH51, Rion, Japan with Rion 7117 force gauge). To achieve free–free conditions, the tubes were suspended from the ceiling at one end using a small chuck, a light spring and thin nylon wire. The experiment did not provide conclusive results in terms of the flexible modes of vibration due to interference from rigid body modes in the useful frequency range. Free–free conditions were subsequently attempted in another setup where the specimens were supported on soft sponges at the ends and then excited by an impact hammer with steel tip, but no conclusive results could be obtained in this case also. It was then decided to conduct tests with fixed-free conditions. The tube was fixed on a cement block at one end to arrest

movement. Figure 14.2 shows this experimental setup. The vibration signals were picked up using a piezoelectric accelerometer (Type 4332, Brüel & Kjær, Denmark) since damping ratios up to around 5 kHz were required. The output of the accelerometer was conditioned using a charge amplifier (Type 2626, Brüel & Kjær, Denmark). A dual channel dynamic signal analyser (35670A, Agilent Technologies, USA), was used to read signals both from the accelerometer and the force transducer of the impact hammer and provide frequency response functions. These tests gave satisfactory vibration FRFs, and damping ratios corresponding to the first four modes of vibration were extracted.
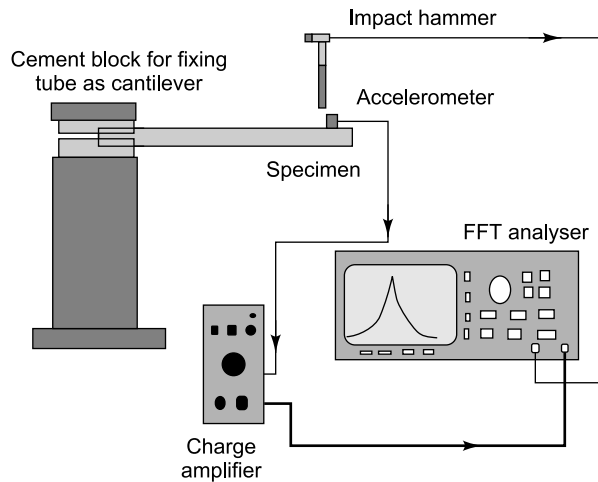
FIGURE 14.2   Test setup for evaluation of damping ratio

Figure 14.3(a) shows a typical accelerance function. Damping ratio corresponding to each mode was obtained using the zoom feature in the analyser and the half power method for computation. Due to the limited useful frequency range of these tubes for various applications, only the first four modes were analysed for vibration damping. The average value of damping ratio for each tube was found for the various modes and is shown in Fig. 14.3(b).

14.3   VIBRATION TEST ON TORSIONAL VIBRATION DAMPER

A torsional vibration damper pulley is used in vehicles for reducing potentially damaging torsional vibrations in the crankshafts of internal combustion engines. This type of damper reduces the vibration when a torsional excitation frequency equals the first natural frequency of the crankshaft, but not at other speeds. If such vibration is not controlled, it can lead to failure of the crankshaft or other accessories. The aim of the experiment described was to determine the natural frequency and damping ratio of the given tuned absorber type of torsional vibration damper pulley. Two lugs were welded to the outer periphery of the damper, diametrically opposite to each other. The damper was fixed to a rigid support by holding the shaft keyed to the inner hub as shown in Fig. 14.4 and was excited in the torsional mode by an electrodynamic exciter (PR 9270/01 Philips, Germany), the excitation point being one of the lugs. Such an arrangement was used due to non-availability of a torsional exciter.

(a)

(b)
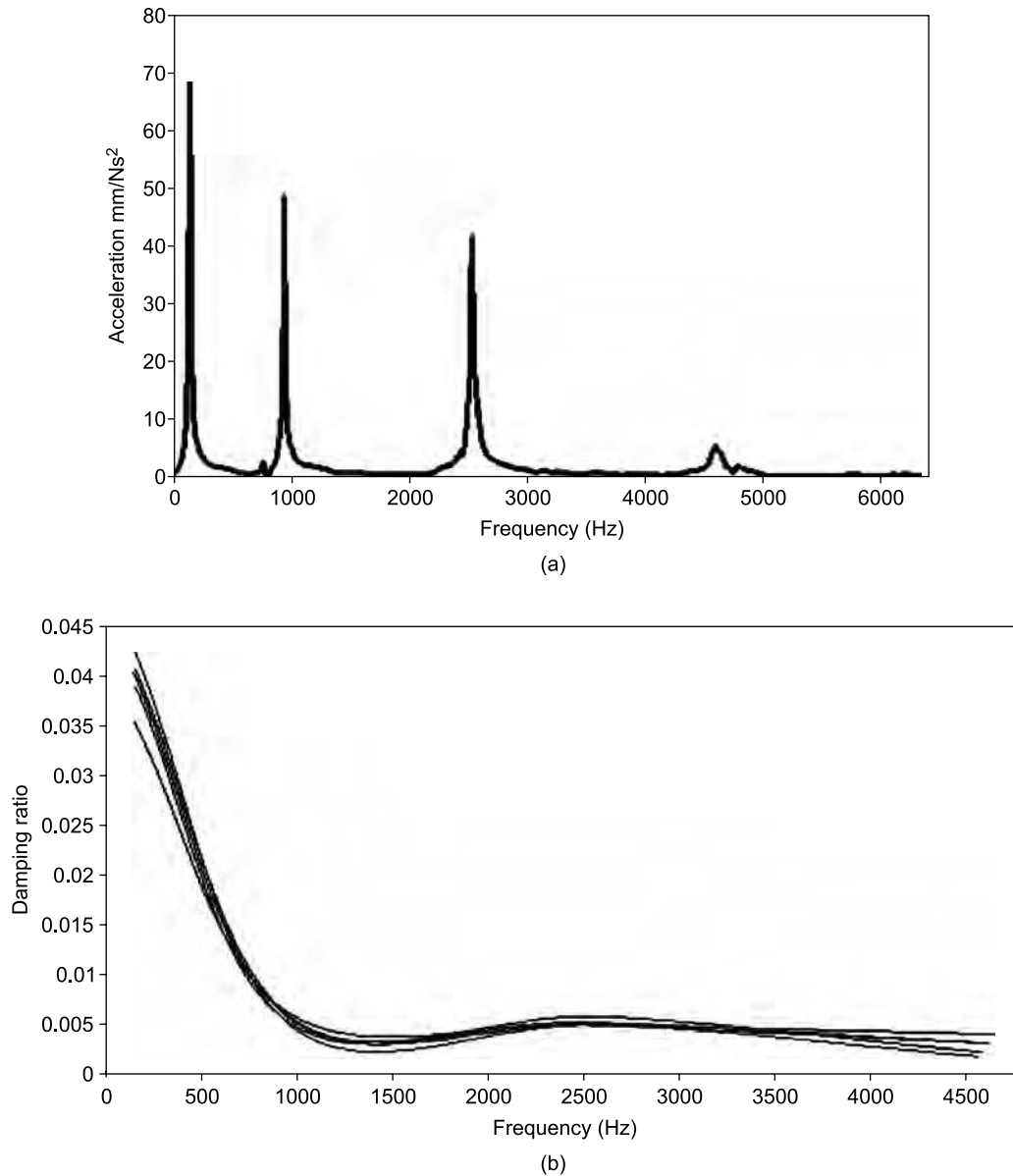
FIGURE 14.3   Co-axial tube characteristics: (a) FRF in 0–6.4 kHz range, (b) damping ratio versus frequency

   Sinusoidal excitation was imparted using a function generator (33120A, Hewlett Packard, USA). The acceleration signal was picked up by a piezoelectric accelerometer (4332, Brüel & Kjær, Denmark) fixed to the other lug. The natural frequency of torsional oscillation was found out corresponding to the maximum in measured output voltage. A storage oscilloscope, (54624A, Agilent Technologies, USA) was used to determine if the vibration corresponded to a bending mode or torsional mode. The phase shift between the acceleration signals at the two lugs was used for this. This phase shift is 0° in the case

of bending vibration and 180° in the case of torsional vibration. The damping ratio was found out by the half-power method using an RMS volmeter (3400A, Hewlett Packard, USA). The first natural frequency and damping ratio of the damper are as shown in Table 14.1. The values reported are the average values taken from three sets of measurements.
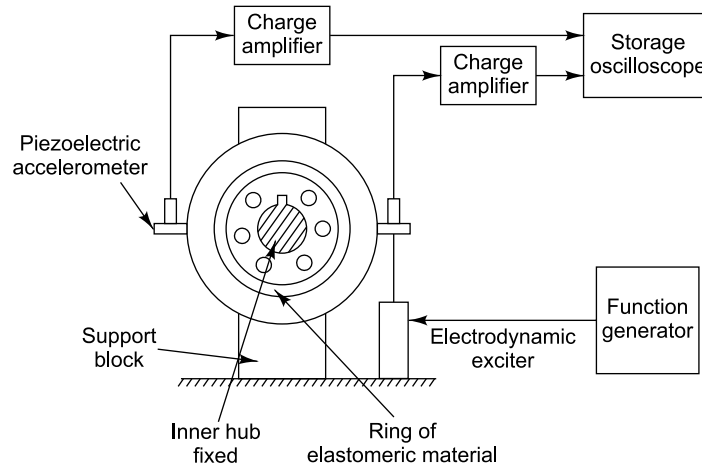
FIGURE 14.4   Setup for testing torsional damper

TABLE 14.1   Damping ratio results

Natural frequency, f_n (Hz)	Lower half power frequency, f_1 (Hz)	Upper half power frequency, f_2 (Hz)	Damping ratio, ζ
118.6	116.9	125.4	0.036

14.4   CALIBRATION OF TRANSDUCERS

The vibration engineer is often faced with the task of calibrating a transducer for which calibration data have been lost, or for which it is a long time since calibration was done. A shaker table with a known output displacement over a range of frequencies may be used for this. Alternately, the pick up to be calibrated and a reference pick up, the sensitivity of which is known and is constant over the required frequency range, may be fixed on the same shaker and the sensitivity of the pick up to be calibrated determined, knowing that of the reference transducer and the table vibration amplitude. In this section, the procedure for calibration of an accelerometer is described.

The accelerometer to be calibrated and the reference accelerometer (with available calibration sheet) were mounted on a mechanical shaker (ST 40, Losenhausenwerk, Germany). The transducer to be calibrated was supplied with appropriate excitation voltage as required. The table was driven from 0 to 60 Hz and the outputs of both transducers were read on an analyser at 5 Hz intervals. The sensitivity of the test accelerometer was calculated in terms of mV/g and plotted as a function of frequency as shown in Fig. 14.5.
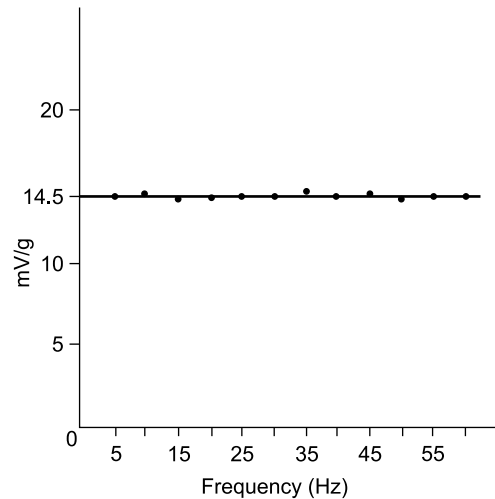
FIGURE 14.5 Calibration curve of accelerometer

14.5 GROUND VIBRATION DUE TO BLASTING

This section describes vibration measurements carried out to ascertain if the cracks that had developed in some residences in the vicinity of a stone quarry, had been caused by ground explosives used by a construction company for blasting operations. A stone quarry located on a hillock about 100 m from ground level was being used for the purpose. As per the existing rules, there should be a safe distance of 300 m between any quarry site and habitation area and this condition was also met. Houses situated near the quarry site (beyond the safe distance of 300 m), and from where there were complaints, were identified to find out the effects of blasting. Ground vibration measurements were carried out at two residences having cracks attributed to the blasting operation, one at a residence about 430 m from the quarry (Site 1) and another at 746 m (Site 2).

It was decided to use inductive accelerometers to capture the low frequency vibrations propagating from the blast, though the data were finally required in the form of vibration velocity spectra. The vibration velocities were obtained by integrating the acceleration signals. The velocity transducer available in the laboratory would not have responded well at low frequencies and hence this decision. At each site, a concrete block with a steel plate bonded on top, was firmly embedded in the ground. Two inductive accelerometers B12/200 (HBM, Germany) were attached to this steel plate, one measuring vibrations in the longitudinal direction (facing the quarry site) and the other measuring vibrations in the vertical direction. At Site 1, the waveform from the blast was recorded using a waveform recorder (HI 8808, Hioki, Japan) and later analysed in the laboratory. At Site 2, the signal was sent to another waveform recorder (HI 8803, Hioki, Japan) and also to a digital signal analyser (Type 35670A, Hewlett-Packard, U.S.A.) and stored for later analysis. The acceleration data were signal conditioned using a six channel carrier frequency amplifier, (KWS 3073, HBM, Germany). Figure 14.6 shows the acceleration measurements made at Site 2 in the longitudinal direction. In this particular measurement, the signal level was found to be comparable to the noise floor. The actual signal extends up to 1.5 s, beyond which is noise. The measured time histories were converted to velocity spectra using MATLAB in order to

calculate the Peak Particle Velocity (PPV) in various frequency bands. Table 14.2 shows the limits on peak particle velocity (PPV) in mm/s that has been set for Mines Safety in India.
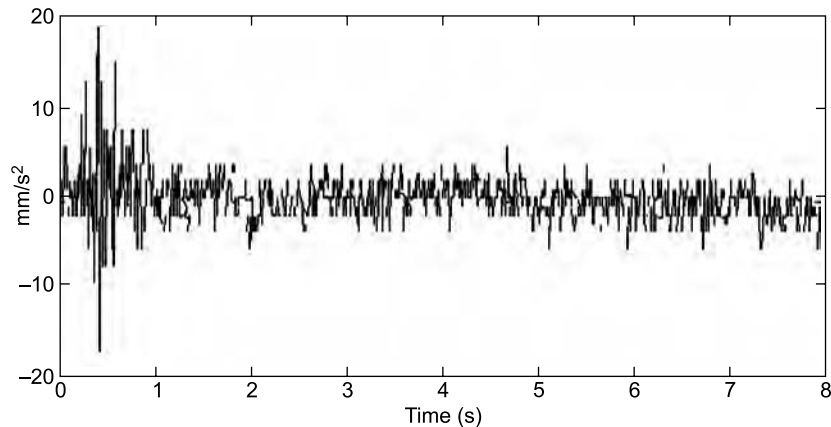
FIGURE 14.6 Longitudinal acceleration during blasting

TABLE 14.2 PPV limits as set by Directorate General of Mines Safety (DGMS)

Type of structure	Dominant frequency (Hz)		
	< 8 Hz	8–25 Hz	> 25 Hz
Buildings/structures			
Domestic houses/structures (Brick/cement)	5 mm/s	10 mm/s	15 mm/s
Industrial buildings (RCC/Framed Structures)	10 mm/s	20 mm/s	25 mm/s
Objects of historical importance/sensitive structures	2 mm/s	5 mm/s	10 mm/s

The vibration measurements made indicated that the PPV at Sites 1 and 2 in the vertical direction were 3.8 and 1 mm/s, respectively in the <8Hz band, while those in the longitudinal directions were 3.4 and 0.9 mm/s, respectively. The values were insignificant in the other frequency bands. From the vibration measurements made, it was seen that the levels did not exceed the limits set by DGMS (PPV of 5 mm/s for frequencies less than 8 Hz) at the residences. Hence it was concluded that the cracks were unlikely to have been caused by ground vibrations.

14.6   A CASE STUDY ON FLOOR VIBRATION MEASUREMENTS

A gear inspection system was to be installed in an engine manufacturing company. Before the installation, vibration measurements had to be made, specifically at a few points on the floor in the 4 m × 4 m area of the transmission shop where the involute and helix angle measuring machines were to be installed. Two linear variable differential transformers (B4 Z and B2/25, HBM, Germany) were used to sense the floor vibration displacement amplitudes. The outputs of the transducers were connected to carrier frequency amplifiers (KWS 3073, HBM, Germany). The conditioned outputs were recorded using a recorder and were also analysed using a real-time analyser (SA 73, Rion, Japan). Most of the vibration levels reached the highest recorded values during the time 12:30 P.M. to 3 P.M. when all the nearby machines were operating. The measured floor vibration levels at six locations during the peak operational

time were found to vary between 0.2 and 0.7 μm (peak to peak) in the vertical direction. The maximum recorded vibration amplitudes in the lateral and longitudinal directions were found to be 0.5 and 0.55 μm (peak to peak). A typical spectrum had vibration components at 4, 9 and 19 Hz with spectral amplitudes of 0.19, 0.18 and 0.026 μm and with an overall value of 0.25 μm. It was clearly observed that the overall vibration level did not reach 1 μm (peak-to-peak), even with the nearby CNC machine and crane operational. This vibration level was well within the maximum value of 12 μm (peak-to-peak) specified by the gear inspection system vibration isolation guide. Hence, it was recommended that the gear inspection system could be safely installed there.

14.7 Vibration Isolation of a 28-HP Engine

14.7.1 Vibration Measurements on Engine

There were complaints of excessive vibration in a 28 HP, two cylinder, four stroke engine (meant to be used in an off-road vehicle) when it was tested on its test bed in the engine factory. It was found to vibrate severely in the 850–1500 rpm range, though its critical speed as per calculation, was supposedly much higher, around 4500 rpm. Measurements were made on the engine test bed to identify the cause of this vibration. The piezoelectric accelerometer of a vibration analyser (CF 250, Onosokki, Japan) was fixed at various locations on the engine base using a magnetic base, in the longitudinal, transverse and vertical directions. The overall vibration levels and relevant spectral values were recorded at 850 and 1600 rpm at no load and full load and are shown in Table 14.3 for the transverse direction. The vibration levels were minimum in the vertical direction (not reported). The levels in the axial direction (not shown) and transverse direction were about 1.5 and 2 times those in the vertical direction. The vibration levels on the dynamometer and propeller shaft were negligibly small (not reported).

TABLE 14.3 Measured vibration levels (*g*) in transverse direction

Location on engine base	At 850 rpm; no load		At 850 rpm; 78.4 Nm load		At 1600 rpm; no load		At 1600 rpm; 87 Nm load	
	overall	*@35.5 Hz*	*overall*	*@35.5 Hz*	*overall*	*@35.5 Hz*	*overall*	*@35.5 Hz*
Rear right	0.237	0.174	0.289	0.168	0.360	0.0482	0.394	0.0420
Front right	–	–	0.257	0.147	0.220	0.0386	0.385	0.0551
Rear left	0.235	0.171	0.313	0.189	0.358	0.0507	0.369	0.0507
Front left	0.212	0.156	0.264	0.15	0.223	0.0319	0.354	0.0290

Figure 14.7(a) shows a typical vibration spectrum in the transverse direction with the engine running at 850 rpm at full load. This figure has a strong peak at 35.5 Hz, clearly indicating that the engine was undergoing resonance. With the engine running at 850 rpm, the (2½) × running speed component at 35.5 Hz was found to set the system into resonance. Subsequently a free vibration test was conducted by impacting the engine and the transient vibration was recorded. Figure 14.7(b) corresponds to a spectrum of the free vibration record clearly showing a natural frequency of the engine in the transverse direction at 35 Hz.
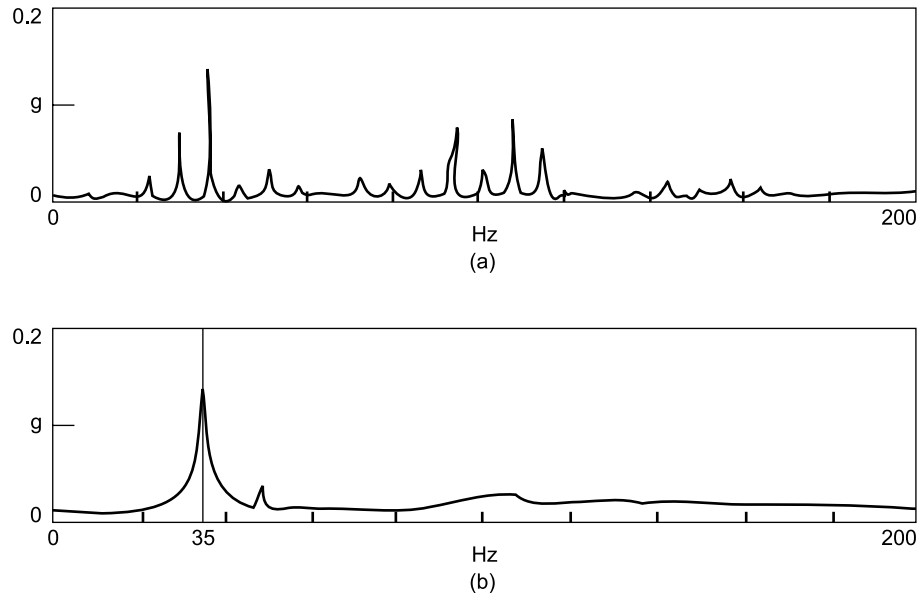
FIGURE 14.7   Typical vibration spectra on engine: (a) spectrum-transverse direction-850 rpm-full load,
(b) spectrum of free vibration record

14.7.2   Transmissibility Measurements of Engine Mounts

The engineers in charge of the test had felt that this resonant condition at 850 rpm could be avoided by
mounting the engine on the test bed through rubber pads, instead of directly bolting it to the test bed
foundation. They had arbitrarily tried using four identical
shear isolators which were available (without doing any
vibration isolation computation or measuring the stiffness
of the pads), to reduce the vibration, but the effect was
adverse. Hence, it was decided to evaluate the stiffness of
these isolators. For this, the isolators were mounted beneath
a double mass reaction exciter of mass 265 kg (almost equal
to the mass of the engine) as shown in Fig. 14.8.

Measurements were made for three different
configurations:

 (i) The exciter was constrained to vibrate in the
    horizontal direction with the shear direction
    perpendicular to the direction of exciting force; the
    vibration velocities in the direction of exciting force
    and perpendicular to it were measured.

(ii) The exciter was constrained to vibrate in the
    horizontal direction with the shear direction along
    the direction of exciting force; the vibration velocities
    in the direction of exciting force and perpendicular
    to it were measured as before.
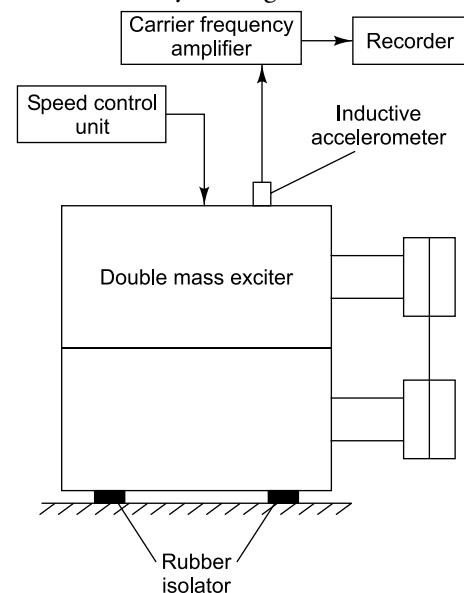
FIGURE 14.8   Setup for determination of
isolator stiffness

(iii) Finally, the exciter was constrained to vibrate in the vertical direction and the vibration velocity in the vertical direction was also measured.

The speed of the exciter was swept from 100 rpm to around 3300 rpm. At each speed, the vibration of the exciter was sensed using the piezoelectric accelerometer and fed to a vibration analyser (with built-in integrator) to display vibration velocity. Ideally a test like this should be done using an exciter which gives a constant excitation force. Since the reaction exciter used gave a force output proportional to ω^2, the square of the running speed (being an unbalance mass exciter), the measured velocity values had to be normalized to simulate the effect of a constant force in the forced vibration test. The normalized velocity was calculated as

$$\text{Velocity}_{\text{normalized}} = \text{Velocity}_{\text{measured}} \times \left(\frac{\omega_{\text{ref}}}{\omega}\right)^2 \tag{14.2}$$

Here the normalization was done taking ω_{ref} = 1000 rpm. The normalized velocities were plotted as a function of running speed. From these plots, the resonant frequencies of the spring-mass system (constituted by the exciter and the four rubber isolators) were found to be at 420 rpm (7 Hz), 980 rpm (16.33 Hz) and 2000 rpm (33 Hz) in the three directions mentioned above, with the last value in the vertical direction. It was concluded that resonant frequencies obtained from this test setup could be expected with the combination of the test engine and isolators also, since the engine mass was almost the same as that of the exciter. The natural frequency in the vertical direction was found to be in the vicinity of 35 Hz, the (2½) × engine running speed component for a speed of around 850 rpm. Since the engine was expected to run at speeds from 800 to 2000 rpm, it was decided that the isolators used were not suitable and that it was imperative to look for isolators which would not result in a system resonance in the operating frequency range.

With newly designed isolators, the test was repeated in the frequency range 100–3000 rpm in steps of 100 rpm and the normalized velocities were calculated as before. From this test, the highest of the three translational natural frequencies of the system was found to be 9.5 Hz (or 570 rpm) as seen from a plot of normalized velocity versus speed shown in Fig. 14.9 and this was in the vertical direction. Since the ratio 35 Hz/9.5 Hz = 3.68 is much greater than 1, the isolators were found to be suitable.
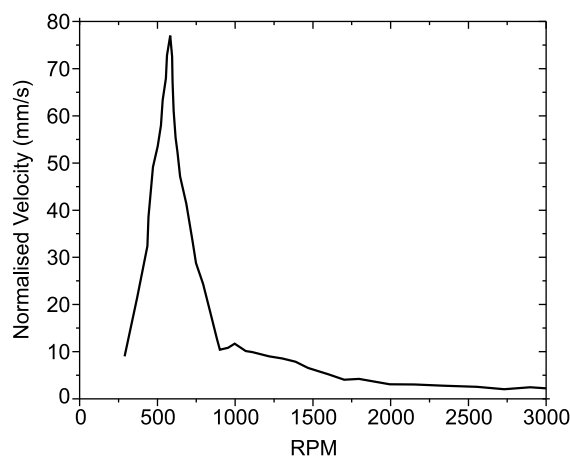
FIGURE 14.9 Normalized velocity versus speed

14.8 Heavy Vehicle Dynamics—Comparison of Ride Comfort Between Vehicle with Leaf Spring and Hydropneumatic Suspensions

This section discusses the results of experimental work carried out for the measurement of vibration response of heavy vehicles to evaluate and compare their ride behaviour with two different types of suspensions, namely the conventional leaf springs and hydropneumatic suspensions. The vertical dynamic response of these vehicles at various locations in the vehicle was found for three different types of terrain under various load and speed conditions. The vehicles studied were (i) a heavy vehicle with conventional leaf springs at the front and rear (ii) the same heavy vehicle with the rear leaf springs replaced by hydropneumatic suspensions. The roads considered for the study were three 1 km stretches of good, bad and paved herringbone roads which were characterized using a fifth wheel bump integrator. The speeds covered were 20, 30, 40 and 60 km/hr (kmph). The tests were carried out for no load, as well as for 1/3, 2/3 and full rated loads, and also for extra load on one side of the vehicle to simulate the weight of extra passengers on one side. For loading, sand bags of weight 400 N each were loaded on the floor of the vehicle. For assessment of ride quality, vibration measurements were made at a passenger's location and the driver's seat. Besides, measurements were also made on the axles, as well as a few points on the chassis longitudes directly above the front and rear suspensions on both sides of the vehicle.

14.8.1 Vehicle Response Measurements

Since the measurements due to road undulations were essentially in the low-frequency range of 0–50 Hz and since ride comfort criteria are generally specified in terms of acceleration, inductive accelerometers (B12/200 and B12/500, Hottinger Baldwin Messtechnik, West Germany) were rigidly fixed to the points of measurement by appropriate clamps. The signals from the transducers were fed to a six-channel carrier frequency amplifier (KWS 673–A2 with KWS 3073, Hottinger Baldwin Messtechnik, Germany) for signal conditioning. Conditioned signals from these channels were recorded on a cassette using an instrumentation tape recorder (V store, RACAL, England) so that detailed spectral and statistical analyses could be carried out later in the laboratory. They were also simultaneously fed to a digital storage oscilloscope (DSS 6520A, Kikusui, Japan) for monitoring the signals. Since all measurements were made in a running vehicle, two 12 V lead acid cells were used to power a sine wave inverter (TIS 24/500, Aplab, India), which in turn powered the entire instrumentation setup. Figure 14.10(a) shows a block diagram of the measurement setup and Fig. 14.10(b) shows the transducer placed on the chassis longitude directly above the hydrogas suspension.

The tape recorder was set for a speed of 15/16 in/s to give measurements in the frequency range of 0–625 Hz. Of the eight channels available for recording in the instrumentation tape recorder, channel one was set to record date and time. Another channel was used for flutter compensation to minimize the effect of tape speed variation due to vibration and noise. Tchebyshev anti-aliasing filters were used for all the channels. All readings were taken after the vehicle had reached steady state speed condition. The recorded vibration data in the form of analogue signals were played back from the instrumentation tape recorder and digitized using an analogue to digital converter interfaced to a personal computer. Data acquisition software with user friendly graphics user interface (GUI) written in Visual Basic language was used. The software had the capability to take multiple samples and average the data. Digitized time domain data were transformed to the frequency domain using programs written in MATLAB 5.1.

Programs were also written for the statistical analysis of the data and to plot spectrum and power spectral density curves. Figure 14.11 shows a block diagram of the instrumentation used for analysis of data.
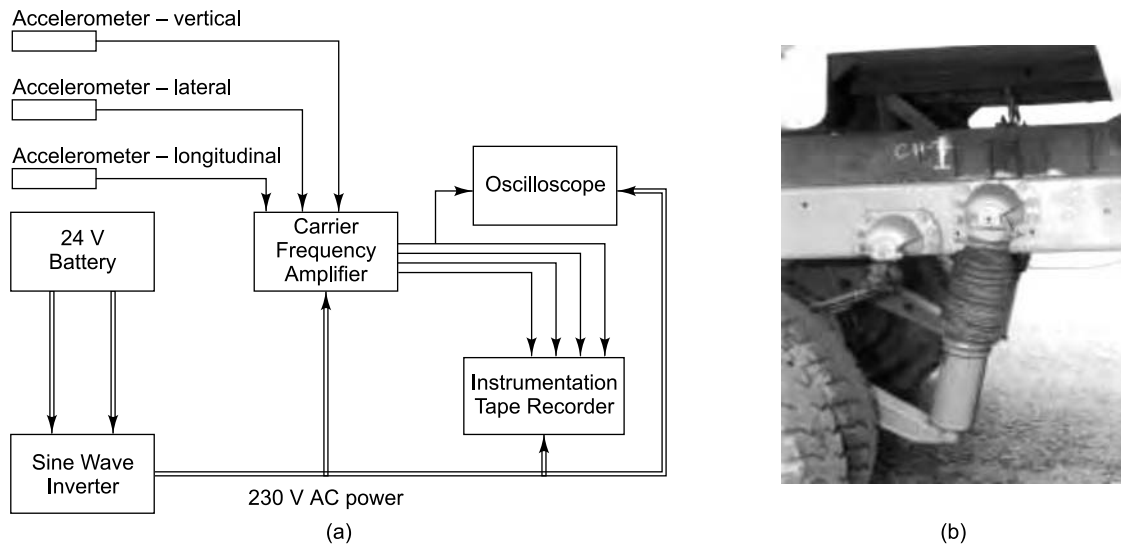
FIGURE 14.10   Vibration measurement setup: (a) block diagram, (b) photograph of transducer placed above suspension
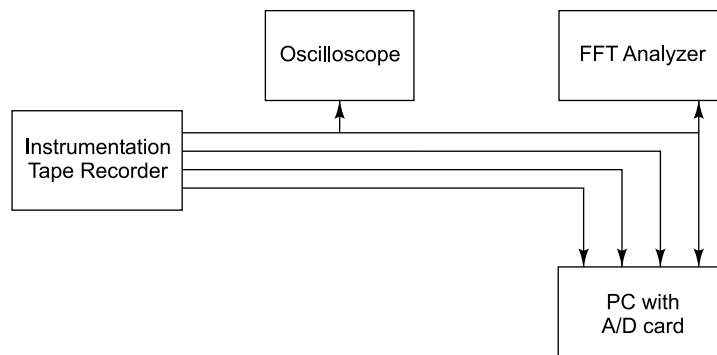
FIGURE 14.11   Instrumentation for data analysis

14.8.2   Road Undulation Measurements

The roughness of the roads traversed by these vehicles was also measured. Road undulation measurements were made on three representative 1 km stretches in and around Chennai, India; they are: (i) good road: premixed bituminous carpet road, (ii) bad road: a water bound macadam gravel road, and (iii) herringbone road: a road paved with granite blocks, typically used for off-road tracked vehicles.

The bump integrator or roughometer consists of a single-wheeled trailer with a hard rubber tyre mounted on a chassis and measures the relative vertical movement between the axle and chassis. The fifth wheel was towed by a van carrying the instruments required for obtaining road undulation records (Fig. 14.12). The operating speed of the machine was maintained at 30 ± 1/2 kmph. The data were

recorded in the form of electrical signals by mounting an inductive accelerometer on the arm carrying the scriber and responding to road undulations. Ground unevenness signals were recorded on the left- and right-hand sides, as well as at the centre of the test roads chosen and averaged. Table 14.4 shows the RMS acceleration values and mean roughness indices indicated by the roughometer for the various roads considered.

FIGURE 14.12   Road unevenness recorder

TABLE 14.4   Road characteristics

Road type	Surface type	RMS acceleration 'g'	Mean roughness index (mm/km)
Good	Premixed bituminous carpet	0.258	2810
Bad	Water bound macadam Gravel	0.399	5230
Herringbone	Granite blocks	0.640	6280

14.8.3   Test Results

Figure 14.13(a) shows a typical displacement power spectral density curve for the good road in terms of spatial frequency, computed knowing the temporal frequency and speed of the fifth wheel bump integrator (30 kmph). Figure 14.13(b) depicts a PSD measured at a location above the suspension for the vehicle with leaf springs. From such PSD plots, the RMS values of acceleration were obtained and these were plotted as a function of load for various road conditions and speeds. Figure 14.13(c) shows RMS acceleration versus load for different speeds for the good road for both suspension systems. These plots were also redrawn as RMS acceleration versus one-third octave centre frequency to study the ride behaviour of the vehicles (Fig. 14.13d). Graphs of RMS acceleration as a function of frequency for various roads, loads and speeds were obtained and compared with the ISO 2631 standards. The vibration levels at the driver's seat were compared with the 8 hours fatigue decreased proficiency boundary of ISO standards in the longitudinal direction a_z (buttocks to head of a seated passenger), while the response at a passenger's location was compared with the 1 hour reduced comfort boundary in the longitudinal direction, on the assumption that the journey lasts an hour. The major frequency components of vibration are in the range 4–8 Hz, which is the most uncomfortable range for a human being.
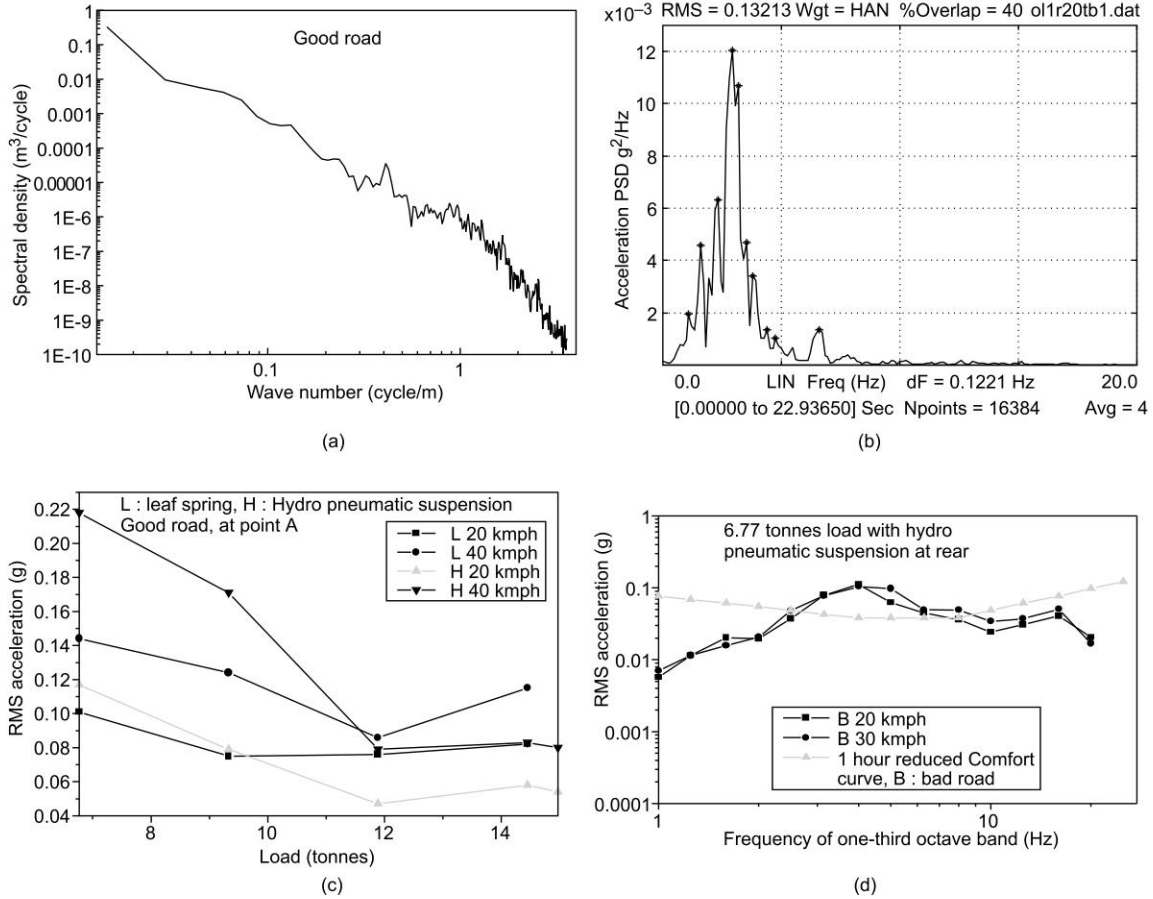
FIGURE 14.13 Typical responses: (a) spatial PSD of good road, (b) acceleration PSD of vehicle with leaf spring, (c) comparison of suspensions for good road, (d) 1/3 octave spectra of vehicle with hydropneumatic suspension

14.9 DYNAMIC STUDIES OF RAILWAYS BOGIES

In this section are described the experiments conducted on Alternating Current/Electrical Multiple Unit/Trailer (AC/EMU/T) coaches of Southern Railways, India running on broad gauge suburban tracks and the subsequent analysis of the measured response. The experiments were conducted over a stretch of approximately 30 km around the metro of Chennai, India. The objectives of the experimental investigation were to:

 (i) Carry out extensive measurement of acceleration and displacement levels on the bogie and car body of a railway coach running on typical suburban tracks under various running conditions (speed and load) to assess the ride quality of the vehicle.
 (ii) Study the pitch, roll and yaw behaviour of the bogie and car body.

   (iii) Measure the response of the rail, in terms of acceleration as well as strain, due to moving railroad vehicles.

The instrumentation used was very similar to that described in Section 14.8. All experiments were conducted during off-peak hours and all readings were taken after the train had attained steady state speed condition. The average speed at which the measurements were made was about 45 kmph. The coaches were loaded with approximately 20 passengers and 200 kg of instrumentation. Acceleration and displacement levels on the underframe or floor of the coach were measured. For acceleration measurements, inductive accelerometers as described earlier were used. MEMS based uniaxial accelerometers (Type-SAA1102, Make-NeuwGhent Technology, USA) were also used since these accelerometers also are meant for low frequencies. For displacement measurements on the coach floor, linear variable differential transformer (LVDT) type displacement pickups (B2/25, Hottinger Baldwin Messtechnik, West Germany) were used. Besides, a triaxial MEMS based rate gyroscope with a dynamic range of ± 150 º/s (Type TAG–31150, Make-NeuwGhent Technology, USA) was also used for the measurement of the angular rates of the bogie and car body about the three axes of motion (pitch, roll and yaw). The MEMs accelerometers and gyros were conveniently powered using a 5 V DC source.

For studies on the track, acceleration response measurement (using inductive accelerometers) on the rails of surface tracks to moving rail vehicles was carried out when express trains passed at an approximate speed of 75 kmph. Besides, 5 mm linear strain gauges (Make Rohit, India) were also bonded to the rail to measure the stress that the rail is subjected to due to a passing train. The outputs of the three inductive accelerometers mounted in three mutually perpendicular directions (longitudinal, lateral and vertical) and that of the LVDT were fed to a six-channel carrier frequency amplifier. The strain gauge outputs were also fed to carrier frequency amplifiers (MGC Plus, HBM, West Germany). Conditioned signals from these amplifiers were fed to different channels of the instrumentation tape recorder and simultaneously to a digital storage oscilloscope. A PC-based data acquisition system with CATMAN software was used in conjunction with MGC plus for recording the outputs of the strain gauges. The outputs of the MEMs based devices were fed to FFT analyser (Type CF250, Onosokki, Japan) and a data recorder (Type 8808-01, Hioki, Japan). Figure 14.14(a) shows instrumentation involving MEMs-based transducers. Figure 14.14(b) shows the complete instrumentation setup for subsequent data analysis in the laboratory.

The tape-recorded vibration data were fed to an FFT analyser (SD 380, Spectral Dynamics, USA) for further analysis. Experimental results have been presented as amplitude vs. frequency (spectrum) plots and PSD plots. Acceleration has been plotted in g and displacement in mm, acceleration PSD in g^2/Hz and displacement PSD in m^2/Hz. For the calculation of RMS values, the frequency range 0–25 Hz was used. The time domain plot of a typical LVDT reading is shown in Fig. 4.15(a). The response spectrum of acceleration in the lateral direction of a point on the bogie frame is shown in Fig. 4.15(b). Figure 14.15(c) shows a typical vertical acceleration PSD of a point on the coach floor and Fig. 4.15(d) a typical vertical acceleration spectrum of a point on the rail.
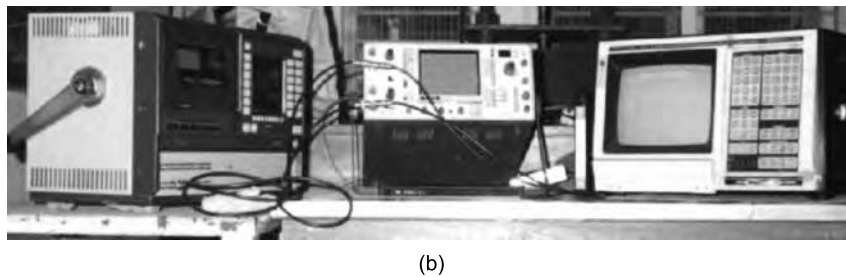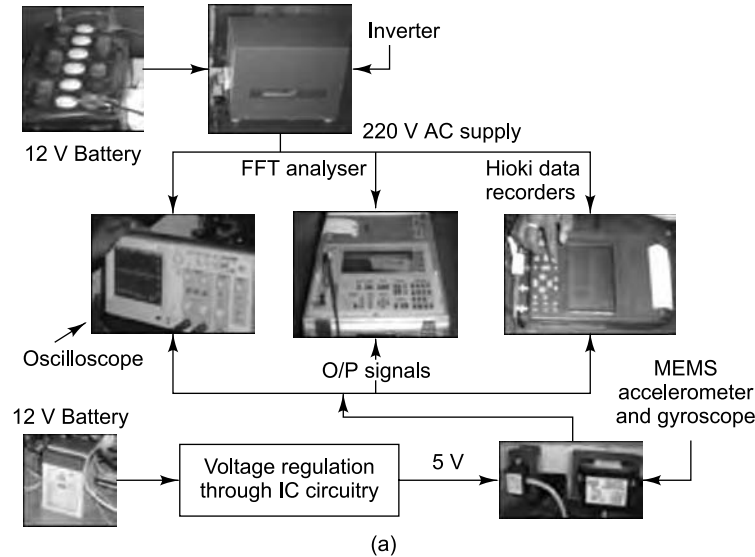
FIGURE 14.14   Instrumentation setup for rail vehicle dynamics: (a) vibration measurement: MEMS transducer-based instrumentation setup, (b) instrumentation setup for data analysis

14.10   VIBRATION AND RIDE COMFORT STUDIES ON A TRACKED VEHICLE

Field tests conducted on a military tracked vehicle to assess the vibration levels on the operating crew are discussed in this section. Such vehicles experience a very severe vibratory environment as operational requirements necessitate their use over rough terrain and for prolonged periods. This vibration is further complicated by multiple track link-terrain interaction, high power-to-weight ratio and high mobility. The vibrations experienced in military tracked vehicles are predominantly of low-frequency and high-amplitude and have a detrimental effect on the crew. The accelerations experienced at specific locations on the vehicle for four different types of terrain were recorded and analysed in spectral and statistical modes. From the point of view of crew comfort, acceleration levels encountered in the crew cabin have been compared with those specified in ISO 2631 standard. Ride quality assessment has also been made on the basis of absorbed power.
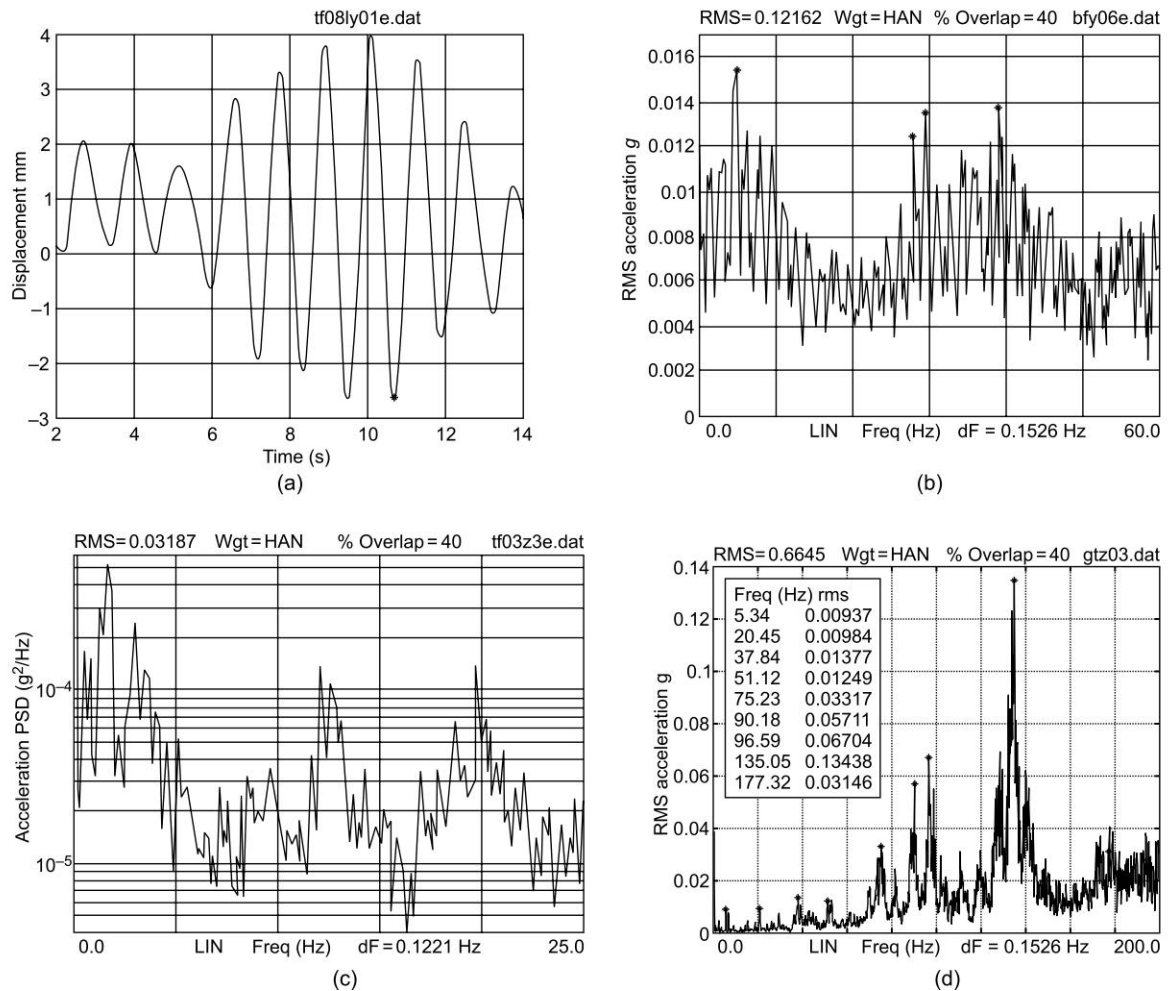
FIGURE 14.15 Typical responses: (a) time domain plot from LVDT (b) lateral acceleration spectrum on bogie frame, (c) vertical acceleration PSD on coach floor, (d) vertical acceleration spectrum on rail

The vehicle used for the field test was a tracked vehicle with seven bogie stations on each side and with torsion-bar suspension systems. The off-road terrain profiles selected for the field tests were: (i) a concrete track of length 1 km, (ii) a sinusoidal track of wavelength 7 m and amplitude 100 mm peak stretching up to 200 m, (iii) a random cross-country course and (iv) a staggered obstacle course consisting of half round obstacles equally spaced on a concrete track. The average vehicle speeds for the concrete, sinusoidal, cross-country and staggered obstacle tracks were 35, 15, 35 and 15 kmph, respectively. All tests were conducted in the unladen condition of the vehicle. The instrumentation and test procedure were identical to those outlined in Section 14.8. All instruments were secured tightly inside the vehicle with sufficient cushioning to prevent vibration. Considerable effort was taken to properly route all the cables connecting accelerometers, carrier frequency amplifiers and instrumentation tape recorder. The signals were subsequently analysed in the laboratory in the spectral and statistical modes using a spectrum analyser. Table 14.5 shows the RMS and peak values of measured vibrations at various locations. Figure 14.16(a)

shows a typical time history recorded at the rear seat of the vehicle on a cross-country run. Figure 14.16(b) shows the transfer function from the torsion bar to the driver's seat for the same run. Figure 14.16(c) shows the probability density function of acceleration at the driver's seat for a concrete track run.

TABLE 14.5   Summary of vibration levels measured on the test tank (*g*)

Sl. No.	Location	Frequency range (Hz)	Concrete		Sine		Cross-country		Staggered obstacle	
			Pk	RMS	Pk	RMS	Pk	RMS	Pk	RMS
1	Driver's seat (Vert)	200	3.37	0.876	1.12	0.246	2.08	0.45	1.76	0.422
2	Driver's seat (Long)	100	1.12	0.226	0.56	0.113	0.72	0.115	0.88	0.138
3	Rear seat (Long)	100	1.44	0.291	0.72	0.137	0.71	0.153	0.72	0.184
4	Rear seat (Vert)	500	–	–	–	–	12.40	2.54	8.02	2.10
5	Rear seat (Lat)	100	2.4	0.316	2.09	0.204	2.73	0.334	1.44	0.194

Note: Vert: vertical, Long: longitudinal, Lat: lateral.

(a)

(b)

(c)
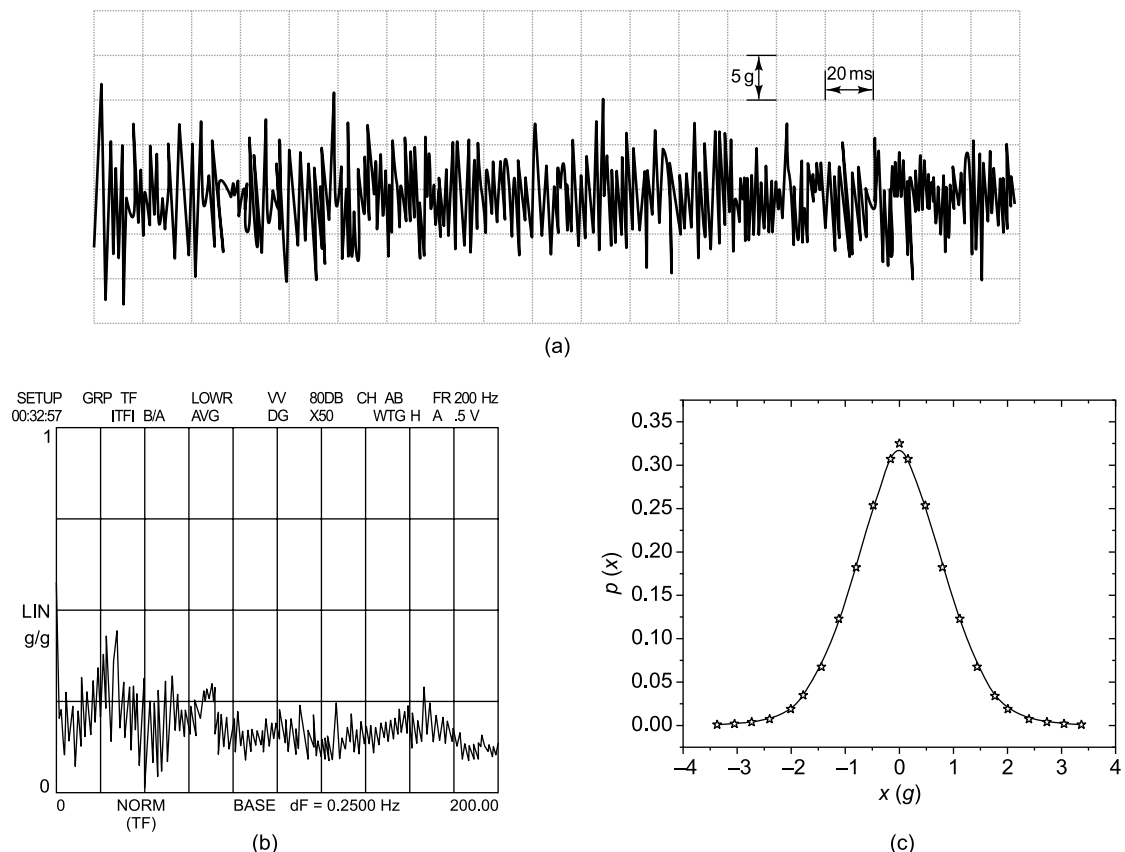
FIGURE 14.16   Tracked vehicle response: (a) time history of acceleration at rear seat (vert.) cross-country, (b) transfer function at driver's seat (vertical), (c) probability density histogram of acceleration at driver's seat-concrete track

A different approach to analysing the ride quality and comfort has also been done using absorbed power as described in Section 9.3.3.3 using Eqs (9.4) and (9.5). Values of K_i for vertical response as a function of frequency have been interpolated for the required frequencies with $\Delta f = 0.625$ Hz and based on these values of K_i, the absorbed power at the rear seat for a cross-country run has been computed and is found to be 66.31 W. This is much higher than the acceptable limit of 6–10 W for off-road vehicles. Table 14.6 shows the computation of absorbed power.

TABLE 14.6   Absorbed power calculations

S. No.	f (Hz)	K_i	$A_{i,\text{rms}}^2 [(ft/s^2)^2]$	$K_i A_{i,\text{rms}}^2$	$K_i A_{i,\text{rms}}^2 \Delta f$
1	0.625	0.0040265	969.3	3.9	2.4
2	1.25	0.015101	944.4	14.3	8.9
3	1.875	0.037814	208.15	7.8	4.87
4	2.5	0.077843	121.16	9.4	5.87
5	3.125	0.127606	184.3	23.5	14.7
6	3.75	0.174830	68.24	11.9	7.4
7	4.375	0.2010225	30.75	6.2	3.87
8	5.0	0.198689	68.35	13.6	8.5
9	5.65	0.17282	38.3	6.6	4.1
10	6.25	0.134302	10.66	1.43	0.89
11	6.875	0.10045	30.6	3.07	1.9
12	7.5	0.075581	27.96	2.1	1.3
13	8.125	0.06107	7.96	0.48	0.3
14	8.75	0.0537	14.5	0.78	0.48
15	9.375	0.0505	20.09	1.01	0.63
16	10.0	0.048340	7.26	0.35	0.2

14.11   EXPERIMENTAL MODAL ANALYSIS

In this section, experimental modal analysis of two very simple structures, namely a beam and plate are described. The specimens used were (i) a fibre reinforced plastic beam of size 210 mm × 25.4 mm × 4.5 mm and (ii) an aluminium plate of size 500 mm × 500 mm × 2 mm. The following sections describe the procedure adopted for obtaining the modal parameters, more importantly, the mode shapes.

14.11.1   Modal Analysis of Beam

To determine the modal parameters of the beam, the experimental setup shown in Fig.14.17 was used. The specimen was clamped rigidly between metal blocks using closely spaced bolts, such that the length of the cantilever beam was exactly 210 mm. The bolts were tightened to a pre-determined torque of 53 Nm to ensure that there was repeatability every time the specimen was unclamped and clamped again.
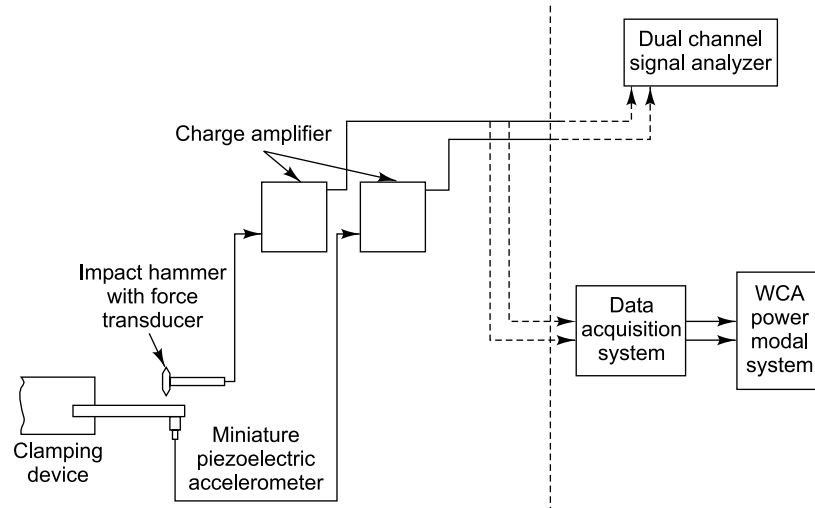
FIGURE 14.17   Test setup for modal analysis

It was decided to use roving impact excitation with a miniature piezoelectric accelerometer for response measurement, keeping the location of the accelerometer fixed. This decision was made based on the fact that the weight of the specimen was quite low, around 55 gm. For such cases, roving impact is found to be better than roving response measurement for obtaining the best FRFs; the latter results in additional (and variable) mass loading effect on the structure requiring corrective action while extracting modal parameters from the FRFs.

Figure 14.18 shows the details of the excitation and response measurement locations; these were marked on the specimen using a marker pen (shown as points 1 to 14 in the figure). The best location for the transducer was found to be point 9 for modes 1 to 4. This location of the accelerometer was selected from among the other locations because it gave sufficient response for all modes, the FRFs were repeatable and all the values of coherence functions were close to 1. The miniature accelerometer (Type 4344, Brüel & Kjær, Denmark) was attached using bee's wax. An impact hammer (Type PH51 with built-in force transducer Type 7117, Rion, Japan) was employed to impact the beam, transverse to its axis. The hammer was used with a plastic tip which allowed the first few modes of the beam to be excited in the frequency range 0 to 2 kHz.
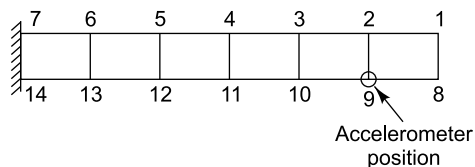
FIGURE 14.18   Excitation and response locations

The signals, both of impact force and acceleration response were fed through charge amplifiers and a data acquisition system to the WCA Power modal system to capture the mode shapes. This system is a Macintosh - based analyzer with the capability to acquire data on up to 32 parallel channels and has a

modal analysis program. The software accurately computes the damped natural frequency, damping ratio and mode shape for each mode. While capturing the data before doing the modal analysis with the WCA, a rectangular weighting function was applied to the impact hammer output and an exponential weighting function to the accelerometer response function. To obtain good results, the specimen was excited at each of the 14 points as shown in Fig.14.18. At each location, a minimum of 6 averages was used to obtain an averaged FRF, the estimator $H_1(f) = \dfrac{S_{xy}}{S_{xx}}$ being used for obtaining the FRFs. For each averaged FRF the coherence function was monitored. The averaged FRF was accepted if the coherence function remained over 90% during the averaging process over the entire frequency range of interest. The averaged FRFs were used to determine the natural frequencies, damping ratios and mode shapes of the first few modes using an MDOF complex exponential curve fitting procedure. Figure 14.19 shows the mode shapes of the first four modes of the beam thus obtained through experimental modal analysis.

Mode 1 (60.15 Hz)

Mode 2 (394.91 Hz)

Mode 3 (479.27 Hz)

Mode 4 (1091.25 Hz)
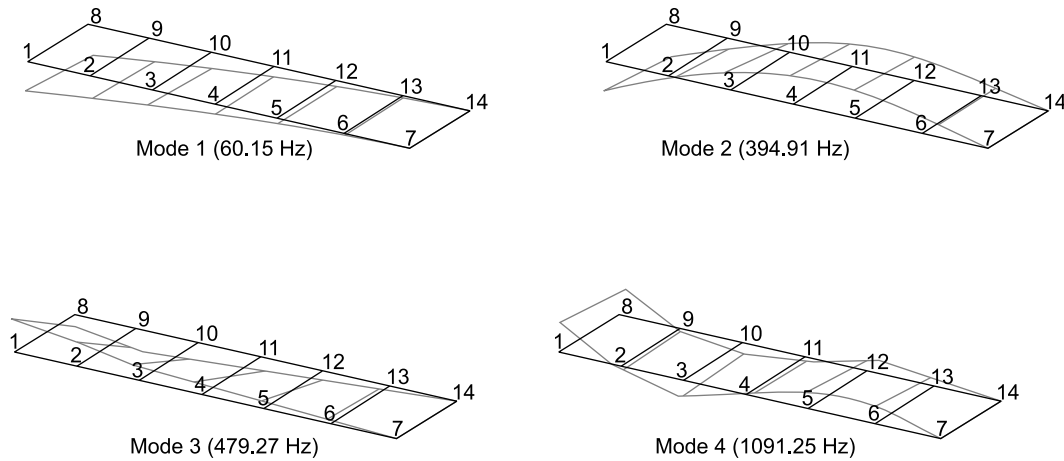
FIGURE 14.19   Mode shapes of beam

14.11.2   Modal Analysis of Plate

Modal analysis was also conducted on an aluminium plate of dimensions 500 mm × 500 mm × 2 mm. The aim of the exercise was to obtain natural frequencies, damping ratios and mode shapes as before. Prior to obtaining the mode shapes through experimental modal analysis, a simple test was conducted to visualize the mode shapes and this has been described in this section. The plate was subjected to sinusoidal point excitation using an electrodynamic exciter (Type PR9271, Philips, Germany). The point of excitation was chosen such that it did not coincide with a nodal point of the first seven modes of vibration. The excitation signal, i.e. the sine wave was fed from the sine wave generator with a frequency resolution of 0.001 Hz. The response of the plate was picked up by a non-contact capacitance pickup (Eichhorn and Hausmann OHG, W. Germany). The excitation signal and the response signal after signal conditioning using displacement measuring unit (Type MW 210-13-3, Eichhorn and Hausmann OHG, W. Germany) were fed to a spectrum analyzer (Type SD 380, Spectral Dynamics, USA). The advantage of the capacitance transducer is that it is non-contact in nature and does not alter the plate dynamics. Even a miniature piezoelectric accelerometer of weight 2 gm was found to change the natural frequencies by 2 – 3 Hz and

hence the need for a non-intrusive type of pickup. To visualize the mode shapes, fine sand was sprinkled over the surface of the plate. The plate was excited at each of the first seven natural frequencies and some of the patterns formed on the surface are shown in Fig. 14.20. Other details regarding the modal analysis are not shown.
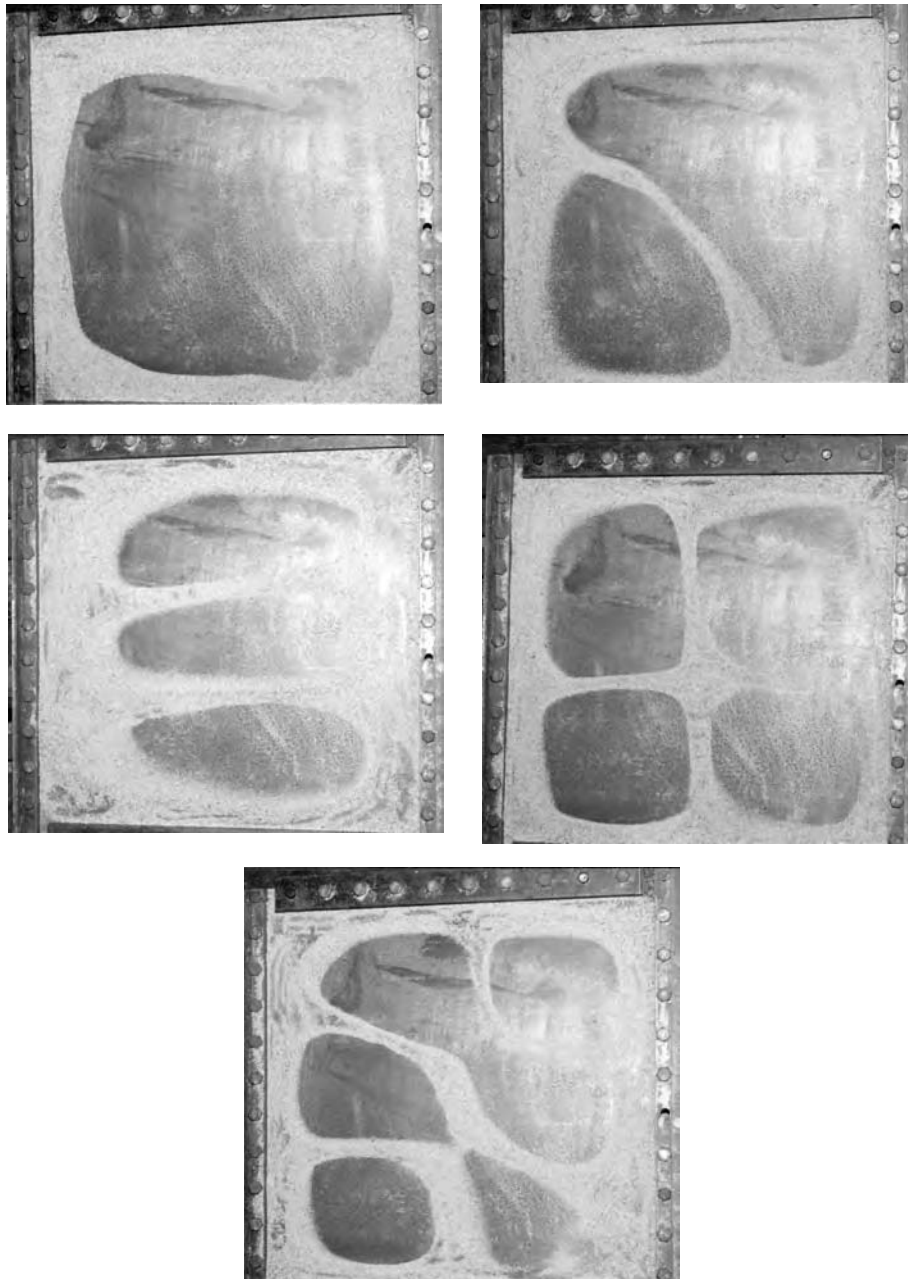
FIGURE 14.20   Mode shapes of plate

14.12   Vibration Measurement on Turbine Blades in a Power Station

In one of the major power plants in India there was an unfortunate accident involving the flying off of one of the turbine blades, leading to a fire accident and havoc. This incident instigated the study of the natural frequencies of such blades during stationary conditions through experimental work and during rotation from a finite element analysis in order to determine 'rogue blades', if any. The study involved observing the natural frequencies of all blades of the low pressure (L.P) stages: L.P.-5 and L.P-4 and also the packet to packet variation in the high-pressure H.P.-3 stage. It was also decided to determine the mode shapes of predominant modes, i.e. axial, tangential, etc. The whole exercise had to be carried out with simple instruments that could be carried to the field, and which could still give reliable results within a short period of time. Free vibration tests were carried out to determine the natural frequencies. Since it was known from finite element analysis that the natural frequencies were in the range of 100s of Hz, a piezoelectric accelerometer (Type 4332, Brüel & Kjær, Denmark), with a magnetic base was decided on for the measurements. The output of the accelerometer was conditioned using a charge amplifier (Type 2626, Brüel & Kjær, Denmark). The blade was given impact excitation with an instrumented hammer (Type PH 71, Rion, Japan). A rubber tip was used for exciting the lower order modes and a steel tip for the higher order modes. The acceleration response from the accelerometer and the force output from the instrumented hammer were fed to two channels of a signal analyser (Type SA 73, Rion, Japan) to get the transfer function between the force and the acceleration. The most predominant peaks in the transfer function, corresponding to the natural frequencies were read off using a cursor and were recorded. This exercise was done for all blades of the LP stage and for only one centrally located blade of each packet of the high-pressure stage to determine packet to packet variation in natural frequencies. Figure 14.21 shows a typical response obtained for L.P.–5 blade.
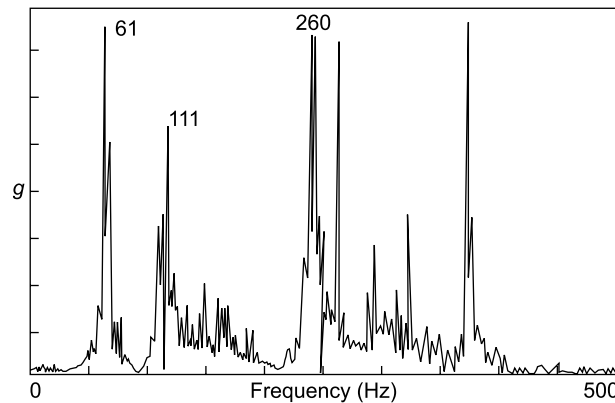
Figure 14.21   Response of L.P.–5 blade

14.12.1   Mode Shape Determination

Efforts were made to determine the mode shapes of the predominant modes for the H.P.-3 blade. Since the frequencies due to the blade alone are very high, it was deduced that the lower order modes observed in the free vibration spectrum were contributed to by the disc. The disc of the high-pressure stage was subjected to forced vibration by exciting it axially through the shroud using an electrodynamic exciter

(Type PR 9270/10, Philips, Germany). Excitation was given at each of the predominant natural frequencies determined from the free vibration test, and at each frequency the disc mode shape was determined. This was done by placing one reference piezoaccelerometer on the disc and by moving a second transducer along the circumference of the disc to determine the phase of vibration. The number of nodal diameters was ascertained from the plots, indicating the number of phase reversals around the circumference of the disk in the output of the roving accelerometer. Since the excitation signal was very weak, the signal to noise ratio was not very good, making the identification of phase reversals difficult, especially at higher frequencies. Figure 14.22 shows phase changes observed as a function of blade number (over a sector of the disc) along the circumference for a mode at 2050 Hz.
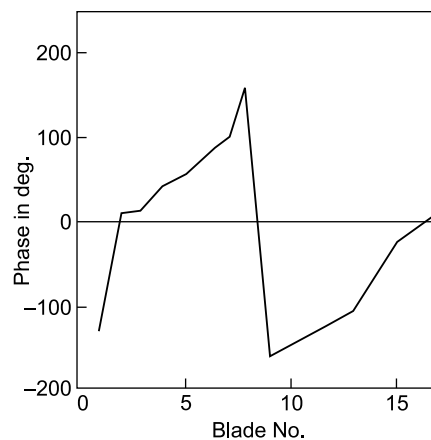
Figure 14.22 Phase plot: 2050 Hz

14.13 Study of Sound from a 106 MW Power Plant

The aim of the study was to determine the sound pressure levels (SPLs) in the houses of a village in the vicinity of a private power corporation generating 106 MW with seven diesel generators, each of capacity 15.143 MW. About 50 m from the plant wall commences the village area from where there were complaints of noise. The idea behind the study was to measure the noise in the village and compare it with national and international standards to check whether it was within accepted norms.

Noise measurements were made within the plant with all seven diesel-generator units running at the rated speed of 480 rpm under full-load condition. Overall sound pressure levels (SPLs) were measured in terms of dB (LIN) and dB (*A*). Spectrum analysis was also done in the A-weighted mode using the 1/1 octave filters with centre frequencies from 8 Hz to 16 kHz to identify the predominant noise frequencies. All these measurements were made at each location using a sound level meter (CEL 393, England) with a ½ in. microphone (CEL 186/3F England). Figure 14.23(a) indicates the SPLs at three locations inside the plant: in the diesel generator hall, immediately outside the hall and about 15 m away from the hall.

Sound level measurements were also made at various locations in the residential area during daytime (between 11 A.M. and 12 noon), during which period there was noise emanating from the factory as well as the school and the residences in the village area. Measurements were also made at night (after 10 P.M. when all other sources of sound had died down). Figure 14.23(b) indicates the SPLs at night time at three residences in the village area outside the plant. The residences have been indicated as Locations 1–3.
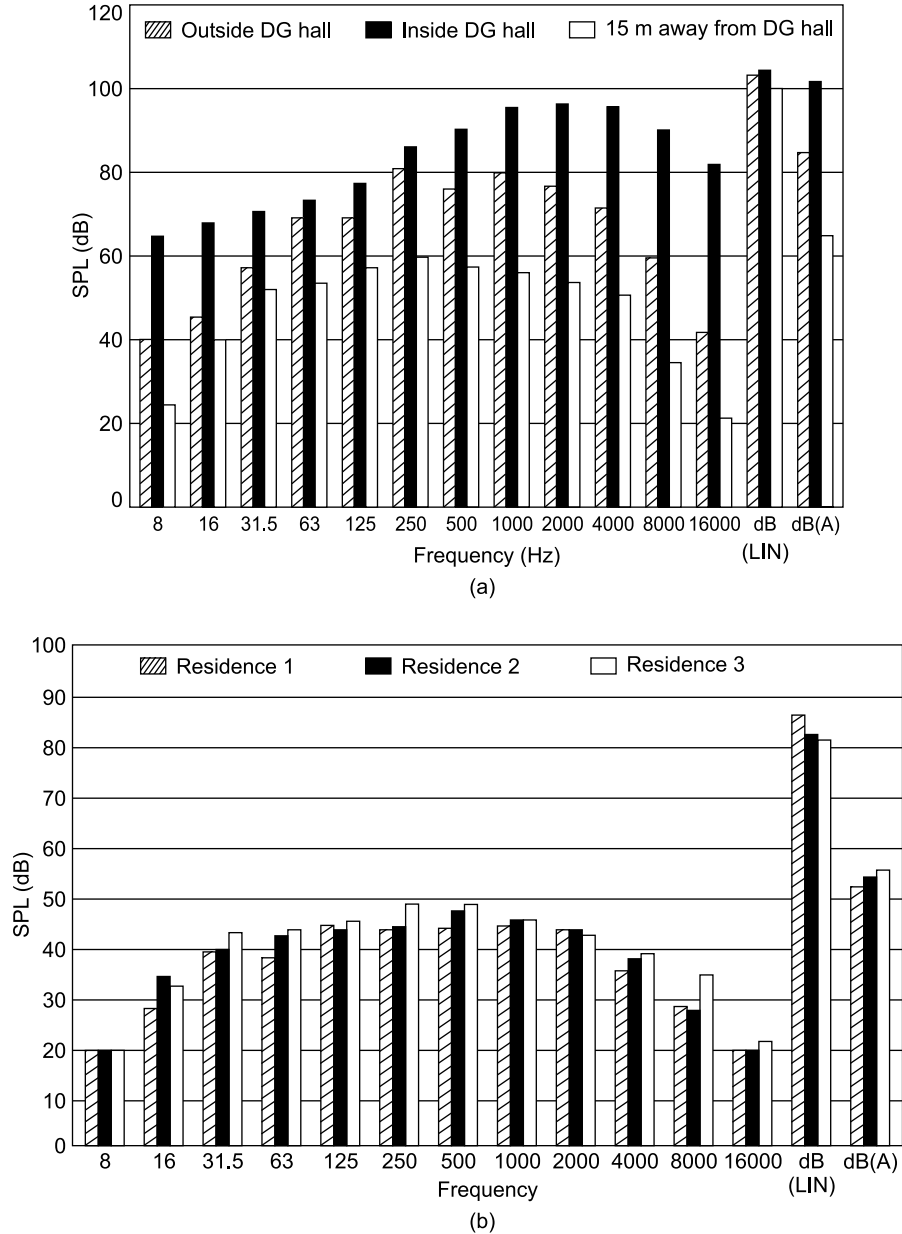
FIGURE 14.23 Typical SPLs: (a) SPLs in plant, (b) SPLs at night in residential area

The noise studies conducted in the village area (which is a declared industrial zone) show overall values ranging from 54 to 63 dB (*A*) during the day and values ranging from 44 to 57 dB (*A*) during night. The allowable values as per national and international standards have been shown in Tables 13.9 and 13.10. The SPLs measured are well within the norms specified by Pollution Control Board of India [75 dB (A) at day time and 70 dB (A) at night time] and within the norms specified by many other countries also.

14.14   Evaluation of Rubber Impregnated Steel Roller Chains for Noise Reduction and Vibration Damping Characteristics

A study was conducted to evaluate and compare the noise and vibration damping characteristics of nitrile rubber impregnated steel roller chains with conventional steel roller chains used for conveyor applications. Two sets of chains each, of rubber impregnated type and conventional steel type were used, having 2". pitch with oversized rollers and moving on steel tracks on a test rig especially fabricated for this purpose.

14.14.1   Noise Measurements

It was decided to compare noise based on sound power levels by measuring sound pressure levels over a hemispherical surface as described in Section 13.2.1. Due to non-availability of an anechoic chamber, the chain test rig was set up in an open field where a clear, open flat surface without obstructions for a radius of about 20 m was available. No trees, poles, buildings or bushes were present in this 20 m radius. The test rig and motor were fixed on a foundation at the centre of this field. The ground in the vicinity of the test rig was flattened to ensure a proper reflecting plane. A schematic of the test setup is shown in Fig. 14.24.
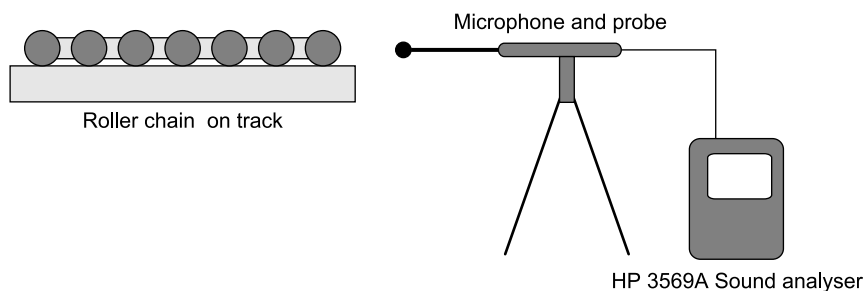
Figure 14.24   Test setup for chain noise measurements

The measurement surface was taken as a hemisphere of radius, r, centred at $x = 0$, $y = 0$, $z = 0$. The dimensions of the test rig were 2 m × 1 m × 1 m, with an additional 0.5 m space in the lengthwise direction being taken by the motor. Hence the characteristic dimension D_o or the equivalent diameter D_e of the source was taken as 2.5 m. Since the radius of the hemispherical measurement surface should be at least three times the equivalent diameter, it was taken as 7.5 m. In order to obtain the average value of the mean square pressure on the surface of the test hemisphere, an array of fixed microphone positions distributed over the hemisphere was used in accordance with Section 8.5.2.1 of British Standard BS 848: Part 2: 1985. The coordinates of the points are shown in Table 14.7 and the points are depicted graphically in Fig. 14.25. The locations of the measurement points were marked on the ground. The heights in the Z-direction were marked on a marker pole, the pole being slim enough as not to cause any appreciable disturbance in the free-field environment. Measurements were made with a real time frequency analyser (HP 3569A, Hewlett Packard, U.S.A.) using a ½ in. condenser microphone (HP 35237 A) attached to a probe (HP 35230 A).
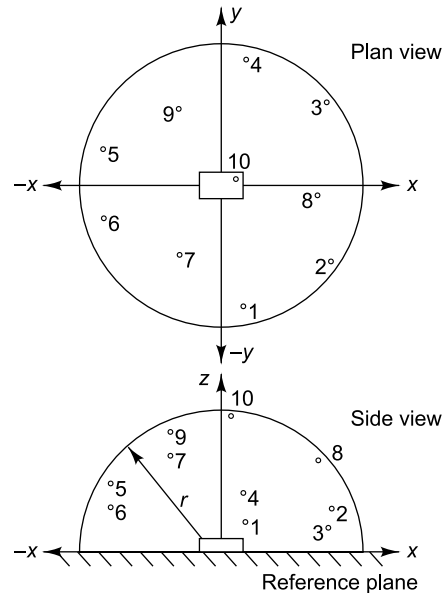
FIGURE 14.25   Location of points on the hemisphere for free-field measurements

TABLE 14.7   Coordinates for microphone locations for hemispherical measurement surface

Position	X-coordinate		Y-coordinate		Z-coordinate		Surface area	
i	x/r	$x(m)$	y/r	$y(m)$	z/r	$z(m)$	S_i	$S_i\ (m^2)$
1	0.16	1.2	−0.96	−7.2	0.22	1.65	$0.2\pi r^2$	35.343
2	0.78	5.85	−0.60	−4.5	0.20	1.5	$0.2\pi r^2$	35.343
3	0.78	5.85	0.55	4.125	0.31	2.325	$0.2\pi r^2$	35.343
4	0.16	1.2	0.9	6.75	0.41	3.075	$0.2\pi r^2$	35.343
5	−0.83	−6.225	0.32	2.4	0.45	3.375	$0.2\pi r^2$	35.343
6	−0.83	−6.225	−0.40	−3.0	0.38	2.85	$0.2\pi r^2$	35.343
7	−0.26	−1.95	−0.65	−4.875	0.71	5.325	$0.2\pi r^2$	35.343
8	0.74	5.55	−0.07	−0.525	0.67	5.025	$0.2\pi r^2$	35.343
9	0.26	1.95	0.50	3.75	0.83	6.225	$0.2\pi r^2$	35.343
10	0.10	0.75	0.10	0.75	0.99	7.425	$0.2\pi r^2$	35.343

Note: r is radius of hemispherical surface.

Sound pressure levels were measured in 1/3 octave bands, on the *A*-weighted scale, from 3.15 Hz to 10 kHz at the 10 points on the hemisphere in the free field condition under various loads and speeds for various chain and track conditions. Tests were conducted for the regular and rubberized chains. Measurements were also made with and without tracks to investigate the major contribution of noise. Sound pressure levels were recorded at each position for all the above conditions and sound power calculated. The sound power levels obtained for the regular and rubberized chains are shown in Fig. 14.26.
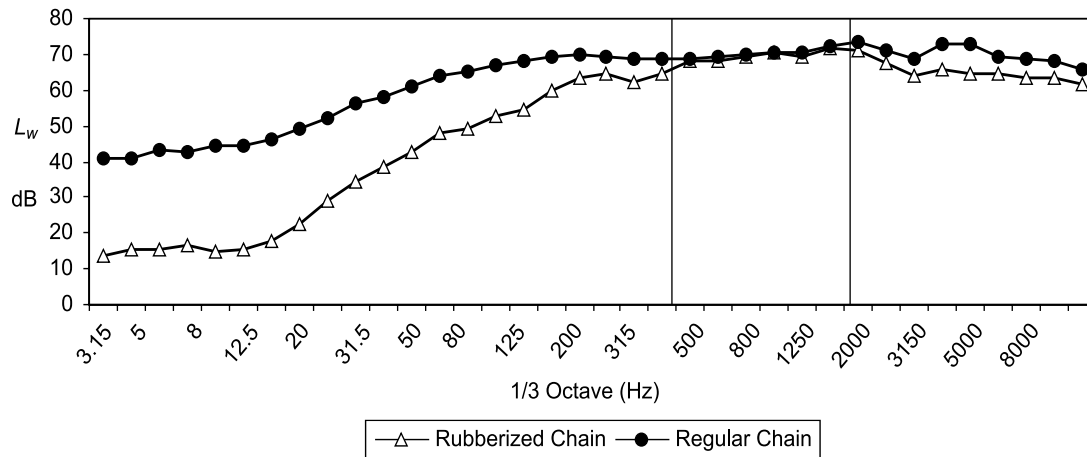
FIGURE 14.26   Comparison of sound power levels—100 rpm, no load, with track

Comparison of overall sound pressure levels for all conditions of load, speed and track showed that the rubberized chain was quieter in general by up to 5.14 dB. Since a reduction of a minimum 6 dB is required for the human ear to perceive the drop in noise, it was concluded that the rubberized chain would not sound quieter than the regular chain for the speeds considered in the study; however the high frequency components (3150 Hz to 10 kHz) in the noise would sound less sharp in the case of the rubberized chain. It was concluded that for high speed applications, the rubberized chain would offer substantial reduction in noise.

14.14.2   Vibration Measurements

Vibration tests were conducted in the laboratory on the chain test rig (Fig. 14.27a), with the aim of comparing the vibration spectra of the rubberized and regular chains to ascertain if the former showed any reduction in vibration amplitude. Vibration power spectra in the 0–10 kHz range and overall g levels of the rubberized and conventional roller chains at 100 and 150 rpm were recorded at various points on the test rig using a piezoelectric accelerometer (CEL 192/2, England) along with a charge amplifier (Type 2626, Brüel & Kjær, Denmark). A dual channel signal analyser was used for analysis. Vibration power spectra in g RMS, for the rubberized chain at 100 and 150 rpm were compared with the corresponding vibration spectra recorded for the regular chain. This comparison was made at 6 locations, four of which were on tracks and two on bearing housings, in all three directions at each point. Figure 14.27(b) shows a comparison of the power spectra for regular and rubberized chains at 100 rpm on the track. It can be seen from this figure that the rubberized chain has lower vibration levels at low frequencies, below 300 Hz, and also at high frequencies, above 2000 Hz. In the mid-frequency range, the two chains have similar g levels, with the rubberized chain displaying considerably lower overall levels of vibration. These results correlate with the comparison of sound power between regular and rubberized chains discussed earlier.
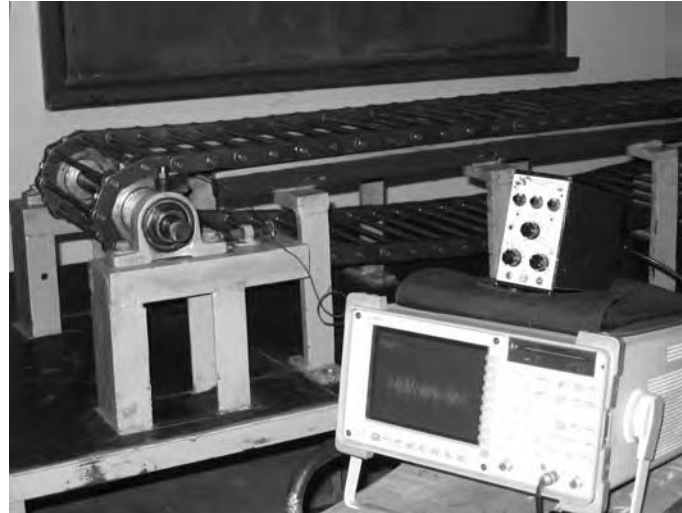
(a)
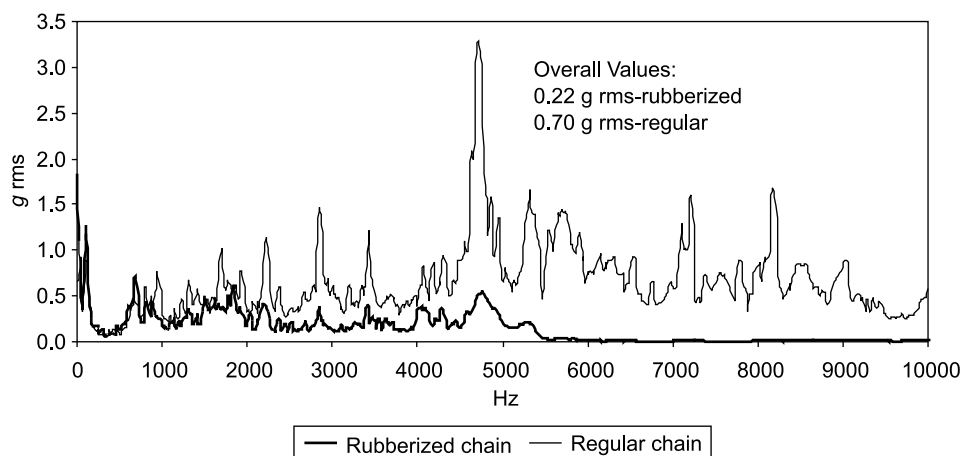
(b)

Figure 14.27 Vibration measurements on chain: (a) photograph of test setup for vibration measurements, (b) comparison of vibration spectra, 100 rpm at a point on track (along movement of chain)

The cross-spectrum between noise and vibration was also measured to find the vibration frequencies resulting in noise. The test setup used was similar to that for vibration measurements alone, except that the dual channel dynamic signal analyser received vibration signals on one channel and noise signals on the other channel. Plots of cross-spectrum between noise and vibration in units of Pa × *g* RMS were obtained for various points on the track, bearing and test table, showing sharp peaks at frequencies common to the noise and vibration signals. These frequencies are essentially those due to roller rotation, meshing impact, their harmonics, or natural frequencies of structural components, including rollers and chain links. Figure 14.28 shows a representative cross-spectrum plot at 150 rpm for the rubberized chain.
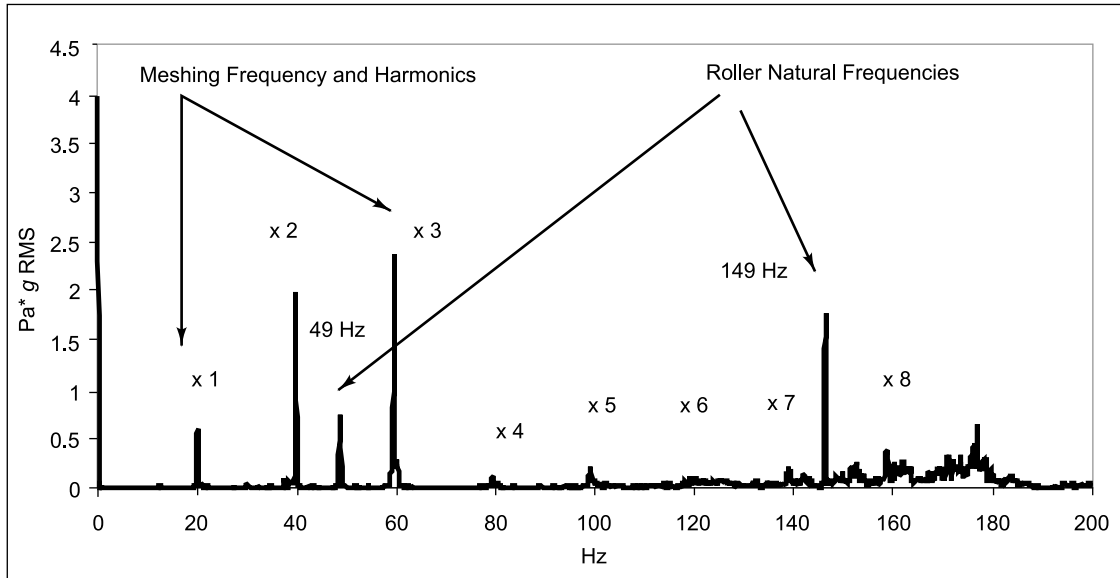
FIGURE 14.28   Cross-spectrum, 150 rpm, NDE Bearing, 0–200 Hz

From the cross-spectrum analysis, it is seen that the radiated noise consists of forcing frequencies, i.e. meshing and roller frequencies, and their harmonics, as well as the natural frequencies of the structural elements excited by these forcing frequencies. These natural frequencies include roller-chain link natural frequencies, bearing race and ball resonance frequencies, as well as those of the supports and table. It is evident from the cross-spectrum plots that the major contribution to noise comes from the lower frequencies. This is expected, as the meshing at the sprockets is considered to be the major source of noise in chains. The roller rotational frequency and its harmonics are the main source of noise as the chain moves over the tracks. Roller natural frequencies are also excited by the meshing impact and contribute to the generated noise. In addition, resonant frequencies of track, supports, bearing housing, shafts, etc. also create noise, but the contributions are much less in magnitude.

14.15   SOUND INTENSITY MEASUREMENT ON HELICOPTER

The study of sound emanating from a helicopter was conducted in order to make objective physical measurements of the sound produced inside the driver's cabin and outside and thereby pinpoint the sources of sound. The aircraft had an engine with a single stage axial compressor, followed by a single stage radial compressor. It had a three stage axial turbine. The engine, the reduction gearbox, clutch unit and main gearbox were mounted on a tubular frame and were located at the centre of the helicopter. For describing the noise characteristics of the helicopter, sound power level data were used since they are independent of the distance between the source and the observer, as well as of the environment in which the measurements are made. For this purpose, intensity measurements were made so that sound power could be calculated by adding the products of the subareas times the corresponding acoustic intensities for these subareas on any hypothetical surface containing the helicopter; besides intensity

measurements helped pinpoint the source. Measurements were made on a hypothetical rectangular box around the helicopter. SPLs were also recorded in the cabin of the aircraft.

Field tests were carried out on the helicopter in the following configuration:

(a) Engine running at maximum rpm, i.e. 33500 ± 200
(b) Main rotor fully engaged and rotating at 353 rpm
(c) Collective fully down.

The intensity measurements were conducted in an open field on an aviation tarmac since there was no anechoic or reverberation chamber available. The measurements were made at a distance of 6 m around the helicopter, which is far field for engine noise and near field for the helicopter main rotor. Since the dominant noise source for the helicopter is the engine, the near field effects from the main rotor were considered negligible. A hypothetical rectangular box of dimensions 13 m × 11 m × 3 m (Fig. 14.29) was chosen around the helicopter to map the intensities. This box was divided into many subareas and was scanned and mapped for intensity. Owing to flight safety reasons, the top surface of the box could not be mapped.

The helicopter with one pilot and co-pilot was started and the engine was throttled to a maximum RPM of 33500. The Sound analyser (HP 3569A, USA) was used after calibrating and configuring for the required test specifications. The probe was kept in the external gated mode, and intensities in the area of subsurfaces were recorded for subsequent power calculations. The subareas were so decided as to
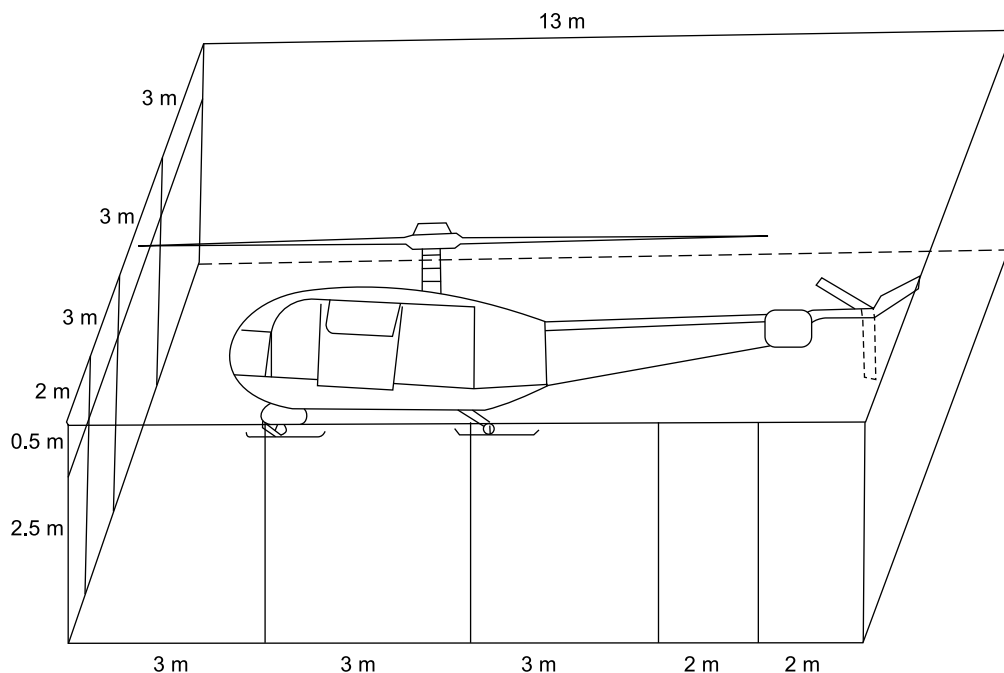
FIGURE 14.29 Hypothetical box with subareas

highlight the probable sound sources to obtain maximum information. They were chosen so as to cover the following sources: main rotor, engine, main gearbox (MGB) and tail rotor. The front and side subareas are shown in Fig. 14.29; the subareas for the other two sides were identical. The rectangular box surrounding the aircraft was marked on the ground around the helicopter for reference. A mild steel bar grid was fabricated and was used for providing a reference grid for scanning the sub areas, with the probe remaining perpendicular to the subareas. This grid also prevented the person making measurements from inadvertently stepping into the main/tail rotor region and thus served as a flight safety measure. Each subarea was scanned with an approximate speed of 0.25 m/s and with 30 cm spacing between the sweeps. The measured data were stored after each measurement and the grid shifted to the next subarea; this process was repeated till the four sides around the helicopter had been scanned and intensity maps recorded for each subarea as a trace to complete the multispectrum in 1/3 octave bands from 50 Hz to 10 kHz. Figure 14.30 shows a typical subarea and the scanning pattern.
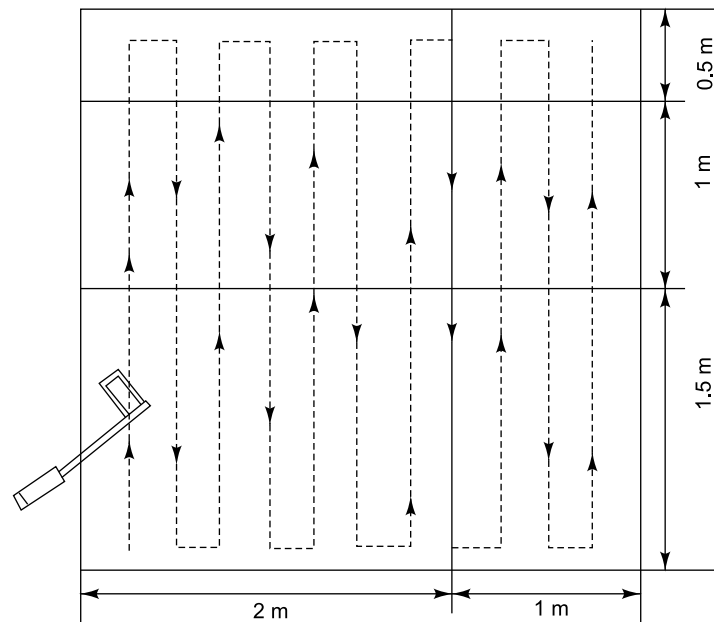
FIGURE 14.30   Typical subarea and scanning pattern
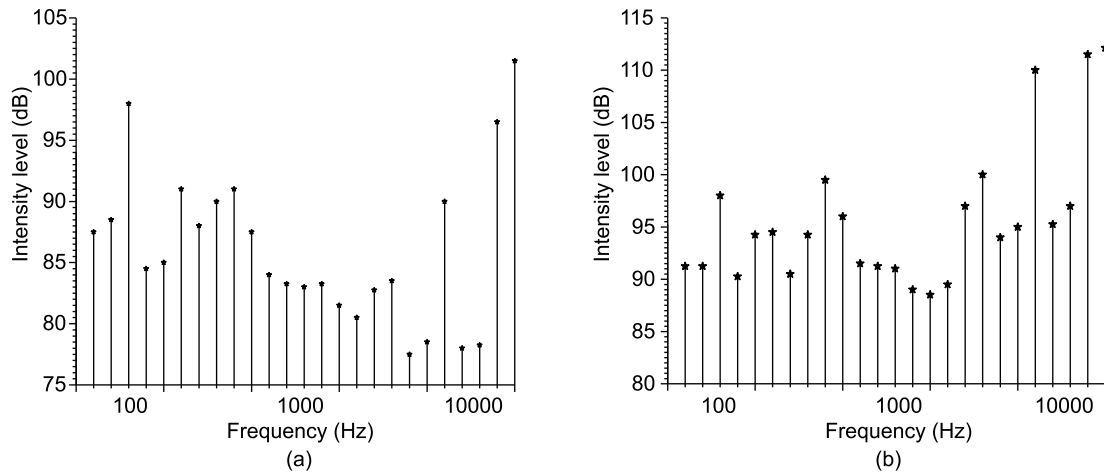
FIGURE 14.31 Intensity levels: (a) front side, (b) advancing side

Sound pressure levels inside the cabin were also measured with the same instrument (HP3569A) in the sound pressure measuring mode. The probe was kept at such a location in the aircraft cabin where it could record the SPL near the pilot's ear. Sound pressure levels were recorded in the 1/3 octave bands ranging from 50 Hz to 10 kHz. Figure 14.31(a) shows the intensity level corresponding to a subarea on the front side of the helicopter which reveals the rotor noise characteristics. Figure 14.31(b) shows the intensity level corresponding to a subarea on the advancing side of the helicopter which reveals the main gearbox noise characteristics. The results were plotted as intensity in dB versus frequency in 1/3 octave bands.

Index