Appendix 3: General Notions Regarding the Diagnosis of the Functioning State of Machine Tools

A3.1 COMPELLED VIBRATIONS

Compelled vibrations appear because of kinematics and/or dynamic factors. They have a permanent occurrence and major consequences on the working of the technological equipment, and on the equipment, technological process, and workpiece quality.

The usual classification of this type of vibration incorporates (a) compelled vibrations that depend on the working process, including such factors as stock variation, the periodic variation of the chip cross-section (milling, stitching), workpiece material hardness variation, and work-speed variation; being dependent on the working process characteristics these vibrations are very difficult to avoid; and (b) compelled vibrations that are not dependent on the working process, which appear because of a deficiency mounting of the equipment, technological and assembling lack of precision of the parts, and because of some specific features. This type of vibration can be observed during idle running of the equipment.

The following are some influence areas of factors that are independent of the working process.

The floor vibrations have a complex spectrum of periodic and random components and also shocks with a frequency range of $1, \ldots, 40$ Hz with amplitudes between $0.5, \dots, 15$ mm. The functioning of electrical engines (especially three-phased asynchronous engines and continuous current engines) induces a wider spectrum, $16, \dots, 530$ Hz, with amplitudes up to 0.4 mm.

A special influence on gearing working is made by pitch errors, profile errors, the wheels' eccentricities, and the axle deformations on which they are mounted. Two critical zones have been observed in experiments: one corresponding to the specific frequency of the elastic system of the teeth and the other corresponding to the specific frequency of the elastic system gear-shaft.

The driving belt transmissions introduce vibrations corresponding to specific throbs. They can be computed using the formula:

$$\omega n = \frac{\pi n}{l}\sqrt{\frac{S}{m}}$$

where l is the length of the belt, m is the weight of the belt, S is the belt tension, and n is the pulsation number ($n = 1, 2, 3, \dots$).

The vibrations caused by the movement of the bearings depends mainly on shaft rotation frequency and on the number of rolling elements. The functioning of hydraulic systems, of cam mechanisms, Malta cross, and the like also produce compelled vibrations of frequencies and amplitudes which can be estimated.

Compelled vibrations, within or outside the process, appear and take place simultaneously, which means that vibrational phenomena have a complex nature and their consequences should be evaluated taking into account all the stages of the working process.

A3.2 MEASUREMENT OF VIBRATIONS, SPECIFIC QUANTITIES

The vibrator signal picked up in the measurement points using translators is, usually, an aleatory sum of periodic and nonperiodic vibrations. Ways of processing and evaluating this signal are described below.

A3.2.1 Periodic Determinist Vibrations

The pure harmonic movement (Fig. A3.1) is characterized by the mathematical equation:

$$x(t) = X_v * \sin(\omega t + \theta)$$

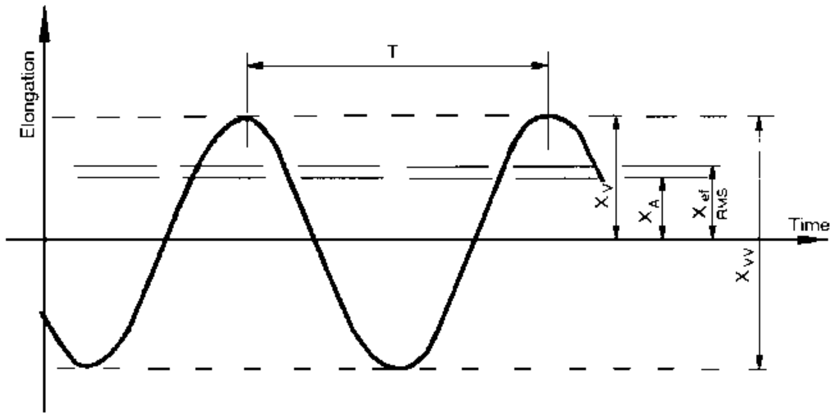

FIGURE A3.1 Pure harmonic movement.

where X_v is movement amplitude (the peak value), θ is phase difference, and ω is movement throb.

By derivation one can obtain the movement speed and acceleration, respectively:

$$v(t) = \frac{dx}{dt} = \omega X_v \cos(\omega t + \theta) = V_v \sin\left(\omega t + \theta + \frac{\pi}{2}\right)$$

$$a(t) = \frac{d^2x}{dt^2} = -\omega^2 X_v \sin(\omega t + \theta) = A_v \sin(\omega t + \theta + \pi)$$

The speed and the acceleration of the movement are also harmonic, having the same throb as the displacement, with a phase difference of $\pi/2$ and π, respectively.

The characterization of periodic determinist vibrations is possible not only by throb and amplitude but also by defining some features related to the process progress during one period:

Absolute average value (mathematical), X_A:

$$X_A = \frac{1}{T}\int_0^T |x(t)|dt$$

Effective value (square average), X_{ef} (RMS):

$$X_{ef} = \sqrt{\frac{1}{T}\int_0^T x^2(t)dt}$$

The configuration element, F_f:

$$F_f = \frac{X_{ef}}{X_a}$$

The peak element, F_v:

$$F_v = \frac{X_v}{X_{ef}}$$

Note: in the case of pure harmonic movement between X_V, X_A, and X_{ef} for F_f and F_V, respectively, the mathematical relationships are:

$$X_{ef} = \frac{\pi}{2\sqrt{2}} X_A = \frac{1}{\sqrt{2}} X_V$$

$$F_f = \frac{\pi}{2\sqrt{2}} = 1.11\ (\approx 1\ \text{dB})$$

$$F_V = \sqrt{2} = 1.414\ (\approx 3\ \text{dB})$$

The deformation energy (elastic) of the elastic system, W_p is

$$W_p = \frac{k}{2} \int_0^T x^2(t)dt$$

Among these parameters the most important is X_{ef} because it is proportional to the vibration power; this can be seen from the previous mathematical formula.

A3.2.2 Aleatory Vibrations

In this case movement is irregular and not repeated in time; the vibration should be monitored constantly (theoretically, infinitely); however, this is impossible. In practical terms it can work at specific intervals of time, which are called “achievements” of the periodic process. All these “achievements,” picked up in similar conditions, form the aleatory process itself. This method requires a very long time; as a result, one can introduce some probabilistic features that are capable of characterization of the aleatory vibrations.

The performances can be interpreted in terms of time or overall (Fig. A3.2). The amplitude in a specific time t_1 is an aleatory feature, and it is characterized by the statistical values during the considered "achievements":

$$x_1(t_1) = [x_1(t_1), x_2(t_1), x_3(t_1) \ldots]$$

If the same probability is considered for every "achievement" $xk(t)$ the following features can be described:

The average value of the amplitude, on time t:

$$m(t_1) = \lim \frac{1}{n} \sum_{k=1}^{n} x_k(t_1)$$

The autocorrelation function, which can appreciate in what measure the aleatory process remains identical with itself:

$$\psi(t_1, \tau) = \lim \frac{1}{n} \sum_{k=1}^{n} x_k(t_1) \cdot x_k(t_1 + \tau)$$

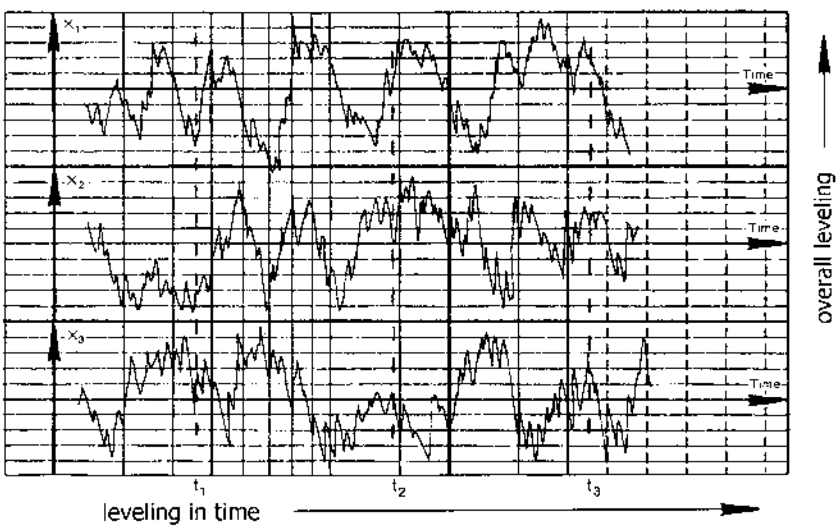

FIGURE A3.2 Aleatory vibrations.

The intercorrelation function:

$$\psi'(t_1, \tau) = \lim \frac{1}{n} \sum_{k=1}^{n} x_k(t_1) \cdot y_k(t_1 + \tau)$$

which estimates the way two aleatory signals $x(t)$, $y(t)$ are similar.

A3.2.3 Ergodic and Stationary Aleatory Vibrations

Stationary processes are those processes for which the average value and autocorrelation function do not depend on the specific time $t1$:

$$m(t1) = m(t2) = \cdots = m$$

$$\psi(t1, \tau) = \psi(t2, \tau) = \cdots = \psi(\tau)$$

For every "achievement" of a stationary aleatory process one can calculate the average value and the autocorrelation function, respectively:

$$m_x(k) = \lim \frac{1}{T} \int_{T/2}^{T/2} x_k(t) dt$$

$$\psi(k, \tau) = \lim \frac{1}{T} \int_{T/2}^{T/2} x_k(t) \cdot x_k(t + \tau)$$

This means that an ergodic aleatory process can be characterized by a "single achievement" which is a great advantage in practice.

Other features for these processes inclue the repartition function of the amplitudes; $F(x)$ is defined by the probability that the movement amplitude is inferior to a given value x:

$$F(x) = P(-\infty < x(t) < x) = P(x(t) < x)$$

This function is a monotone increasing function (Fig. A3.3) and it helps determine the probability that the movement amplitude can be found in a given range (a, b):

$$P(a < x(t) < b) = F(b) - F(a)$$

In addition, the amplitude probability density is the limit ratio between the probability that the momentary amplitude of the aleatory movement

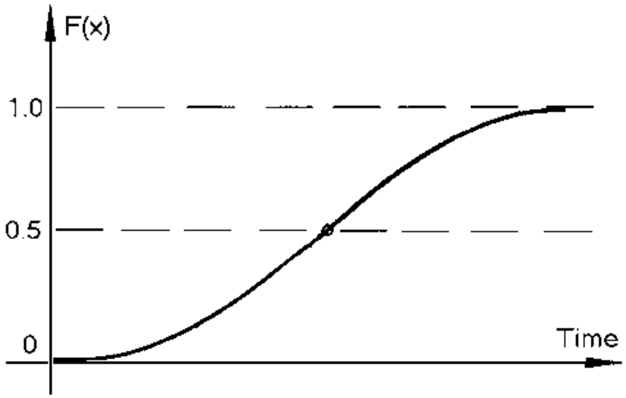

FIGURE A3.3 Repartition function of the amplitudes, $F(x)$.

can be found stuck between a given interval and the magnitude of this interval when it tends to zero:

$$p(x) = \lim \frac{P(x \leq x(t) \leq x + \Delta x)}{\Delta x}$$

In conformity with the previous relation it can be inferred that

$$p(x) = \lim \frac{F(x + \Delta x) - F(x)}{\Delta x} = \frac{dF(x)}{dx}$$

The most well-known analytical form for the amplitude probability density function is the Gaussian (Fig. A3.4).

A3.2.4 Noise and Acoustic Emission

A3.2.4.1 Sound and Noise

Sound is the sensation that is perceptible by the human ear as a result of rapid fluctuations of air pressure; it represents the mechanical vibration of an elastic medium in which the energy can be propagated from the source by progressive sound waves. The noise is usually described as a sound or a sum of undesirable sounds. It is considered to be a byproduct of daily activity.

The characteristic of sound waves is that substantial particles oscillate with respect to an equilibrium position and the wave propagation speed (the sound speed) is significantly higher than the oscillation speed

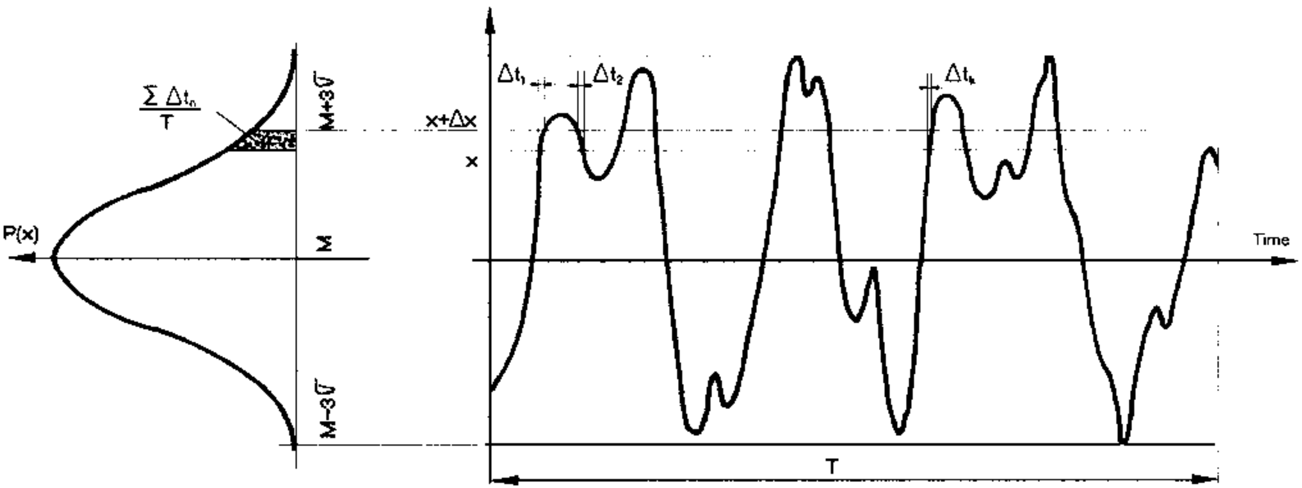

FIGURE A3.4 Gauss form for the amplitude probability density function.

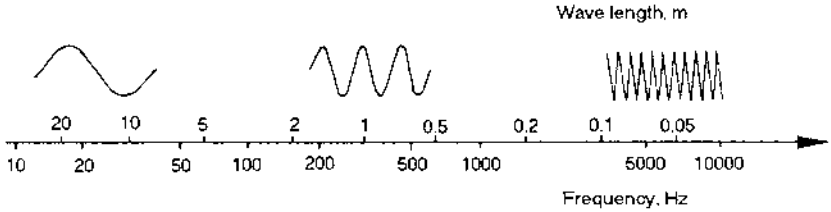

FIGURE A3.5 Dependency of wavelength by frequency.

of the particles. The distance traveled by the front wave during an oscillation period is called the wavelength: $\lambda 25 = c/f = c * T$, where c is the propagation of the sound speed, f is the frequency, and T is the period or the duration of a complete oscillation.

The dependency of the wavelength on the frequency is shown in Figure A3.5, where the medium of sound propagation is air; for this medium the speed propagation is given by:

$$c = f\sqrt{\gamma_0 \frac{p}{\rho}},$$

where $\gamma 0$ is the ratio between the specific heat on constant pressure and the specific heat on constant volume, p is static pressure, and ρ is the mass of the volume unit of the medium. Depending on the distance between source and receiver, the sound waves are considered to propagate in the shape of progressive plane waves or progressive spherical waves while the source can be pointlike or linear.

Decibel. The introduction of a decibel measurement scale (dB) for a subjective evaluation of sound power is based on Weber–Fechner physiological law. According to this law, the subjective sensation is proportional to the decimal logarithm of the excitation when the reference level of the acoustic intensity is zero. The lowest acoustic pressure that the human ear can sense is 20 mPa, which is $5 * 109$ times weaker than normal atmospheric pressure.

The sonorous pressure, expressed in mPa, varies in a very large range from 20 to 108 mPa. Therefore, it is difficult to do mathematical calculus using such a scale. The decibel scale (Fig. A3.6) avoids this difficulty because it uses an audibility threshold of the value 20 mPa. This value is considered to be 0 dB. Consequently, every time the acoustic pressure (in Pa) is multiplied by 10, 20 dB will be added to the decibel

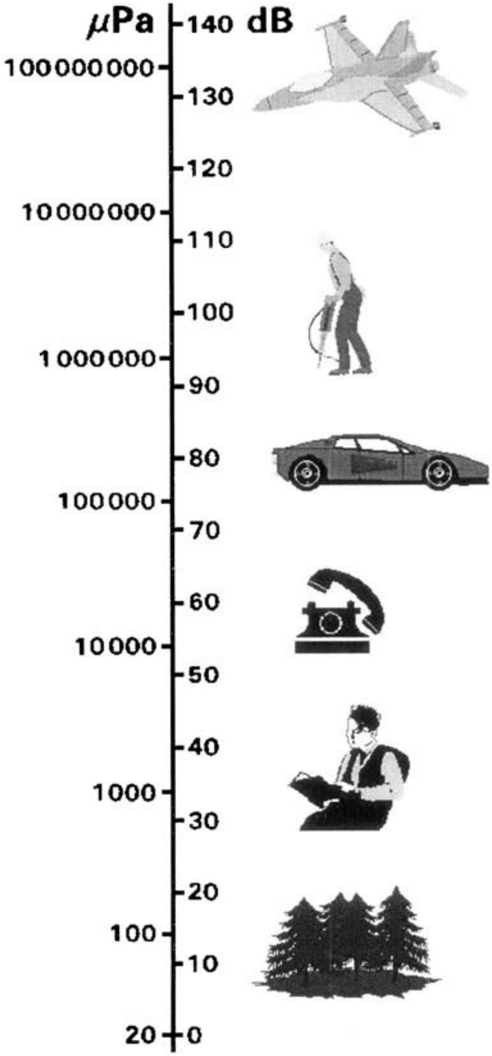

FIGURE A3.6 Decibel scale.

level; hence, 200 mPa corresponds to 20 dB, 2000 mPa corresponds to 40 dB, and so on. So, the scale in decibels compresses the values from 20 to 20 million mPa in a range of 0 to 120 dB.

Another useful aspect of the decibel scale is that it offers a better approximation of the noise threshold by human perception; the human ear reacts to relative changes of threshold. On this scale, 1 dB variation is the same relative variation no matter where it is placed on the scale, 1 dB being also the lowest variation that humans can sense. An increase of 6 dB represents twice the threshold of acoustic pressure and an increase of 10 dB is necessary in order to obtain a sound twice as strong.

Physiological Perception of the Noise. The human auditory apparatus, the ear, allows the perception of sounds produced by different sound waves with frequencies between 16 Hz and 20 kHz (the audibility domain). The maximum sensitivity of the ear is in the range of 2000 and 6000 Hz.

If two sounds have frequencies $f1$ and $f2$ it is said that they are separated by the interval $f1/f2$. If $f2 > f1$ those two frequencies have the bandwidth $\Delta f = f1 - f2$. In the acoustic industry, bandwidths having an octave and a tierce (a third of an octave) which have the corresponding intervals of 2 and $\sqrt[3]{2} = 1.26$, respectively, are very important. The central frequency fc of an octave is the frequency that has the limited frequencies $f1$ and $f2$ such as $fc = \sqrt{2} * f1$. When bandwidths have an octave, the normalized central frequencies are 31, 63, 125, 250, 500, 1000, 2000, 4000, 8000, 16,000 Hz and upward.

In order for a sound to be perceptible, it is necessary that its sonorous intensity have a specific minimal level that depends on sound frequency and on the sensibility of the ear. The lower limit of the acoustic pressure, for a given frequency that can be heard by a human being is called the audibility threshold. It is considered to be a sound with the frequency of 1000 Hz and sonorous pressure of $2*10-5$ N/m2. For lower frequency sounds (<1000 Hz) the audibility threshold increases.

On the other hand, very strong sounds create pressure on the eardrum that can induce pain. The threshold where the pain appears is $2 * 10$ N/m2 for the frequency of 1000 Hz. The power of the sounds represents a feature according to which the sounds can be arranged from weak to strong. Subjective perception of a sound or noise strength depends on the acoustic pressure level and on its spectral characteristic.

Figure A3.7 illustrates Fletcher–Munson izosonic curves. Standard sound has the frequency 1000 Hz for an acoustic pressure of 1 dB.

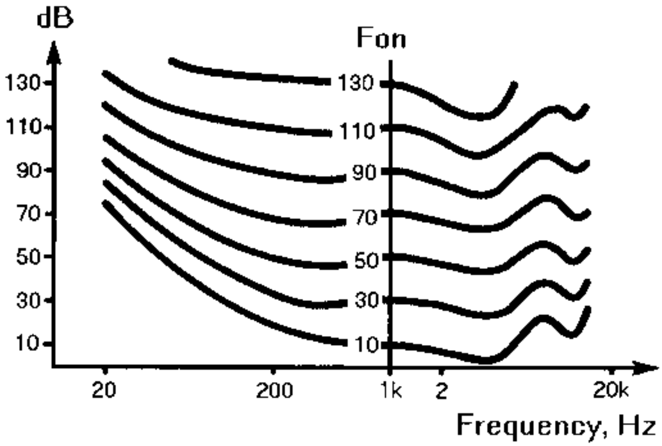

FIGURE A3.7 The Fletcher–Munson izosonic curves.

In order to appreciate the acoustic pressure level the term moderate acoustic level has been used. Hence, the measurement apparatus is equipped with balanced filters given by the A, B, C, or D curves in Figure A3.8. The most used moderate acoustic level is represented by the A curve, especially in industry and transportation. In the aerospace industry the moderate acoustic level is illustrated by the D curve.

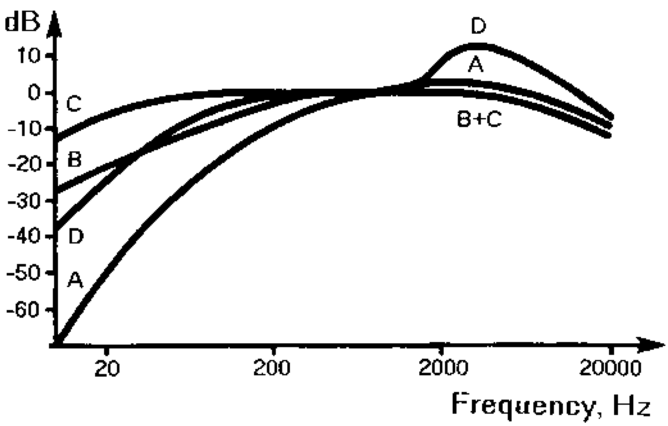

FIGURE A3.8 Balanced filters for measurement apparatus.

A3.2.4.2 Acoustic Emission

Acoustic emission is the sequence of elastic waves generated by the release of the internal energy stored in a structure. It becomes manifest in the higher frequency domain ($f > 100$ kHz) by elastic waves detected as vibrations on the structure surface. Acoustic emission represents a nondestructive method used in order to perceive when and where a flaw or a crack appears. The nature and the causes of these shortcomings are determined using complementary methods.

There are four main sources of acoustic emission:

Movements of structural dislocation
Phase transformations
Friction mechanisms (microfrictions, microcollisions)
Formation and development of cracks

In the case of dislocations (movement of a line imperfection within a crystalline structure) which develop like an avalanche, the signal is continuous while in the case of phase transformations (the formation of martensite in carbon steel) the signal is impulse type and can be detected for every transformed grain.

Flaws come into view in the material where the stress outruns strength tension. New surfaces appear and there is a release of energy that is partially transformed in acoustic emission. The signal is impulse type having high frequency. At the same time, friction mechanisms also emanate acoustic signals. The signal amplitudes in acoustic emission cover a large area; in relative units these amplitudes are 1 to 10 for structural movements, 5 to 1000 for phase transformations, and 20 to 1000 for flaws. For machining with machine tools these sources are: continuous and discontinuous chip forming; deformation of the workpiece material; cracking of the workpiece or of the tool; or friction between workpiece, tool, chip breaker, breakage, and collision of the chip. There are also sources that appear from the functioning of the mechanical subsystem (bearings, gears) and high-frequency electrical sources.

The propagation of acoustic emission is similar to radio waves. The source emits spherical wave packages that are influenced by the surfaces which are intersected; this creates reflections and surface waves (Fig. A3.9). The heterogeneities of the propagation medium distort the front waves. Consequently, the mathematical relationships that describe the real propagation phenomena become very complicated when one needs to locate and measure the sources and effects of the phenomena.

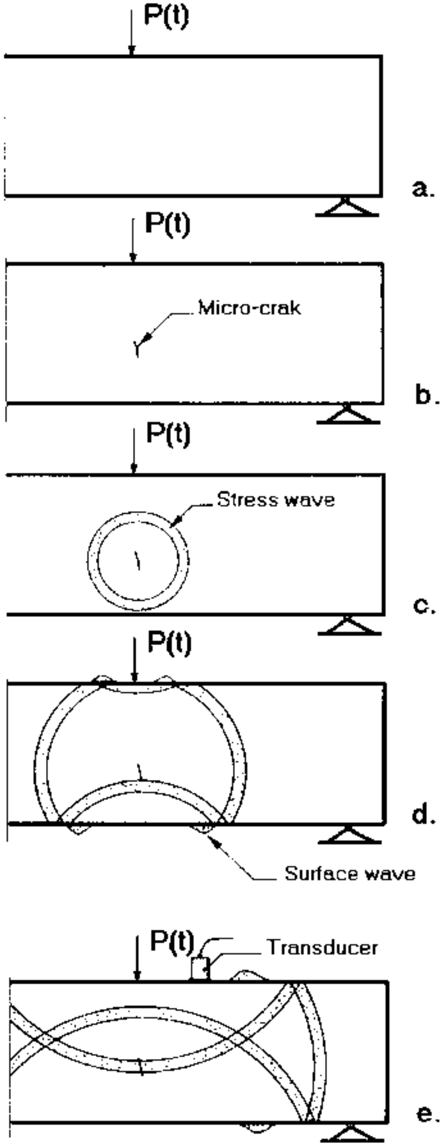

FIGURE A3.9 Wave packages influenced by surfaces they intersect.

Further on, the area of utilizing these methods (based on acoustic emission) on large steel structures where an uncertainty coefficient can be accepted, becomes very narrow.

In industry acoustic emission is used for controlling wear and tool breakage and also for supervising the bearings and hydrodynamic bearings.

A3.2.4.3 Acoustic Emission Measurement

As shown in the previous section, acoustic emission is constituted from surface waves. Special piezoelectric transducers connected to the structure that is to be monitored can perceive them very well. The signal monitored by the transducer can be processed in two ways: for a continuous emission the most useful information is given by the voltameter; or for impulse type signals one can use the impulse analyzer. Another technique is based on counting the impulses using a peak indicator with an adjustable threshold level.

A3.2.4.4 Acoustic Parameters

Propagation Velocity of Sonic Waves. Sound waves in solids follow the equation

$$c_{l(t)} = \sqrt{\frac{E(G)}{\rho}}$$

where E is the longitudinal elasticity module, G is the transversal one, and $\rho 1$ is the density of the medium, which is homogeneous and isotropic.

The propagation of the sound waves in liquids follows a similar equation, where ko is the compression modulus:

$$c = \sqrt{\frac{k_0}{\rho}}$$

Sound wave propagation in gases is adiabatic and follows, therefore, the formula $pV^\gamma = \text{const.}2$, and the propagation velocity is given by the relation:

$$c = \sqrt{\frac{\gamma p}{\rho}} = \sqrt{\frac{\gamma RT}{\mu}}$$

where $\gamma 3 = cp/cv$ is the adiabatic exponent, while R is the constant of perfect gases. For the case of sound wave propagation through air, the relation remains

$$c = 332\sqrt{1 + \frac{t}{273}}$$

where t is the air temperature (°C), which leads to determination of the sound wave propagation at a temperature of 20°C to be approximately 340 m/s.

Sonic pressure, is, by definition, the average of pressure variation, meaning:

$$p_s = \sqrt{(\overline{\Delta p})^2}$$

The acoustic level (level of sonic pressure) is determined as

$$L_p = 10 \log \left(\frac{p}{p_0}\right)^2 = 20 \log \left(\frac{p}{p_0}\right)$$

where $po = 20$ mPa is the threshold, considered to be the audibility limit for a sound having the frequency of 1 Hz.

The intensity of the sound can be defined as the average value of the acoustic energy that travels through the surface unit, on a direction perpendicular to the propagation, in a unit of time. The equation of the sound intensity depends on the propagation area: for the progressive wave in a free field it is

$$I = \frac{p_{ef}^2}{\rho c} 4$$

and for the progressive wave in a fuzzy field it is

$$I = \frac{p_{ef}^2}{4\rho c} 5$$

where $p_{ef}^2 6$ is the square average value of the sonic pressure.

The sound intensity level can be described similarly to the acoustic level; that is,

$$L_i = 10 \log \left(\frac{I}{I_0}\right)$$

where Io is the threshold of the acoustic intensity; that is, $Io = 10$ to 12 W/m2. It can be pointed out that in normal conditions ($po = 1$ atm. and $to = 22°\text{C}$), the difference between the value of the acoustic pressure and acoustic intensity level is, for the same sound, 0.16 dB, and, therefore, it can practically be ignored.

The level of acoustic power of a sonic source can be evaluated by using the relation

$$L_w = 10 \log \left(\frac{P}{P_0} \right)$$

where Po is the threshold of the acoustic power level, $Po = 10$ to 12 W.

A3.3 CORRELATION BETWEEN SOUND ANALYSIS AND MACHINE TOOL PERFORMANCE: TECHNICAL DIAGNOSIS EQUIPMENT

A3.3.1 Time Analysis

The acoustic signal can be considered to be created by a sum of pure harmonic pulsations having different intensities and frequencies, and therefore

$$x(t) = x_0 + \sum_{i=1}^{n} x_i \sin \frac{2\pi t}{T_i}.$$

As previously presented, it is possible to characterize a signal by using a series of parameters that define the progress of the signal in time (usually, for one period): absolute arithmetic mean, xA; effective value (of the squared average), xef; and pick value, xv, $xv = \max[x(t)]$.

Correlation analysis is the most utilized tool for time analysis, being recommended for linear systems that operate either with continuous or discrete signals, especially when the ratio between signal and noise has a small value. Acceptable input values are either stochastic signals or periodic signals, leading, as a result, to the correlation functions and, as a particular case of these, the weighted function. The correlation functions define the level of similarity between two given signals in accordance with the time delay between them (intercorrelation function):

$$R_{ab}(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} a(t) \cdot b(t + \tau) dt$$

or, even, the level of similarity of a signal with itself, in accordance with the time delay (self-correlation function):

$$R_{aa}(\tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} a(t) \cdot a(t+\tau) dt$$

For the case of a invariant stable and linear system, by noting with $h(k)$ the weighted function or the reaction of the considered system at an impulse, the self-correlation function and the intercorrelation function are connected by a relation having the form:

$$R_{ab}(\tau) = \sum_{k=0}^{\infty} h(k) \cdot R_{aa}(\tau - k)$$

A correlation graph is a graphical representation of the self-correlation function in accordance with the time delay. Its form suggests the contents of the signal's frequencies. The correlation graph gives a peak value for $t = 0$, which is sharper as the contents in high frequencies are richer; this value is equal to the square mean value of a random process, being, therefore:

$$R_{aa}(0) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} a(t)a(t)dt = \overline{x^2(t)}$$

For a linear system, the integral is proportional to the energy along the interval considered; by dividing this value by the length of the interval one can obtain the average power along that interval, and this is the physical significance of the value of the square mean.

Figure A3.10 presents some significant types of correlation graphs: (a) the correlation graph of a random and stationary ideal process (no noise); (b) the correlation graph of a nonperiodic process for large values of the time displacement t (the self-correlation functions tend to attend the square mean value); (c) idem, when the square average value is zero; and (d) the correlation graph of a random process that conceals a periodic phenomenon (the correlation graph tends to become a periodic function in time t).

By calculating the Fourier transform of the self-correlation function one can obtain the spectral density function of the square mean value (power spectral density):

$$\Im[R(\tau)] = \int_{-\infty}^{\infty} R(\tau) e^{-i2\pi f \tau} d\tau = S(f)$$

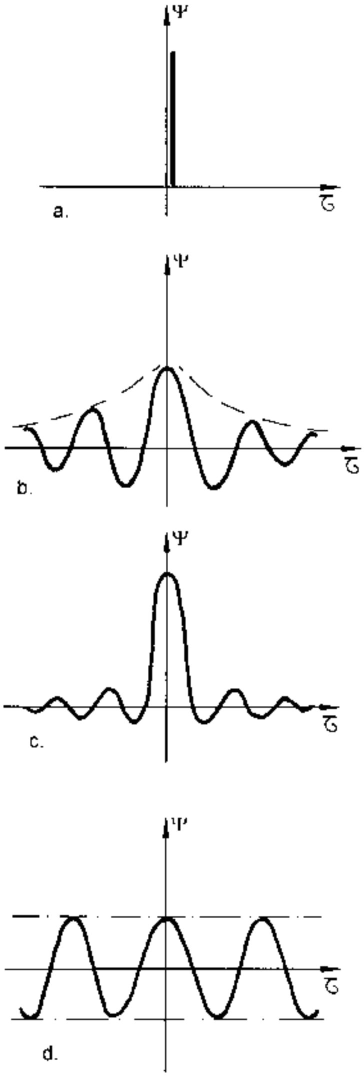

FIGURE A3.10 Significant types of correlation graphs: (a) correlation graph of random and stationary ideal process (no noise); (b) correlation graph of nonperiodic process for large values of time displacement t (self-correlation functions tend to attend the square mean value); (c) idem, when the square average value is zero; (d) correlation graph of random process that conceals a periodic phenomenon (correlation graph tends to become a periodic function in time t).

This function depicts the manner in which the square mean value (and the average power, also) is distributed in the domain of frequencies. A periodic vibration can be represented in the frequency domain by a band formed by discrete lines, each line representing the square mean value of the respective harmonic component, while a random vibration determines a continuous band in the frequency domain; so, the measured value of the square mean for a given frequency depends on the utilized bandwidth; that is why the spectral power density is useful. The self-correlation function and the spectral density power form a pair of Fourier transforms in such a manner that the first one can be rapidly obtained by applying an inverse Fourier transformation on the second one.

A3.3.2 Analysis in the Frequency Domain

The acoustic signals can not be analyzed by studying only the amplitude-time characteristic because this does not furnish sufficient data for a diagnostic interpretation.

The separation of the vibrations in individual frequency components is called frequency analysis (Fig. A3.11), this being a technique that can be considered as the fundament of the diagnosis, which is based on studying vibrations and acoustic signals. The curve that indicates the amplitude of the vibrations versus frequency is called a spectrogram.

The most sophisticated and most precise techniques for establishing the diagnosis of the performance of machine tools and equipment are based on signal analysis in the frequency domain. This analysis can be accomplished using various protocols, depending on the element studied, diagnosis method, and apparatus controlled by the researcher or existing in a research laboratory.

Gauge apparatuses for measuring vibrations and noise indicate a unique band of signals evaluated along the entire bandwidth (Fig. A3.12a). To evaluate the individual frequency components a filter is utilized, which will allow passing only those signals' components within a tide bandwidth. The filter's passband is successively displaced along the entire domain and as a result a reading of the level of vibration is obtained in the bandwidth. The filter can be formed either by a series of individual filters for fixed frequencies, which are adjacent and successively scanned (Fig. A3.12b), or by a unique adjustable filter displaced within the studied frequency domain (Fig. A3.12c).

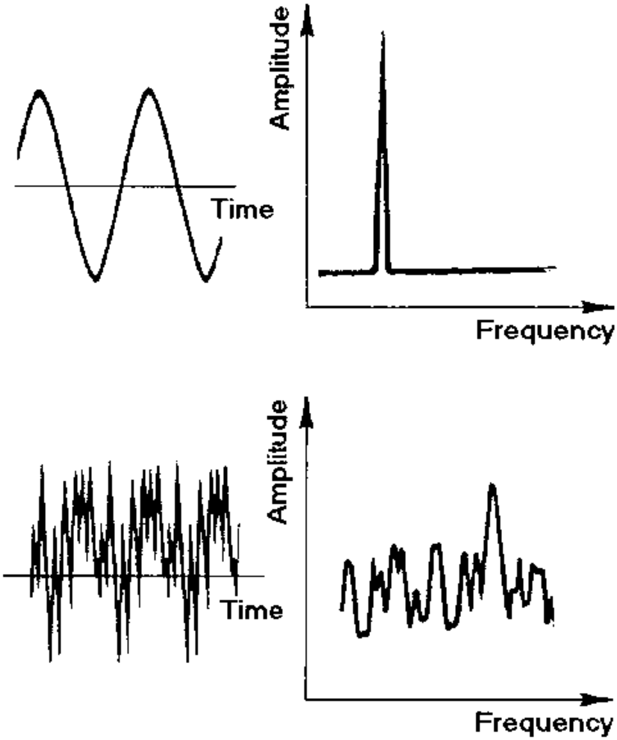

FIGURE A3.11 Frequency analysis.

A3.3.2.1 Frequency Analysis for Tight Bandwidth

This kind of analysis is the most common vibration processing procedure and, due to its precision and the accuracy of the components of the processed signal, the most utilized method for monitoring and diagnosis.

One can utilize two basic types of filters: constant bandwidth type, having an absolute bandwidth of 3 Hz, 10 Hz, and so on, and constant proportional bandwidth type, having a bandwidth expressed as a percentage (3%, 10%, etc.) of the selected central frequency.

In Figure A3.13 the difference between these two types of filters is presented. It should be stressed that the constant proportional bandwidth filters were built in order to maintain a constant bandwidth on logarithmic scales of frequencies, which are ideal for large bandwidths.

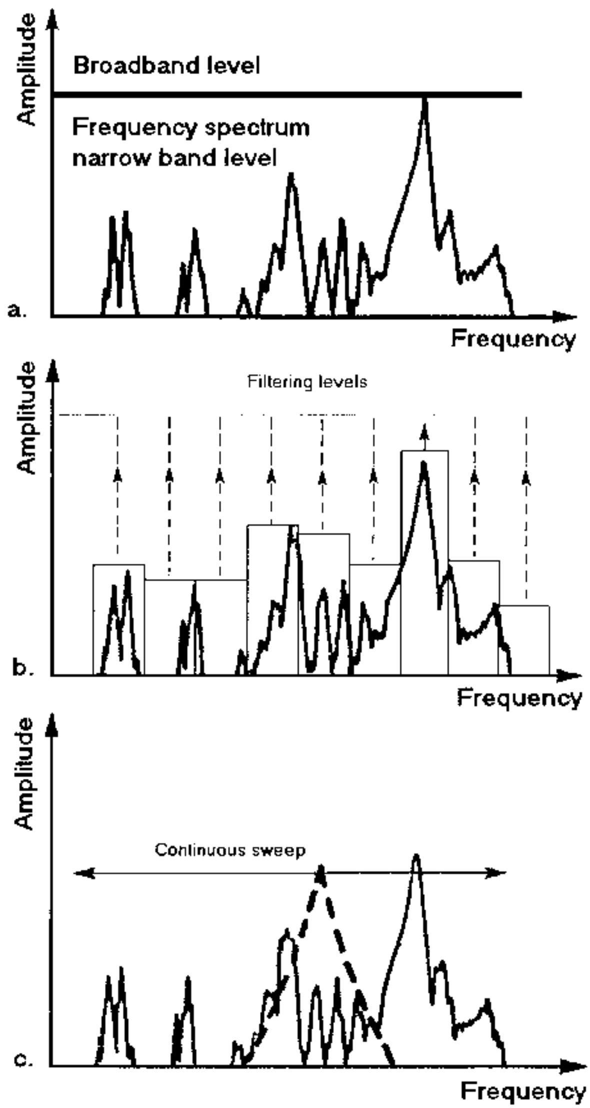

FIGURE A3.12 (a) Unique band of signals evaluated along entire bandwidth; (b) series of individual filters for fixed frequencies, adjacent and successively scanned; (c) unique adjustable filter displaced within studied frequency domain.

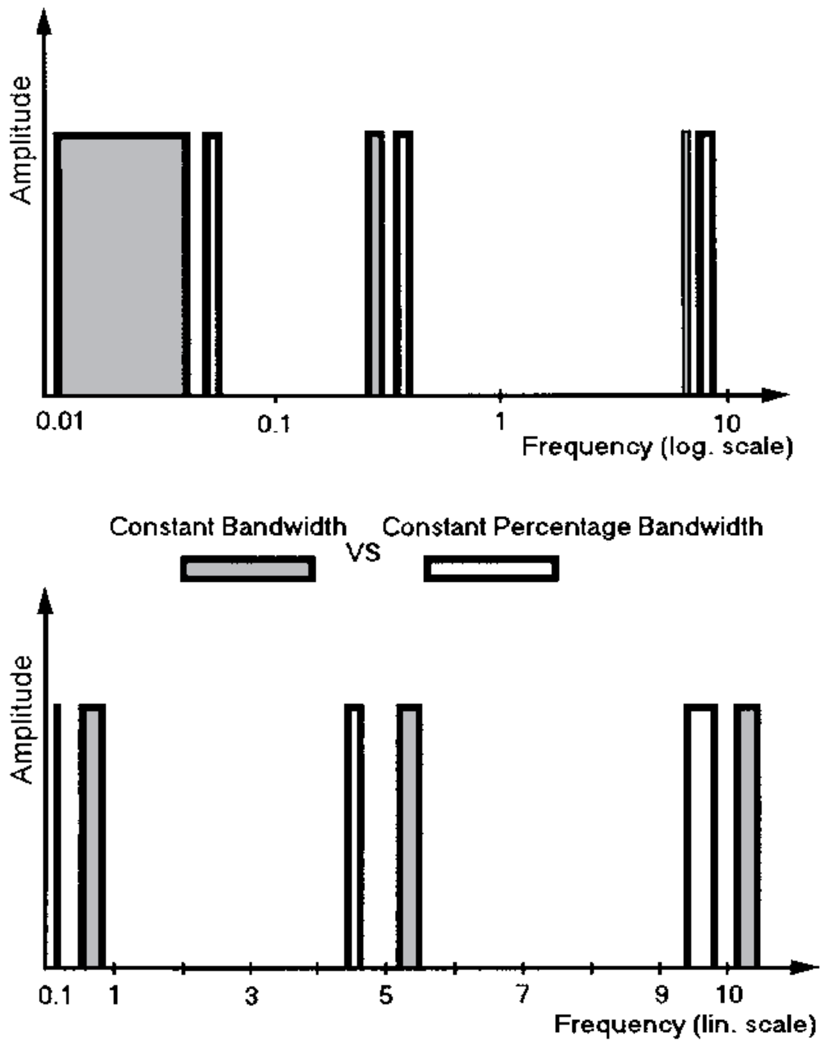

FIGURE A3.13 Difference between constant bandwidth and constant proportional bandwidth types of filters.

However, if the frequency scale is linear, a constant bandwidth filter will give a constant representation, while the constant proportional bandwidth filter will show an enlarged bandwidth, which is not an advantage in practical cases.

Constant proportional bandwidth filter analysis shows the natural response of the systems to mechanical vibrations and permits a compact representation of a large bandwidth; that is why this is the most commonly utilized method for measuring the vibrations.

Constant bandwidth filter analysis is utilized for high frequencies, especially for a linear scale, in order to distinguish the harmonic components.

The filter-pass bandwidth establishes the resolution of the frequency analysis to be obtained. The employment of a filter having a tight bandwidth offers numerous details and permits the isolation of individual peaks within the band, but has, at the same time, the disadvantage of a increased processing time together with a narrowed bandwidth. Sometimes a reduction of the processing time is possible through time compression, that is, fast-forward playing of the recorded signal. In that way the filter's bandwidth increases proportionally, and this leads to an increased scanning rate along the frequency domain.

The ideal filter should let pass all the frequency components that appear in the bandwidth while eliminating all other frequency components. Practically, electronic filters have no perfectly vertical limits, and therefore do not totally eliminate the components outside the bandwidth domain.

In practice two methods are utilized for measuring the filter's bandwidth: as the bandwidth of an ideal filter that lets pass the same quantity of power coming from a white noise source such as the described filter (Fig. A3.14a); and as the bandwidth of a filter that shows an altering of 3 dB in rapport with the normal transmission level (Fig. A3.14b).

The 3 dB bandwidth will differ considerably from the bandwidth of effective noise only for low selectivity filters, and therefore one can consider that the two definitions lead to the same practical result.

A3.3.2.2 Frequency Analysis Using Fourier Series

By using the Fourier theorem and for the given Dirichlet conditions, the periodic deterministic vibrations can be looked upon as a sum of harmonic pulsations, having frequencies equal to multiples of a fundamental frequency. In this case, the transform in the frequency domain can be accomplished using Fourier series by decomposing; during the

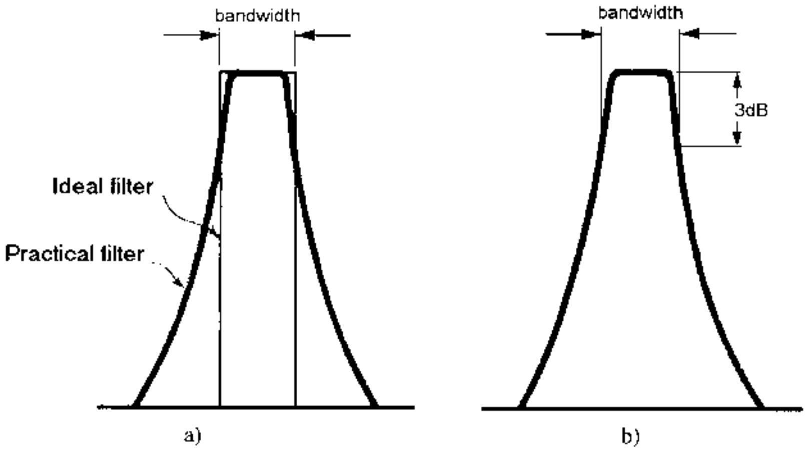

FIGURE A3.14 Two methods for measuring filter bandwidth used in practice.

transforming process only one period of the signal is usable. The transformation relations (Fig. A3.15a) convert the continuous periodic signal from the time domain into a discrete band in the frequency domain, a band that encloses all the harmonics of the signal. The inverse situation is also possible by which a discrete signal in time is transformed in a periodic frequency (Fig. A3.15b). One can observe that due to the symmetry and the periodicity of the frequency band a component having the frequency fc in the continuous signal will appear in the discrete signal at the frequencies $fd = n * fs \pm fc$, where fs is the discrete frequency and $n = 0, \pm 1, \pm 2, \ldots$. In this case, in order to avoid ambiguities in the frequency contents of the continuous signal, a pass filter is utilized, which is not as broad as the half value of the discrete frequency.

The transform relations shown in Figure A3.15 prove the basic symmetry of the Fourier transform between the time domain and the frequency domain.

A3.3.2.3 Frequency Analysis Using Discrete Fourier Transform

The discrete Fourier transform (DFT) is applied to discrete and periodic signals in the time domain and the result is, also, a discrete and periodic signal, but in the frequency domain (Fig. A3.16). Due to the periodicity in both domains, only a finite number of samples are employed, therefore

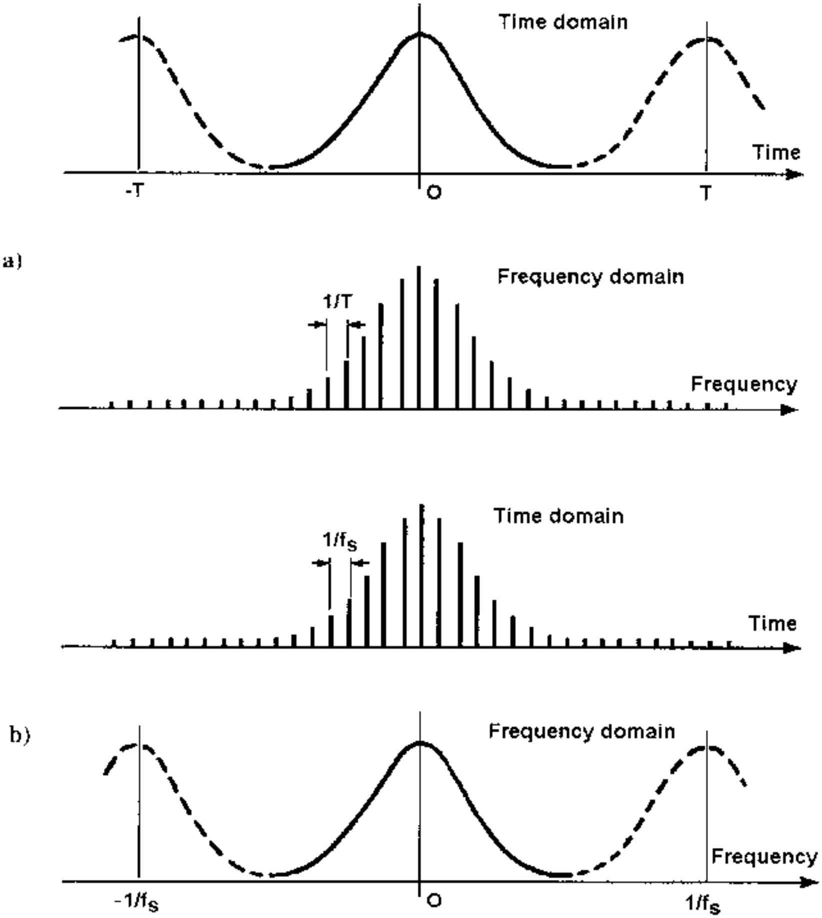

FIGURE A3.15 Transform in frequency domain can be accomplished using Fourier series.

the transform can be calculated using digital processing, for instance, if a period of a signal is described in the time domain by N samples along one period.

The determination of a discrete Fourier transform having N values within a sequence implies $N * N = N2$ multiplication and summation operations of complex numbers, which determine a rapid increase of the calculation time for a larger number of samples N. Nevertheless,

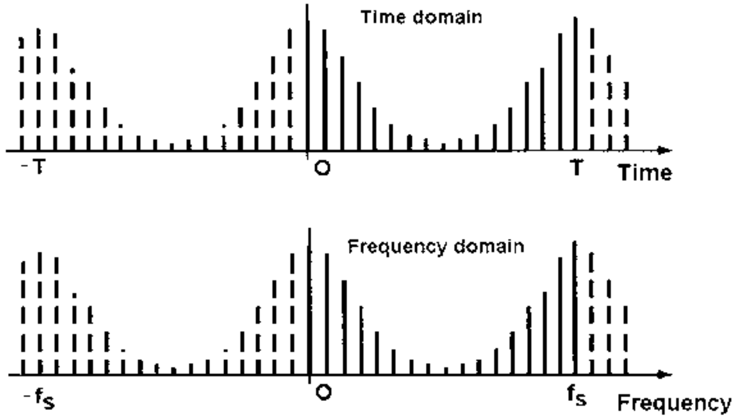

FIGURE A3.16 Discrete Fourier transform (DFT) is applied to discrete and periodic signals in the time domain.

due to the symmetry of the frequency band, only $N/2$ of them will be independent. In addition, by utilizing an anti-over imposing filter (having a bandwidth narrower than half of the sampling frequency) the number of significant frequency components is reduced even further. Usually, for a signal in time having 1024 samples only 400 frequency lines from the band will need to be processed.

A3.3.2.4 Frequency Analysis Utilizing Fast Fourier Transform

The algorithm of the fast Fourier transform (FFT) is based on several features of the complex exponential function in order to shorten the calculation time of a regular Fourier transform. In this case, the number of operations is diminished from $N2$ to $N * \log 2N$. The signal processing procedure for the case of FFT analysis is presented in Figure A3.17. One can observe that it is necessary to input the recorded signal into a sampling block having an analogue-to-digital conversion element because the output signal obtained at the output of the transducer has a continuous variation.

The FFT algorithm considers a record from the time domain as being a block of N samples equally separated in time, and which can be transformed into another block of N samples also equally separated but in the frequency domain (Fig. A3.18). The lowest frequency to be

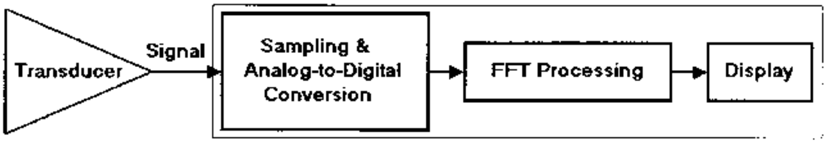

FIGURE A3.17 Signal processing procedure for the case of FFT analysis.

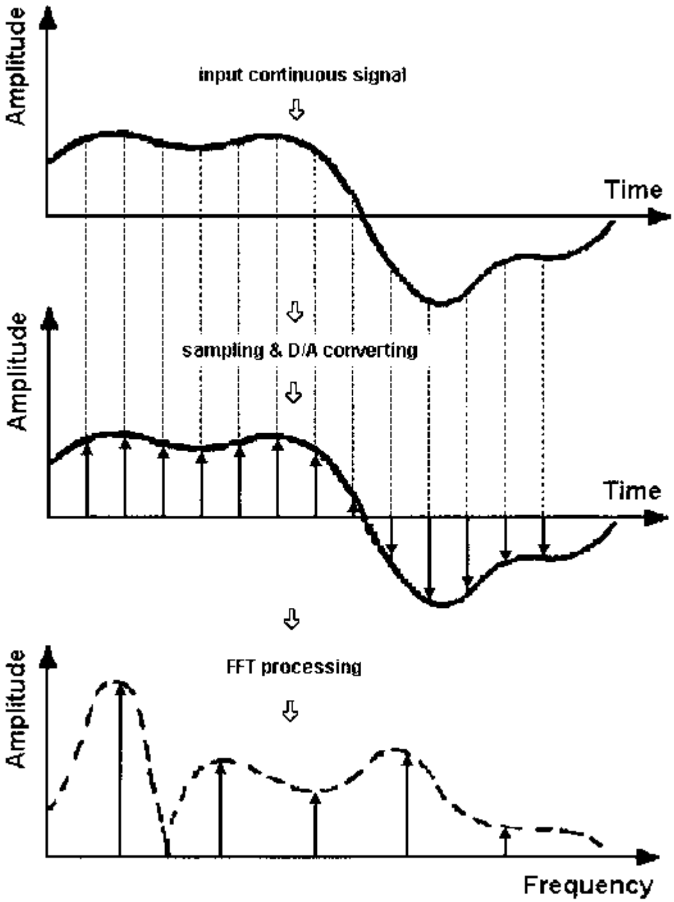

FIGURE A3.18 Block of N samples equally separated in time transformed into another block of N samples also equally separated but in the frequency domain.

analyzed by FFT is determined by the length of the record in the time domain, while the maximum value of the frequency that can be separated is $f\,\mathrm{max} = (N/2) * (1/TR)$, TR being the record duration.

The necessary calculations for using the FFT algorithm require a finite time, in terms of the number of samples (usually $N = 1024$), but also of the speed of the processor implied. If this calculation (TFFT) is less than the duration for the recorded signal (TR), the operation is called the analysis in real-time. In this case, the analyzer processes, along the duration of a record, a FFT analysis of the preceding record, so no information from the time domain is lost, a situation that can occur if TFFT is larger than TR.

The basic relation of the FFT algorithm is still the one known from the discrete Fourier transform; that is,

$$G(k) = \frac{1}{N} \sum_{n=0}^{N-1} g(n) e^{-j(2\pi kn/N)}$$

a relation that can be written using matrix form as

$$\{G_k\} = \frac{1}{N} * \{A_{kn}\} * \{g_n\}$$

where $\{Gk\}$ and $\{gn\}$ are the column vectors that include N samples from the time domain, and $\{Akn\}$ is an N-square matrix that contains the complex unit vectors $e - j2pkn/N$.

For example, the matrix equation is shown in Figure A3.19 for $N = 8$. Each arrow in the square matrix depicts a complex unit vector, with respect to the attached coordinate system. One can observe that a direct calculation of this matrix implies $N * N$ complex multiplication operations, which are time consuming. By using the FFT algorithm the

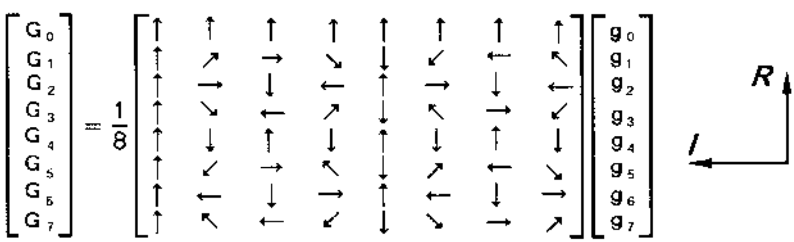

FIGURE A3.19 Matrix equation for $N = 8$.

number of multiplication operations reduces to $N * \log 2N$, with the condition that N is a power of 2. Typically, when $N = 1024$, the processing time is reduced approximately 100 times.

Errors Introduced by FFT Analysis. The aliasing effect occurs due to signal sampling in the time field; it shows up by the appearance of some high frequencies within the field of low frequencies after sampling. This effect can be removed by purposely installing a low-pass filter before sampling, to eliminate frequencies larger than one half of the sampling frequency.

The time window effect is the result of the finite length of recording in the time field; the FFT algorithm deals with this record as a periodic signal, with the period TR (Fig. A3.20). This approach is good for transitory signals with a period less than TR; the effect is, however, harmful for signals with a period larger than TR. Since the signal is "cut off" through a rectangular window and then is introduced in a loop in order to apply the FFT algorithm, distortions and discontinuities occur in transitory areas (Fig. A3.21a). Consequently, the frequency specter will have a few components that do not exist in the original signal. The solution is to use a "smooth" window, which has both its value and slope

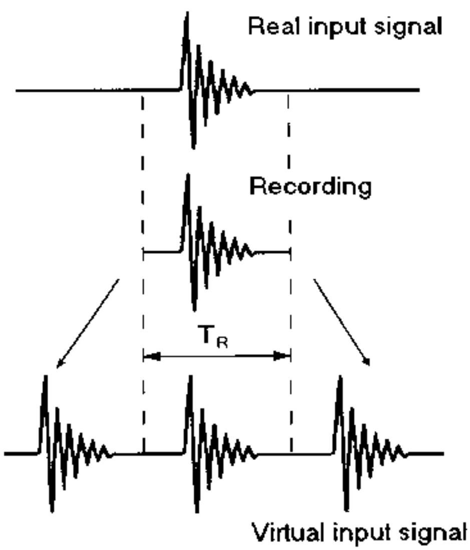

FIGURE A3.20 Result of finite recording length in time field treated by FFT algorithm as a periodic signal with period TR.

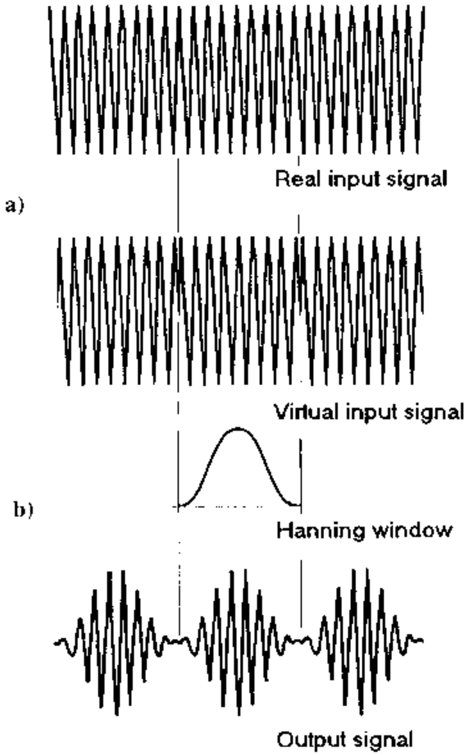

FIGURE A3.21 (a) While applying the FFT algorithm, distortions and discontinuities occur in transitory areas; (b) Hanning window.

equal to zero at both ends. Usually, this window is a Hanning window with period $\cos 2(2pt/TR)$, as in Figure A3.21b. The round peak of a Hanning window can lead to amplitude estimation errors up to 1.5 dB (16%). For other types of windows, errors can go down to 0.1 dB (1%).

The picket fence effect is the result of the discrete sampling of the specter in the frequency field. It appears as if the specter is seen through the slits of a fence. Consequently, some values, such as peak values, can not be observed. The possible resultant error depends on how the characteristics of the adjacent filters overlap (Fig. A3.22). This effect is much less as the overlap is bigger. For a Hanning window, the distortions introduced in this way do not exceed 1.4 dB, while for a rectangular window they reach 3.9 dB. The error can be compensated where a frequency component "fits" between two spectral lines. The

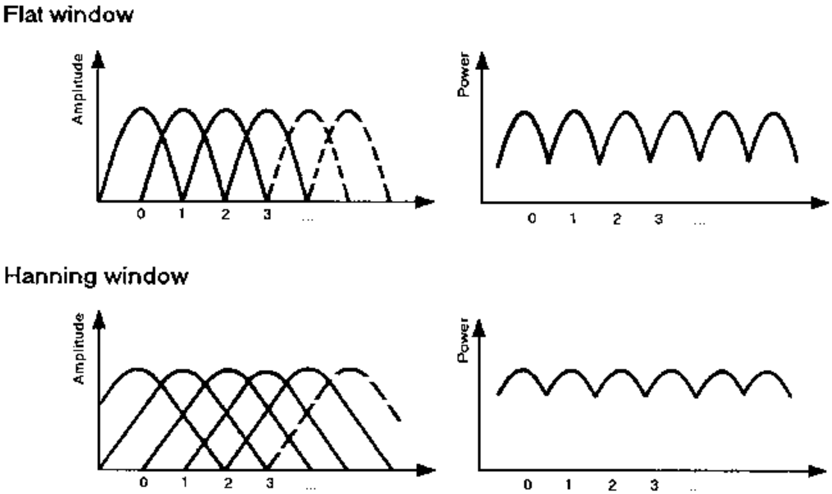

FIGURE A3.22 Possible resultant error depends on how characteristics of adjacent filters overlap.

picket fence effect also occurs in a TFD analysis and it is typical when third-octave filters are used.

A3.3.2.5 Zoom FFT Analysis

In a FFT analysis, the resolution of the result is determined by the Nyquist frequency (equal to one half of the sampling frequency), and by the number of lines in the specter up to the Nyquist frequency. When a resolution higher than the one offered by the 400 lines of the basic specter is desired, a Zoom FFT analysis is used. The interest area, which is between $f1$ and $f2$, is selected by moving the origin of the frequency representation to $f1$, simultaneously with the passing of the signal through a low-pass filter, in order to eliminate all components, except for the range between $f1$ and $f2$ (see Fig. A3.23). Zoom FFT analysis is useful for processing both low-frequency modulated signals, and signals with a large number of harmonics, as well as for separating some vibratory phenomena that have very close frequencies.

A3.3.2.6 Frequency Response Function

The Fourier transform allows a theoretic frequency response of a mechanical system for different types of excitations. For a linear system

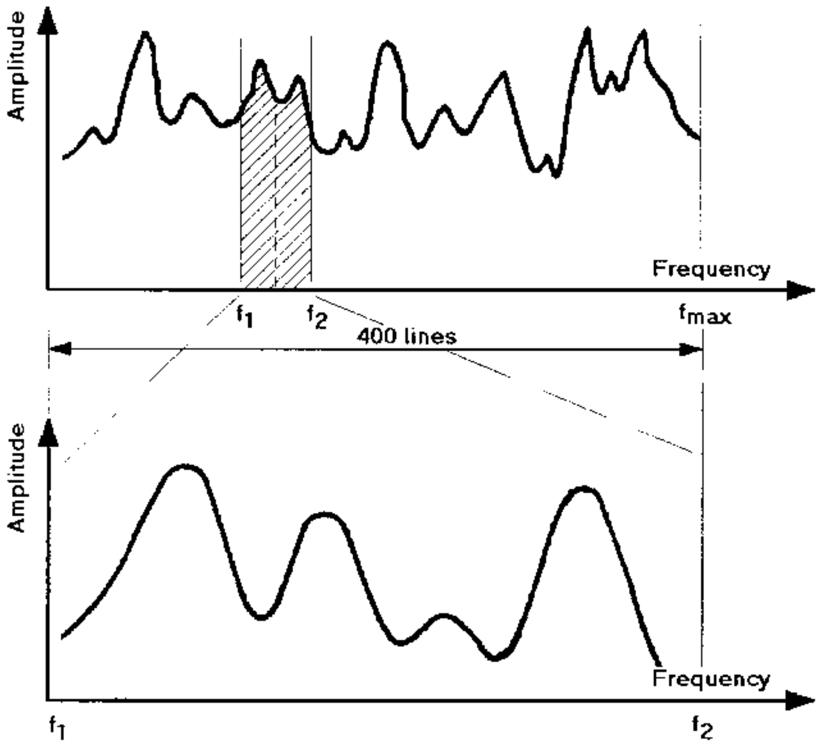

FIGURE A3.23 When a resolution higher than that offered by 400 lines of the basic specter is desired, a Zoom FFT analysis is used.

with one degree of freedom, the relationship between the excitation $f(t)$ and the displacement $x(t)$ of a mass m is given by a linear equation with constant coefficients:

$$A_n \frac{d^n x}{dt^n} + \cdots + A_l \frac{dx}{dt} + A_0 x = B_m \frac{d^m f}{dx^m} + \cdots + B_l \frac{df}{dx} + B_0 f$$

Applying the Fourier transform to both members of this relation, it yields:

$$X(f) \sum_{k=1}^{n} A_k (if)^k = F(f) \sum_{j=1}^{m} B_j (if)^j$$

where $F(t)$ and $X(t)$ are the Fourier transforms of the excitation and the response, respectively. The complex function defined through the relation

$$H(f) = \sum_{j=1}^{m} B_j(if)^j \Bigg/ \sum_{k=1}^{n} A_k(if)^k = |H(f)|e^{i\phi(f)}$$

is called the frequency response function.

The relation $X(f) = H(f) * F(f)$ shows that the response specter is easy to obtain when multiplying the excitation specter by the frequency response specter. The amplitude of the frequency response for each frequency band is the product of the excitation amplitude by the frequency response amplitude. The response phase is the sum of the excitation phase with the response phase. By squaring the relation between amplitudes, the power specter formula is derived:

$$|X(f)|^2 = |F(f)|^2 * |H(f)|^2$$

The frequency response function is most often represented through separating the real member from the imaginary member. For a proportional damping, the real member is zero, and the imaginary member reaches its maximum (Fig. A3.24a). Another way to represent the frequency response function is the Nyquist diagram, by representing the real member in terms of the imaginary member (Fig. A3.24b).

The graphical analysis of these diagrams, as proposed by Kennedy and Pancu as early as 1947 [82], proved to be the most accurate method to determine the dynamic parameters and the vibration mode shapes of a complex structure. The frequency response function is also useful for studying the effects of various excitations upon a mechanical system, for determining the mechanical impedance (if the excitation is a force and the response is a velocity), as well as for removing the effects introduced by signal propagation in order to rebuild the excitation.

A3.3.2.7 Analysis in the Amplitude Field

In general, defects caused by harmful low-speed phenomena (e.g., wear) have a typical way of evolution: initially they show up as local singular defects, which generate impulse excitations (shocks); the frequency of these excitations increases in time and determines the appearance of a few peaks in the frequency specter. In order to monitor and diagnose the signal in such situations, an analysis in the amplitude field is indicated,

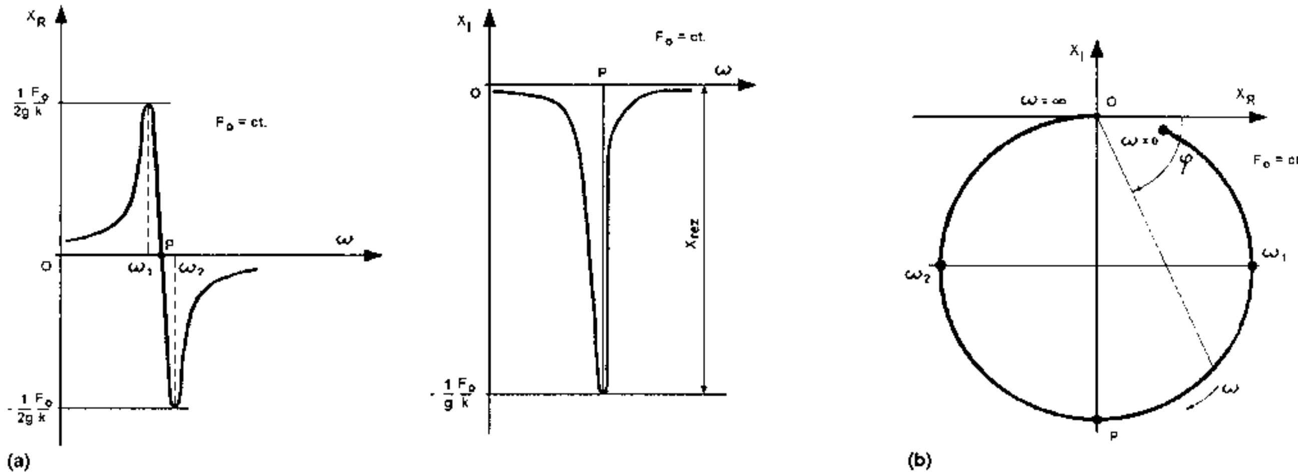

FIGURE A3.24 Frequency response function: (a) for a proportional damping, the real member is zero, and the imaginary member reaches its maximum; (b) another way to represent the frequency response function is the Nyquist diagram.

by means of two specific functions: the peak factor F_v and the Kurtosis function β_2, defined by the relations:

$$F_v = \frac{X_v}{X_{ef}}$$

$$\beta_2 = \frac{1}{\sigma^4}\int_{-\infty}^{\infty}(X - \overline{X})^4 p(x)dx$$

where X_v is the amplitude of the peak signal, X_{ef} is the effective value (mean square) of the signal with the amplitude X, $\overline{X}$ is the average value of the signal's amplitude, $p(x)$ is the probability density of the amplitude, and σ is the amplitude dispersion.

The peak factor is utilized especially in the amplitude analysis of determinist signals; the Kurtosis function can be used for random signals, too. For different situations, the Kurtosis function is given in tables in the literature. Note that, for machines and equipment in normal operating status, this function is 3, which is considered a reference value. An increase of this value in time, determined by the appearance of impulse-type processes, shows the appearance and further evolution of a defect.

A3.3.3 Acoustic Emission Analysis

More methods are available for acoustic emission study, depending on the type of the selected signal. Therefore, a general evaluation is first necessary.

A3.3.3.1 Method of Counting Impulses

This method is useful for impulse-type signals. An evaluation of impulses that pass over a previously set threshold value is done. For the evaluation, the measuring chain consists of a device that differentiates the amplitude level of impulses, followed by an impulse counter (per time unit or per total).

The simple counting of impulses (Fig. A3.25c) can be improved through an evaluation of the impulse area (Fig. A3.25a), in order to take into consideration the duration of the impulse, possibly by introducing a combination of thresholds (Fig. A3.25b).

A3.3.3.2 Method of Impulse Amplitude Mediation

This method is utilized when the acoustic emission shows up through continuous signals. Calculation of the effective value (RMS) is significant

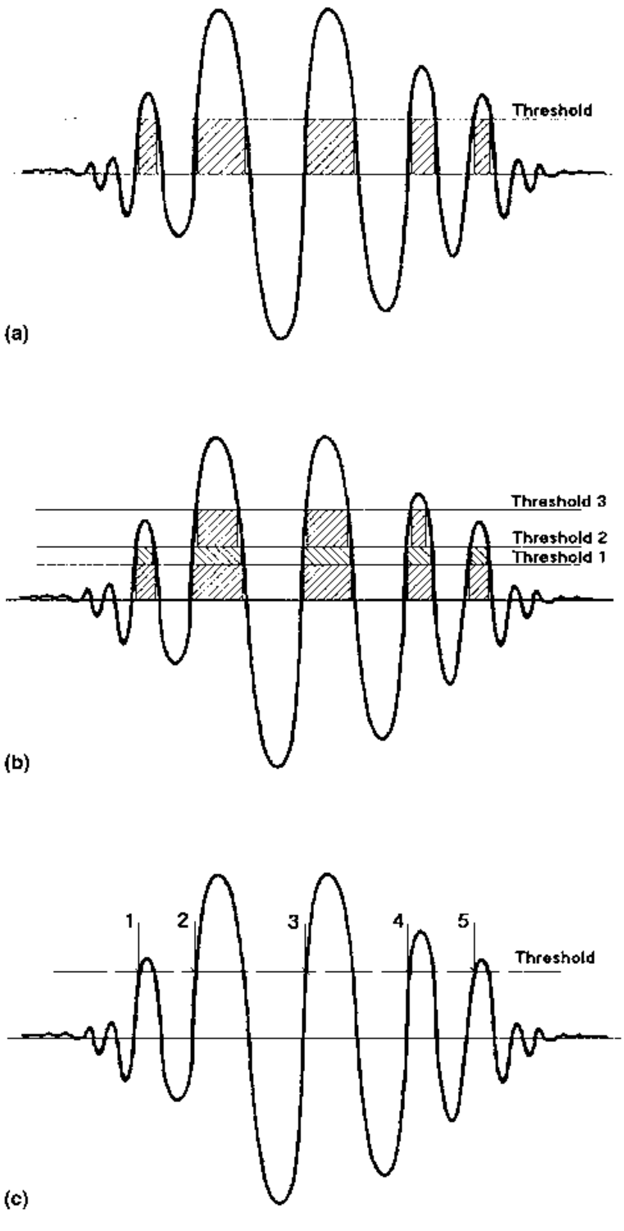

FIGURE A3.25 Evaluation of impulses: (a) Evaluation of impulse area; (b) combination of thresholds; (c) simple counting of impulses.

because, as shown in Section A3.2.2.1, this value is proportional to the signal's power/energy. This method is sometimes called the energetic analysis of impulses.

A3.3.3.3 Method of Source Localization

This method provides information about characteristics and changes of the source and of the wave trajectory. Controlled elastic waves are cre-

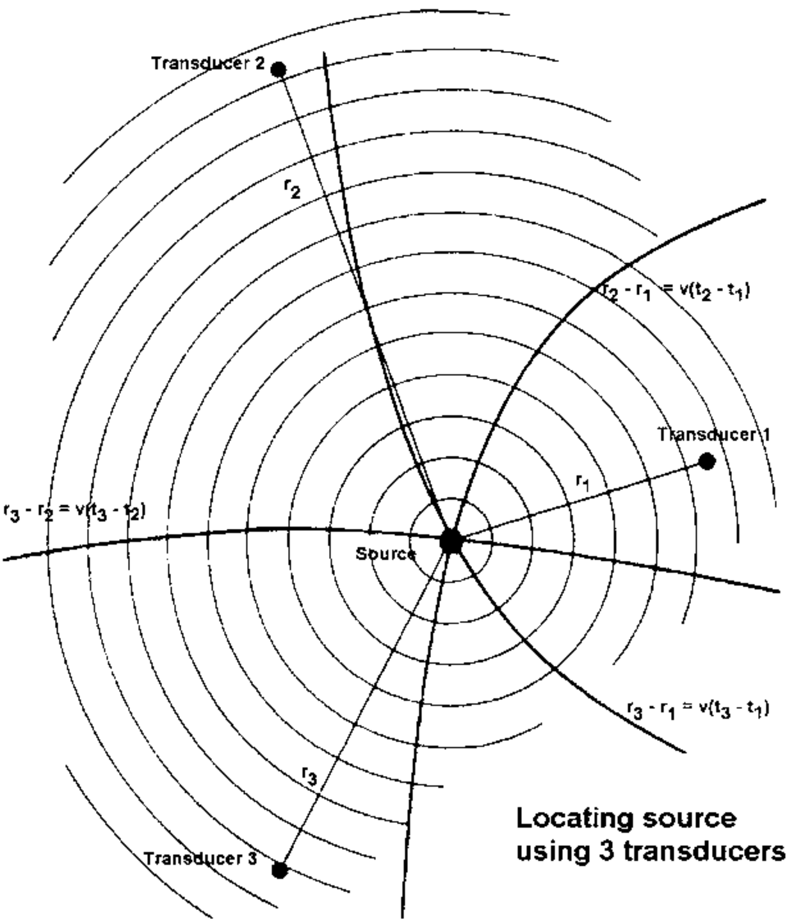

FIGURE A3.26 Locating a source using two transducers.

ated by means of ultrasound sources. The ultrasonic wavetrain is characterized by a factor of the simultaneous wave, which depends on the succession of impulse frequencies, impulse duration, and the number of peaks that exceed a certain threshold. In this way, fracture areas in composite materials can be localized in advance.

In Figure A3.26, the way in which a source is localized is presented, in two dimensions, by means of two transducers. The difference between the arrival time of the signal at two transducers determines a plane hyperbola, provided the propagation velocity of the signal is known. The intersection of the hyperbolas obtained from the transducer pairs (1,2), (1,3), and (2,3) defines the real location of the source. This method is extremely useful for diagnosing, for localizing the source of primary defects, and for reducing the time to restore the operational status.

A3.3.4 CONCLUSIONS

A3.3.4.1 Requirements for Diagnostic Systems

Usage of diagnostic systems should be related to the importance of the system monitored within a fabrication process, taking into account its complexity and performance. Usually, a monitoring process is conducted for complex or continuously working equipment.

For efficiency in monitoring and detection of defects, the analysis and diagnostic have to respect a minimum requirement set:

Must be adequate for the monitored equipment and have the ability and enough accuracy to detect defects.

Notification of defects must be in due time; false alarms must be avoided.

Must allow localization of defects in order to minimize the intervention time for repairs.

Must ensure, as much as possible, a correlation of parameters that accompany the equipment work (vibration, noise, temperature, pressure, etc.), in order to gather complete and correct information.

Must be easy to use and must not raise maintenance problems.

Must be resistant to shipping and handling, to dust, moisture, and industrial liquids, to low and high temperatures, and sometimes to radiation.

Must not be supplied from the same energy sources as the monitored system, but from special, stabilized, protected sources.

The reliability of the diagnostic system must be clearly higher than the reliability of the monitored system.

The cost of the diagnostic system and its installation must not exceed 10% of the cost of the monitored system.

A3.3.4.2 Implementation Stages

Finding a technical diagnostic for complex systems is possible only after their dynamics and kinematics are well known. This will lead to determination of factors to be monitored, types of transducers to be used, as well as locations in which those transducers would be installed.

When processing signals, the normal and maximum admissible levels of the monitored parameters should be taken into consideration. The sensitive points of diagnostic systems are: the signal processor, which extracts the necessary data from the raw signal, and the diagnostic processor, which uses these data in order to identify the status of the monitored system and to localize and isolate the defect.