CHAPTER 14
VIBRATION ANALYZERS AND THEIR USE

Robert B. Randall

INTRODUCTION

This chapter deals primarily with frequency analysis, but a number of related analysis techniques—namely, synchronous averaging, cepstrum analysis, and Hilbert transform techniques—are considered.

With the increase in availability of signal processing packages, virtually all of the techniques discussed, and a large number of others, can now be directly programmed by the user on a general purpose computer (see Chap. 27), but dedicated analyzers still have a number of advantages, as follows:

- Dedicated hardware for preprocessing signals before they are actually stored in the analyzer's memory. This includes real-time zoom with decimation to a lower sampling frequency (vastly reducing the amount of data to be stored), real-time digital resampling for order analysis, and even something as trivial as real-time triggering. If the data only has to be processed after the occurrence of some event that can be used as a trigger, the latter can avoid the storage and postprocessing of vast amounts of useless data.
- Fractional octave digital filter analyzers decimate the sampling frequency of low-frequency signal components as part of their operation. If the equivalent analysis over three frequency decades were to be carried out by postprocessing of an already digitized signal, approximately one million samples would be required to obtain a single one-twelfth-octave spectrum with sufficient averaging for a random signal.
- Dedicated analyzers are more likely to provide error-free results in terms of correct scaling as rms spectra, power spectra, power spectral density, or energy spectral density, while compensating for the data windows used. They also often indicate if insufficient averaging has been used for random signals, etc.

Most frequency analysis is now done digitally, using the FFT (fast Fourier transform) for constant bandwidth analysis on a linear frequency scale, and recursive digital filters for constant percentage bandwidth (fractional octave) analysis on a logarithmic frequency scale; since the latter behave essentially in the same way as analog filters, the chapter starts with a general discussion of filters and their use for

frequency analysis, and later covers FFT analysis. Although spectrum analysis can be done in other ways, such as autoregressive (AR) analysis, moving average (MA) analysis, and their combination (ARMA analysis), these methods are not yet incorporated in spectrum analyzers, and so have not been treated in this chapter.

ELECTRICAL FILTERS

An ideal bandpass filter is a circuit which transmits that part of the input signal within its passband and completely attenuates components at all other frequencies. Practical filters differ slightly from the ideal, as discussed below. An analysis may be performed over a frequency range either by using a single filter with a tunable center frequency which is swept over the entire frequency range or by using banks of fixed filters having contiguous (or overlapping) passbands.

For general vibration analysis it used to be common to use tunable filters whose center frequency was either tuned by hand or synchronized with the X position of the pen on a graphic recorder so that the spectrum was plotted automatically by sweeping the center frequency over the desired frequency range. Alternatively, the center frequency could be synchronized with an external signal, e.g., a trigger pulse once per revolution of a shaft, in which case the filter became a tracking filter which could be used to filter out the component corresponding to a designated harmonic, or multiple, of the synchronizing signal.

In the past, banks of filters with fixed center frequencies, each with its own detector, were widely used for parallel analysis of all frequency bands in real time. This arrangement is costly, however, and has largely been superseded by digital filter analyzers (described in the following section). If real-time analysis is not required, a less-expensive alternative is to switch the output of each filter in turn to a single detector and record the outputs sequentially on paper or on a display. The individual filters can then also share many components, which are selected in appropriate combinations by the switching process. Sequentially stepped fixed filters are typically used for relatively broad-band analysis and are rarely used with less than one-third-octave bandwidth. This type of analysis finds most application in acoustics and in studies of the effects of vibration on humans (Chap. 42).

Digital Filters. Digital filters (in particular, recursive digital filters) are devices which process a continuous digitized signal and provide another signal as an output which is filtered in some way with respect to the original. The relationship between the output and input samples can be expressed as a difference equation (in general, involving previous output and input values) with properties similar to those of a differential equation which might describe an analog filter. Figure 14.1 shows a typical two-pole section used in a one-third-octave digital filter analyzer (three of these are cascaded to give six-pole filtration).

Two ways of changing the properties of a given digital filter circuit such as that shown in Fig. 14.1 are:

1. For a given sampling frequency, the characteristics can be changed by changing the coefficients of the difference equation. (In the circuit of Fig. 14.1 there are three, effectively defining the resonance frequency, damping, and scaling.)
2. For given coefficients, the filter characteristic is defined only with respect to the sampling frequency. Thus, halving the sampling frequency will halve the cutoff frequencies, center frequencies, and bandwidths; consequently, the constant-

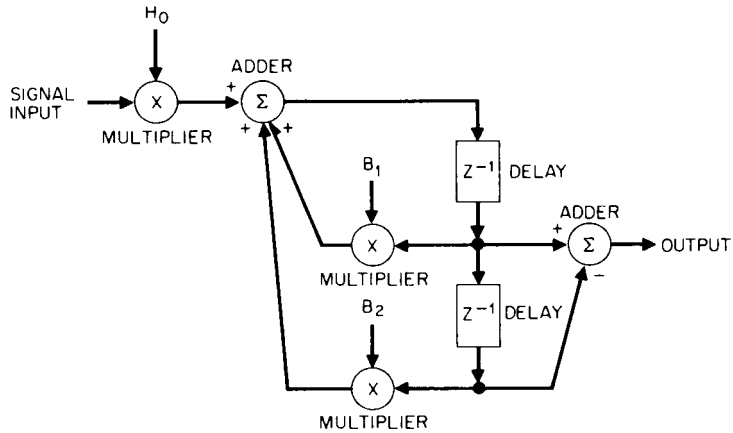

FIGURE 14.1 Block diagram of a typical two-pole digital filter section, consisting of multipliers, adders, and delay units. H_0, B_1, and B_2 are constants by which the appropriate signal sample is multiplied. Z^{-1} indicates a delay of one sample interval before the following operation.

percentage characteristics are maintained one octave lower in frequency. For this reason, digital filters are well adapted to constant-percentage bandwidth analysis on a logarithmic (i.e., octave-based) frequency scale.

Thus, the 3 one-third-octave characteristics within each octave are generated by changing coefficients, while the various octaves are covered by repetitively halving the sampling frequency. Every time the sampling frequency is halved, it means that only half the number of samples must be processed in a given time; the total number of samples for all octaves lower than the highest is $(\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots)$, which in the limit is the same as the number in the highest octave. By being able to calculate twice as fast as is necessary for the upper octave alone, it is possible to cover any number of lower octaves in real time. This is the other reason why digital filters are so well adapted to real-time constant-percentage bandwidth analysis over a wide frequency range.

Filter Properties. Figure 14.2 illustrates what is meant by the 3-dB bandwidth and the effective noise bandwidth, the first being most relevant when separating discrete frequencies, and the second when dealing with random signals. For filters having good selectivity (i.e., having steep filter flanks), there is not a great difference between the two values, and so in the following discussion no distinction is made between them.

The response time T_R of a filter of bandwidth B is on the order of $1/B$, as illustrated in Fig. 14.3, and thus the delay introduced by the filter is also on this order. This relationship can be expressed in the form

$$BT_R \approx 1 \tag{14.1}$$

which is most applicable to constant-bandwidth filters, or in the form

$$bn_r \approx 1 \tag{14.2}$$

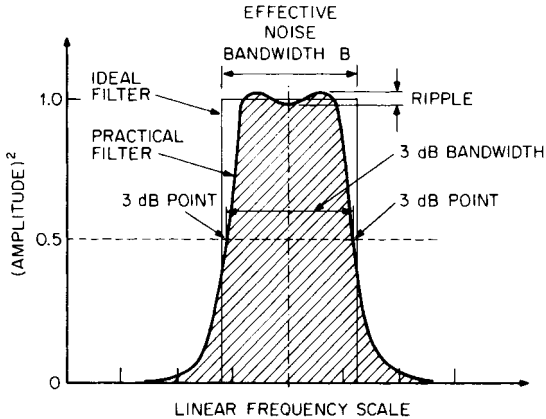

FIGURE 14.2 Bandwidth definitions for a practical filter characteristic. The *3-dB bandwidth* is the width at the 3-dB (half-power) points. The *effective noise bandwidth* is the width of an ideal filter with the same area as the (hatched) area under the practical filter characteristic on an amplitude squared (power) scale.

where $b = B/f_0 =$ relative bandwidth
$n_r = f_0 T_R =$ number of periods of frequency f_0 in time T_R
$f_0 =$ center frequency of filter

This form is more applicable to constant-percentage bandwidth filters. Thus, the response time of a 10-Hz bandwidth filter is approximately 100 milliseconds, while the response time of a 1 percent bandwidth filter is approximately 100 periods. Figure 14.3 also illustrates that the effective length of the impulse T_E is also approximately $1/B$, while to integrate all of the energy contained in the filter impulse response it is necessary to integrate over at least $3T_R$.

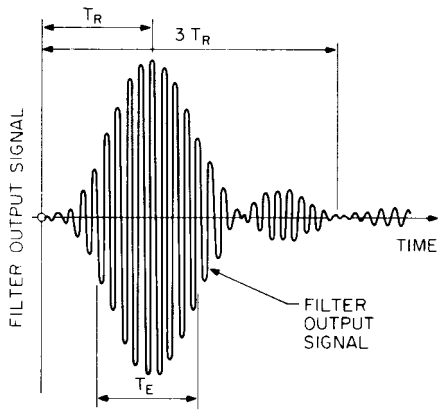

FIGURE 14.3 Typical filter impulse response. T_R = filter-response time ($\approx 1/B$); T_E = effective duration of the impulse ($\approx 1/B$); B = bandwidth.

Choice of Bandwidth and Frequency Scale. In general it is found that analysis time is governed by expressions of the type $BT \geq K$, where K is a constant [see, for example, Eq. (14.1)] and T is the time required for each measurement with bandwidth B. Thus, it is important to choose the maximum bandwidth which is consistent with obtaining an adequate resolution, because not only is the analysis time per bandwidth proportional to $1/B$ but so is the number of bandwidths required to cover a given frequency range—a squared effect.

It is difficult to give precise rules for the selection of filter bandwidth, but the following discussion provides some general guidelines: For stationary deterministic and, in particular, periodic signals containing equally spaced discrete frequency components, the aim is to separate adjacent components; this can best be done using a constant bandwidth on a linear frequency scale. The bandwidth should, for example, be chosen as one-fifth to one-third of the minimum expected spacing (e.g., the lowest shaft speed, or its half-order if this is to be expected) (see Fig. 14.4*A*). For sta-

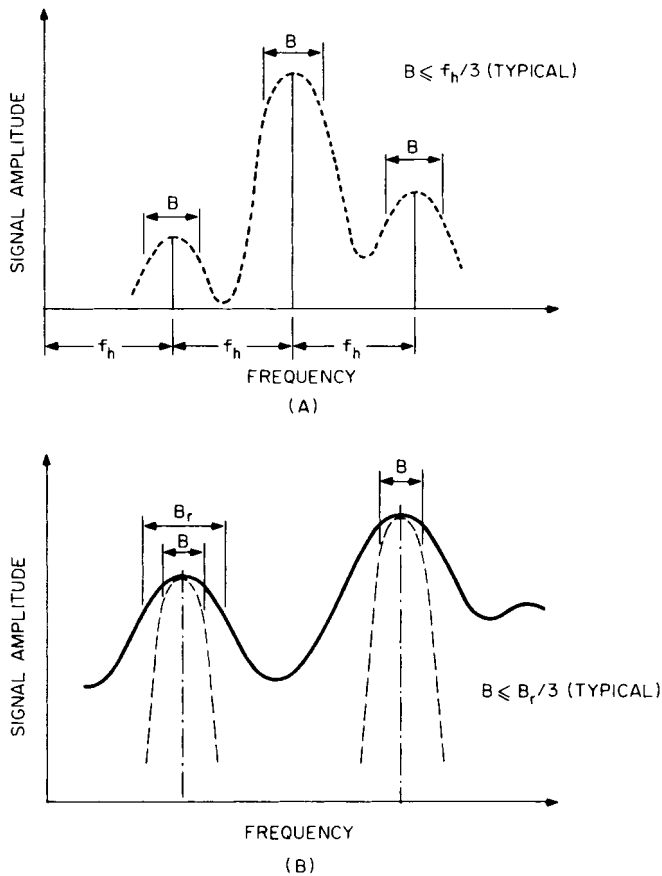

FIGURE 14.4 Choice of filter bandwidth B for different types of signals. (*A*) Discrete frequency signals—harmonic spacing f_h. (*B*) Stationary random and impulsive signals.

tionary random or transient signals, the shape of the spectrum will most likely be determined by resonances in the transmission path between the source and the pickup, and the bandwidth B should be chosen so that it is about one-third of the bandwidth B_r of the narrowest resonance peak (Fig. 14.4*B*). For constant damping these tend to have a constant Q or constant-percentage bandwidth character, and thus constant-percentage bandwidth on a logarithmic frequency scale often is most appropriate.

A linear frequency scale is normally used together with a constant bandwidth, while a logarithmic frequency scale is normally used together with a constant-percentage bandwidth, as each combination gives uniform resolution along the scale. A logarithmic scale may be selected in order to cover a wide frequency range, and then a constant-percentage bandwidth is virtually obligatory. A logarithmic frequency scale may, however, occasionally be chosen in conjunction with a constant bandwidth (though over a limited frequency range) in order to demonstrate a relationship which is linear on log-log scales (e.g., conversions between acceleration, velocity, and displacement).

Choice of Amplitude Scale. Externally measured vibrations, on a machine casing for example, are almost always the result of internal forces acting on a structure whose frequency response function modifies the result. Because the structural response functions vary over a very wide dynamic range, it is almost always an advantage to depict the vibration spectra on a logarithmic amplitude axis. This applies particularly when the vibration measurements are used as an indicator of machine condition (and thus, internal forces and stresses) since the largest vibration components by no means necessarily represent the largest stresses. Even where the vibration is of direct interest itself, in vibration measurements on humans, the amplitude axis should be logarithmic because this is the way the body perceives the vibration level.

It is a matter of personal choice (though sometimes dictated by standards) whether the logarithmic axes are scaled directly in linear units or in logarithmic units expressed in decibels (dB) relative to a reference value. Another aspect to be considered is dynamic range. The signal from an accelerometer (plus preamplifier) can very easily have a valid dynamic range of 120 dB (and more than 60 dB over three frequency decades when integrated to velocity). The only way to utilize this wide range of information is on a logarithmic amplitude axis. Figure 14.5 illustrates both these considerations; it shows spectra measured at two different points on the same gearbox (and representing the same internal condition) on both logarithmic and linear amplitude axes. The logarithmic representations of the two spectra are quite similar, while the linear representations are not only different but hide a number of components which could be important.

An exception where a linear amplitude scale usually is preferable to a logarithmic scale is in the analysis of relative displacement signals, measured using proximity probes, for the following reasons: (1) The parameter being measured is directly of interest for comparison with the results of rotor dynamic and bearing hydrodynamic calculations. (2) The dynamic range achievable with relative shaft vibration measurements (as limited by mechanical and electrical runout) does not justify or necessitate depiction on a logarithmic axis.

Analysis Speed. There are three basic elements in a filter analyzer which can give rise to significant delays and thus influence the speed of analysis.

The *filter* introduces a delay on the same order as its response time T_R (see Fig. 14.3). This is most likely to dominate in the analysis of stationary deterministic signals, where the filter contains only one discrete frequency component at a time and only a short averaging time is required.

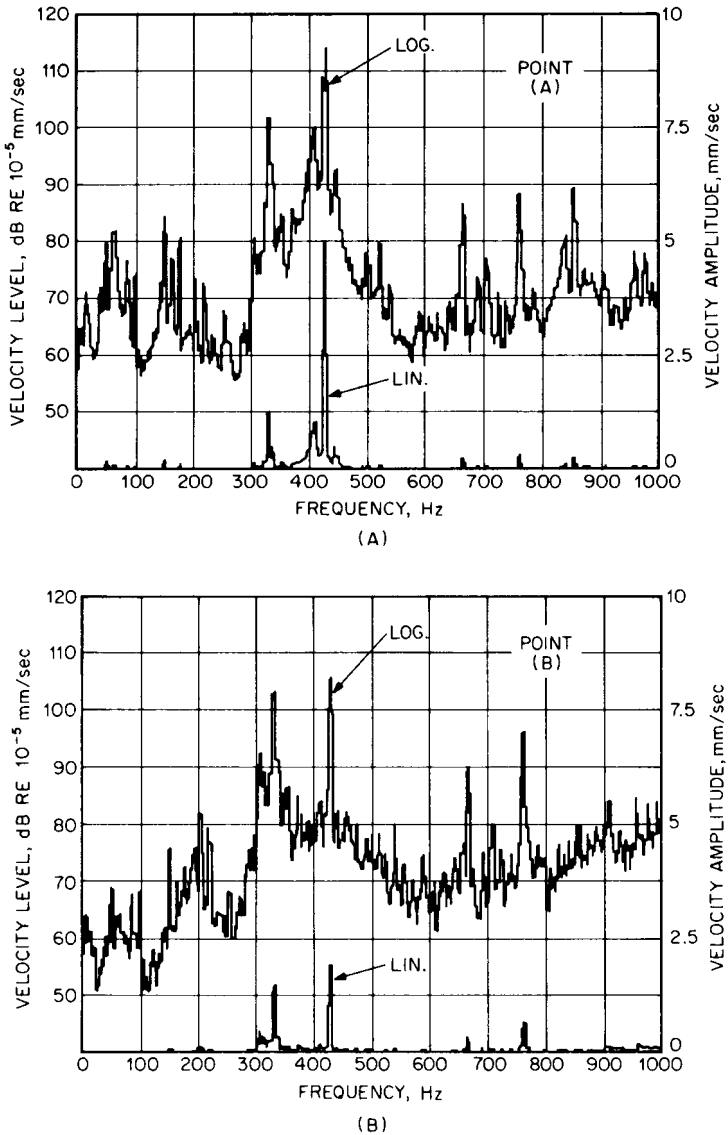

FIGURE 14.5 Comparison of rms logarithmic and rms linear amplitude scales for the depiction of vibration velocity spectra from two measurement points [(*A*) and (*B*)] on the same gearbox (thus representing the same internal condition). The logarithmic representations in terms of velocity level are similar and show all components of interest. The linear spectra in terms of velocity amplitude are quite different, and both hide many components which could be important.

The *detector* introduces a delay on the same order as the averaging time T_A. The choice of averaging time depends on the type of signal being analyzed, namely, stationary deterministic (discrete frequency) or stationary random.

Choice of Averaging Time. For deterministic signals, made up entirely of discrete frequency components, the minimum averaging time required when there is only one component in the filter passband (e.g., for a one-third-octave filter containing the first, second, or third harmonic of shaft speed) comprises three periods of this frequency. However, since a result is obtained only after the filter response time ($1/B$) the averaging time should be set at least equal to this for exponential averaging, or double that value for linear averaging. When a filter contains two to five discrete frequencies (e.g., a one-third-octave filter in the range from the fourth to the twentieth harmonic of shaft speed) there will possibly be a beat frequency equal to the difference between adjacent components (i.e., the shaft speed), and an averaging time five times the beat period (reciprocal of the beat frequency) will be required to smooth the result. In theory, the same applies with more components in the passband (e.g., a one-third-octave filter at higher frequencies), but the bandpassed signal will then resemble a pseudo-random signal, and can be treated as a truly random signal for analysis. If a single frequency component dominates a higher frequency band (e.g., a gearmesh frequency without sidebands), it is possible to revert to the requirement given above for a single component.

For random signals, it is necessary to limit the standard deviation of the error to an acceptable value. The standard deviation of the error is given by the formula:

$$\varepsilon = \frac{1}{2\sqrt{BT_A}} \tag{14.3}$$

where B is the filter bandwidth, and T_A the averaging time. This error corresponds to approximately 1 dB when the BT_A product is 16. To halve the error, the averaging time must be increased by a factor of 4, etc.

Table 14.1 summarizes the above information; further detailed information is given in Ref. 7.

Scaling and Calibration for Stationary Signals. *Scaling* is the process of determining the correct units for the Y axis of a frequency analysis, while *calibration* is the process of setting and confirming the numerical values along the axis. In the most general case, spectra can be scaled in terms of mean-square or rms values at each frequency (or, strictly speaking, for each filter band). For signals dominated by discrete frequency components, with no more than one component per filter band, this yields the mean-square or rms value of each component.

TABLE 14.1 Choice of Averaging Time for Filter Analysis of Stationary Signals

	Signal type			
	Deterministic—1 component in band	Deterministic—2–5 components in band	Deterministic—>5 components in band	Random*
Averaging time T_A	$T_A > 3/f_1 +$	$T_A > 5/f_{beat} +$		$T_A > 16/B$
Exponential	$T_A > 1/B$	$T_A > 1/B$	Treat as random	Ditto
Linear	$T_A > 2/B$	$T_A > 2/B$		Ditto

Legend: f_1 = single frequency in band, f_{beat} = minimum beat frequency in band, B = filter bandwidth.
* for error s.d. ≈ 1 dB.

A spectrum of mean-square values is known as a *power spectrum* since physical power often is related to the mean-square value of parameters such as voltage, current, force, pressure, and velocity.

For random signals, the power spectrum values vary with the bandwidth but can be normalized to a *power spectral density* $W(f)$ by dividing by the bandwidth. The results then are independent of the analysis bandwidth, provided the latter is narrower than the width of peaks in the spectrum being analyzed (e.g., following Fig. 14.4*B*). As examples, power spectral density is expressed in g^2 per hertz when the input signal is expressed in *g*s acceleration, and in volts squared per hertz when the input signal is in volts.

The concept of power spectral density is meaningless in connection with discrete frequency components (with infinitely narrow bandwidth); it can be applied only to the random parts of signals containing mixtures of discrete frequency and random components. Nevertheless, it is possible to calibrate a power spectral density scale using a discrete frequency calibration signal. For example, when analyzing a $1g$ sinusoidal signal with a 10-Hz analyzer bandwidth, the height of the discrete frequency peak may be labeled $1^2g^2/10\text{ Hz} = 0.1g^2/\text{Hz}$.

For constant-bandwidth analysis, the scaling thus achieved is valid for all frequencies; for constant-percentage bandwidth analysis, the bandwidth and power spectral density scaling vary with frequency. On log-log axes, it is possible to draw straight lines representing constant power spectral density, which slope upwards at 10 dB per frequency decade from the calibration point.

Real-Time Digital Filter Analysis of Transient Signals. Suppose a digital filter analyzer has a constant-percentage bandwidth (e.g., one-third-octave or one-twelfth-octave) and covers a frequency range of three or four decades. Because the bandwidth varies with frequency, the filter output signal also varies greatly. At low frequencies (where B is small) the filter output resembles its impulse response, with a length dominated by the filter response time T_R. At high frequencies (where T_R is short) the filter output signal follows the input more closely and has a length dominated by T_I, the duration of the input impulse.

This is illustrated in Fig. 14.6, which traces the path of a typical impulsive signal (an *N*-wave) through the complete analysis system of filter, squarer, and averager for both a narrow-band (low-frequency) and a broad-band (high-frequency) filter.

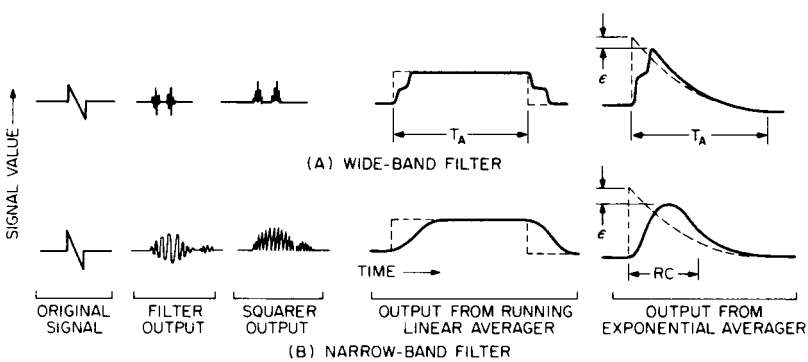

FIGURE 14.6 Passage of a transient signal through an analyzer comprising a filter, squarer, and averager (alternatively running linear averaging and exponential averaging). The dotted curves represent the averager impulse responses. *RC* is the time constant for exponential averaging. ε is the error in peak response. (*A*) With a wide-band filter. (*B*) With a narrow-band filter.

The averaging time T_A must always satisfy

$$T_A \geq T_I + 3T_R \tag{14.4}$$

Thus the averaging time is determined by the lowest frequency to be analyzed. The ideal solution would be running linear integration [with T_A selected using Eq. (14.4)] followed by a maximum-hold circuit (which retains the maximum value experienced). The output of such a running linear averager is shown in Fig. 14.6. Note that during the time the entire filter output is contained within the averaging time T_A, the averager output provides the correct result, which is held by the maximum-hold circuit. However, a running linear average is very difficult to achieve, and normally it is necessary to choose between fixed linear averaging and running exponential averaging.

The problem with fixed linear averaging is that it must be started just before the arrival of the impulse and thus cannot be triggered from the signal itself (unless use is made of a delay line before the analyzer). It is, however, possible to record the signal first and then insert a trigger signal (for example, on another channel of a tape recorder).

In order to extract all the information from a given signal, it may be necessary to make the total analysis in two passes. For example, Fig. 14.7 shows the analysis of a 220-millisecond *N*-wave (the pressure signal from a sonic boom). For an averaging time $T_A = 0.5$ sec, the spectrum is valid only down to about 50 Hz, but it includes frequency components up to 5 kHz. This illustration also shows an analysis of the same signal using $T_A = 8$ sec; this is valid down to about 1.6 Hz. However, as a result of this longer averaging time, there is a 12-dB loss of dynamic range, and so all the frequency components above 500 Hz are lost. The result (with scaling adjusted by 12 dB) is given as a dotted line in Fig. 14.7; it shows that the two spectra are identical over the mutually valid range.

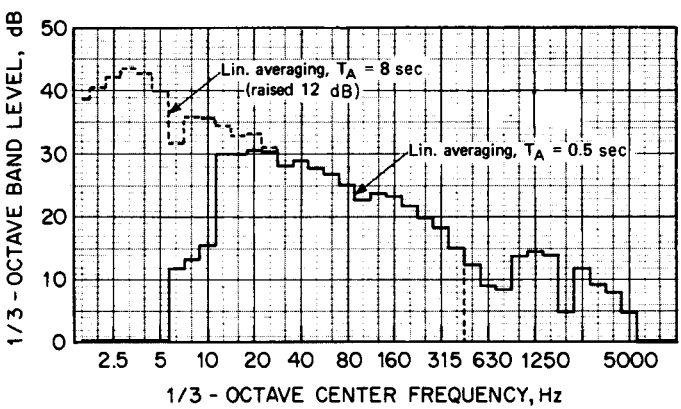

FIGURE 14.7 Transient analysis of a sonic boom (length 218 milliseconds) using a one-third-octave digital filter analyzer. T_A = selected averaging time. The dotted curve (T_A = 8 sec) has been raised 12 dB to compensate for the longer averaging time.

Where the analysis is carried out in real time on randomly occurring impulses, exponential averaging may be used followed by a maximum-hold circuit, but then there is the added complication that the averager leaks energy at a (maximum) rate

of 8.7 dB per averaging time T_A, and thus the total impulse duration must be short with respect to T_A. The error is less than 0.5 dB if

$$T_A \geq 10(T_I + T_R) \tag{14.5}$$

Note that the peak output of an exponential averager is a factor of 2 (i.e., 3 dB) higher than that of the equivalent linear averager (Figs. 13.7 and 14.6); thus the equivalent averaging time to be used in converting from power to energy units is $T_A/2$ for exponential averaging and T_A for linear averaging. Conversion from energy to energy spectral density is valid only for that part of the spectrum where the analyzer bandwidth is appreciably less than the signal bandwidth, although outside that range the results may be interpreted as the mean energy spectral density in the band.

FFT ANALYZERS

FFT analyzers make use of the FFT (fast Fourier transform) algorithm to calculate the spectra of blocks of data. The FFT algorithm is an efficient way of calculating the discrete Fourier transform (DFT). As described in Chap. 22, this is a finite, discrete approximation of the Fourier integral transform. The equations given there for the DFT assume real-valued time signals [see Eqs. (22.26)]. The FFT algorithm makes use of the following versions, which apply equally to real or complex time series:

$$X(m) = \Delta t \sum_{n=0}^{N-1} x(n\,\Delta t) \exp(-j2\pi m\,\Delta f\, n\,\Delta t) \tag{14.6}$$

$$x(n) = \Delta f \sum_{m=0}^{N-1} X(m\,\Delta f) \exp(j2\pi m\,\Delta f\, n\,\Delta t) \tag{14.7}$$

These equations give the spectrum values $X(m)$ at the N discrete frequencies $m\,\Delta f$ and give the time series $x(n)$ at the N discrete time points $n\,\Delta t$.

Whereas the Fourier transform equations are infinite integrals of continuous functions, the DFT equations are finite sums but otherwise have similar properties. The function being transformed is multiplied by a rotating unit vector $\exp(\pm j2\pi m\,\Delta f\, n\,\Delta t)$, which rotates (in discrete jumps for each increment of the time parameter n) at a speed proportional to the frequency parameter m. The direct calculation of each frequency component from Eq. (14.5) requires N complex multiplications and additions, and so to calculate the whole spectrum requires N^2 complex multiplications and additions.

The FFT algorithm factors the equation in such a way that the same result is achieved in roughly $N \log_2 N$ operations.[1] This represents a speedup by a factor of more than 100 for the typical case where $N = 1024 = 2^{10}$. However, the properties of the FFT result are the same as those of the DFT.

Inherent Properties of the DFT. Figure 14.8 graphically illustrates the differences between the DFT and the Fourier integral transform.

Because the spectrum is available only at discrete frequencies $m\,\Delta f$ (where m is an integer), the time function is implicitly periodic (as for the Fourier series). The periodic time

$$T = N\,\Delta t = 1/\Delta f \tag{14.8}$$

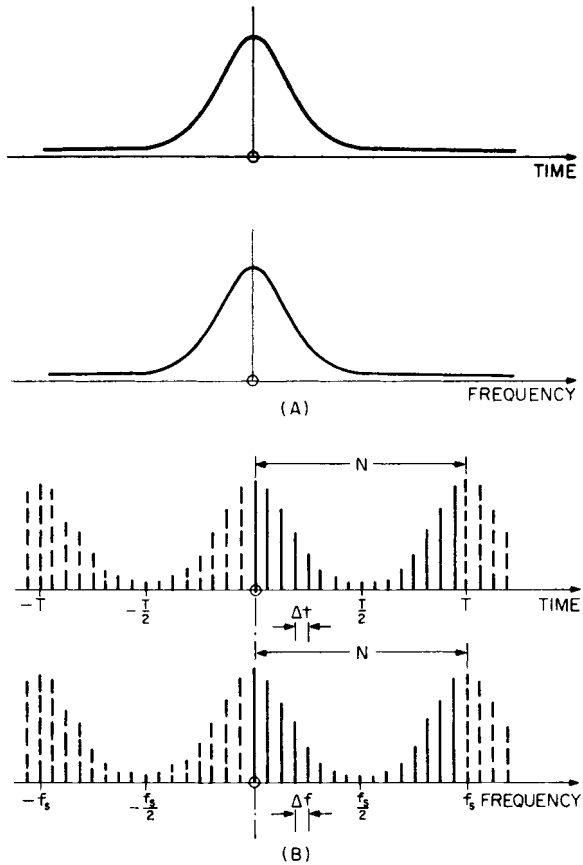

FIGURE 14.8 Graphical comparison of (*A*) the Fourier transform with (*B*) the discrete Fourier transform (DFT) (see text). Note that for purposes of illustration, a function has been chosen (Gaussian) which has the same form in both time and frequency domains.

where $N =$ number of samples in time function and frequency spectrum
$T =$ corresponding record length of time function
$\Delta t =$ time sample spacing
$\Delta f =$ frequency line spacing $= 1/T$

In an analogous manner, the discrete sampling of the time signal means that the spectrum is implicitly periodic, with a period equal to the sampling frequency f_s, where

$$f_s = N\,\Delta f = 1/\Delta t \tag{14.9}$$

Note from Fig. 14.8 that because of the periodicity of the spectrum, the latter half ($m = N/2$ to N) actually represents the negative frequency components ($m = -N/2$ to 0). For real-valued time samples (the usual case), the negative frequency components are determined in relation to the positive frequency components by the equation

$$X(-m) = X^*(m) \tag{14.10}$$

and the spectrum is said to be *conjugate even.*

In the usual case where the $x(n)$ are real, it is only necessary to calculate the spectrum from $m = 0$ to $N/2$, and the transform size may be halved by one of the following two procedures:

1. The N real samples are transformed as though representing $N/2$ complex values, and that result is then manipulated to give the correct result.[2]
2. A zoom analysis (discussed in a later section) is performed which is centered on the middle of the base-band range to achieve the same result.

Thus, most FFT analyzers produce a (complex) spectrum with a number of spectral lines equal to half the number of (real) time samples transformed. To avoid the effects of aliasing (see next section), not all the spectrum values calculated are valid, and it is usual to display, say, 400 lines for a 1024-point transform or 800 lines for a 2048-point transform.

Aliasing. *Aliasing* is an effect introduced by the sampling of the time signal, whereby high frequencies after sampling appear as lower ones (as with a stroboscope). The DFT algorithm of Eq. (14.6) cannot distinguish between a component which rotates, say, seven-eighths of a revolution between samples and one which rotates a negative one-eighth of a revolution. Aliasing is normally prevented by low-pass filtering the time signal before sampling to exclude all frequencies above half the sampling frequency (i.e., $-N/2 < m < N/2$). From Fig. 14.8 it will be seen that this removes the ambiguity. In order to utilize up to 80 percent of the calculated spectrum components (e.g., 400 lines from 512 calculated), it is necessary to use very steep antialiasing filters with a slope of about 120 dB/octave.

Normally, the user does not have to be concerned with aliasing because suitable antialiasing filters automatically are applied by the analyzer. One situation where it does have to be allowed for, however, is in tracking analysis (discussed in a following section) where, for example, the sampling frequency varies in synchronism with machine speed.

Leakage. *Leakage* is an effect whereby the power in a single frequency component appears to leak into adjacent bands. It is caused by the finite length of the record transformed (N samples) whenever the original signal is longer than this; the DFT implicitly assumes that the data record transformed is one period of a periodic signal, and the leakage depends on what is actually captured within the time window, or data window.

Figure 14.9 illustrates this for three different sinusoidal signals. In (A) the data window corresponds to an exact integer number of periods, and a periodic repetition of this produces an infinitely long sinusoid with only one frequency. For (B) and (C) (which have a slightly higher frequency) there is an extra half-period in the data record, which gives a discontinuity where the ends are effectively joined into a loop, and considerable leakage is apparent. The leakage would be somewhat less for intermediate frequencies. The difference between the cases of Fig. 14.9*B* and *C* lies in the phase of the signal, and other phases give an intermediate result.

When analyzing a long signal using the DFT, it can be considered to be multiplied by a (rectangular) time window of length T, and its spectrum consequently is convolved with the Fourier spectrum of the rectangular time window,[3] which thus acts like a filter characteristic. The actual filter characteristic depends on how the resulting spectrum is sampled in the frequency domain, as illustrated in Fig. 14.10.

In practice, leakage may be counteracted:

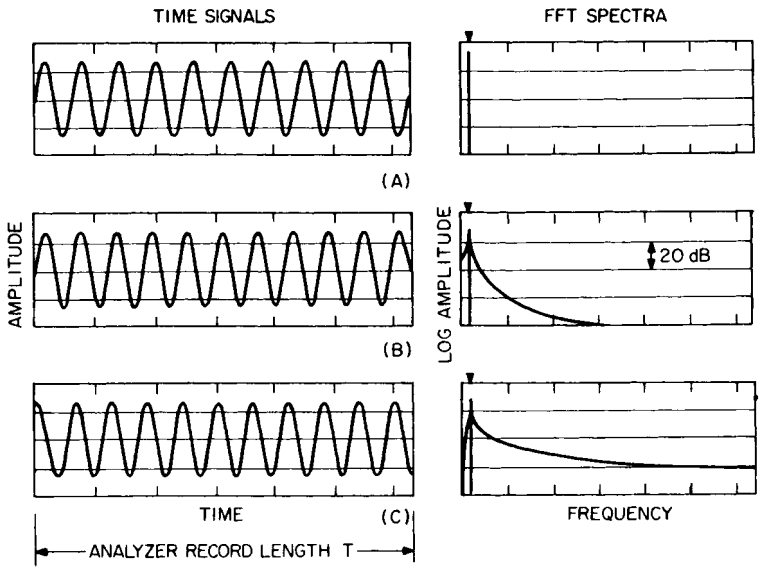

FIGURE 14.9 Time-window effects when analyzing a sinusoidal signal in an FFT analyzer using rectangular weighting. (*A*) Integer number of periods, no discontinuity. (*B*) and (*C*) Half integer number of periods but with different phase relationships, giving a different discontinuity when the ends are joined into a loop.

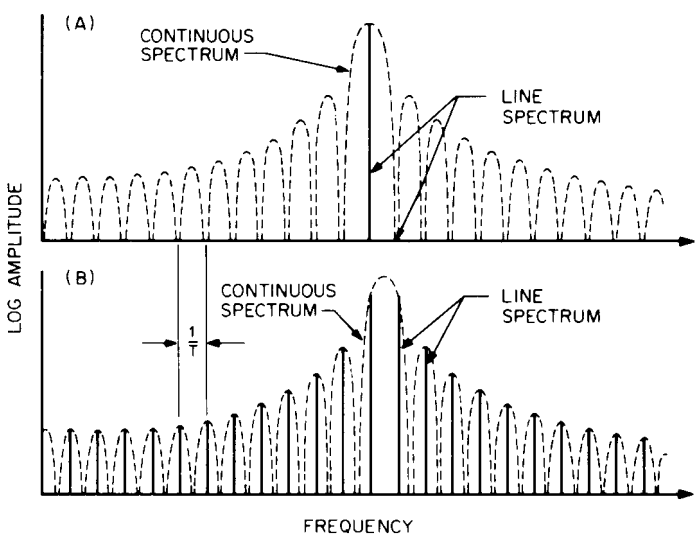

FIGURE 14.10 Frequency sampling of the continuous spectrum of a time-limited sinusoid of length *T*. (*A*) Integer number of periods, side lobes sampled at zero points (compare with Fig. 14.9*A*). (*B*) Half integer number of periods, side lobes sampled at maxima (compare with Fig. 14.9*B* and *C*).

1. By forcing the signal in the data window to correspond to an integer number of periods of all important frequency components. This can be done in tracking analysis (discussed in a later section) and in modal analysis measurements (Chap. 21), for example, where periodic excitation signals can be synchronized with the analyzer cycle.
2. (For long transient signals) By increasing the length of the time window (for example, by zooming) until the entire transient is contained within the data record.
3. By applying a special time window which has better leakage characteristics than the rectangular window already discussed.

Later sections deal with the choice of data windows for both stationary and transient signals.

Picket Fence Effect. The *picket fence effect* is a term used to describe the effects of discrete sampling of the spectrum in the frequency domain. It has two connotations:

1. It results in a nonuniform frequency weighting corresponding to a set of overlapping filter characteristics, the tops of which have the appearance of a picket fence (Fig. 14.11).
2. It is as though the spectrum is viewed through the slits in a picket fence, and thus peak values are not necessarily observed.

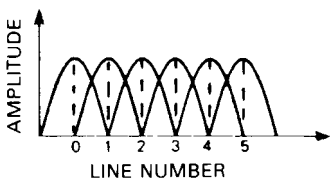

FIGURE 14.11 Illustration of the picket fence effect. Each analysis line has a filter characteristic associated with it which depends on the weighting function used. If a frequency coincides exactly with a line, it is indicated at its full level. If it falls midway between two lines, it is represented in each at a lower level corresponding to the point where the characteristics cross.

One extreme example is in fact shown in Fig. 14.10, where in (*A*) the side lobes are completely missed, while in (*B*) the side lobes are sampled at their maxima and the peak value is missed.

The picket fence effect is not a unique feature of FFT analysis; it occurs whenever discrete fixed filters are used, such as in normal one-third-octave analysis. The maximum amplitude error which can occur depends on the overlap of the adjacent filter characteristics, and this is one of the factors taken into account in the following discussion on the choice of data window.

Data Windows for Analysis of Stationary Signals. A *data window* is a weighting function by which the data record is effectively multiplied before transformation. (It is sometimes more efficient to apply it by convolution in the frequency domain.) The purpose of a data window is to minimize the effects of the discontinuity which occurs when a section of continuous signal is joined into a loop.

For stationary signals, a good choice is the *Hanning window* (one period of a sine squared function), which has a zero value and slope at each end and thus gives a gradual transition over the discontinuity. In Fig. 14.12 it is compared with a rectangular window, in both the time and frequency domains. Even though the main lobe (and thus the bandwidth) of the frequency function is wider, the side lobes fall off much more rapidly and the highest is at –32 dB, compared with –13.4 dB for the rectangular.

Other time-window functions may be chosen, usually with a trade-off between the steepness of filter characteristic on the one hand and effective bandwidth on the other. Table 14.2 compares the time windows most commonly used for stationary

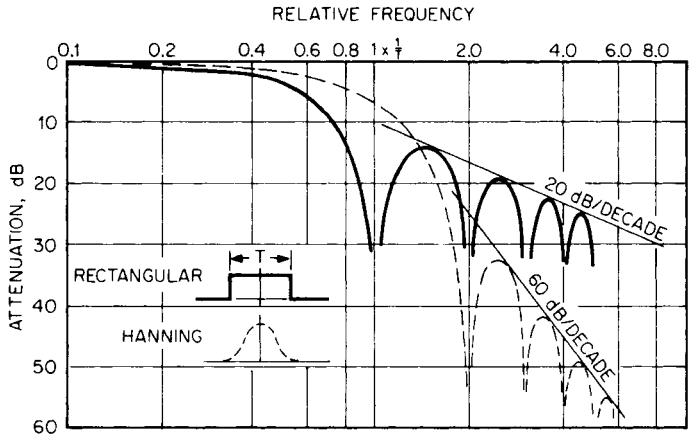

FIGURE 14.12 Comparison of rectangular and Hanning window functions of length T seconds. Full line—rectangular weighting; dotted line—Hanning weighting. The inset shows the weighting functions in the time domain.

signals, and Fig. 14.13 compares the effective filter characteristics of the most important. The most highly selective window, giving the best separation of closely spaced components of widely differing levels, is the *Kaiser-Bessel window.* On the other hand, it is usually possible to separate closely spaced components by zooming, at the expense of a slightly increased analysis time.

Another window, the *flattop window,* is designed specifically to minimize the picket fence effect so that the correct level of sinusoidal components will be indicated, independent of where their frequency falls with respect to the analysis lines. This is particularly useful with calibration signals. Nonetheless, by taking account of the distribution of samples around a spectrum peak, it is possible to compensate for picket fence effects with other windows as well. Figure 14.14, which is specifically for the Hanning window, is a nomogram giving both amplitude and frequency corrections, based on the decibel difference (ΔdB) between the two highest samples around a peak. For stable single-frequency components this allows determination of the frequency to an accuracy of an order of magnitude better than the line spacing.

TABLE 14.2 Properties of Various Data Windows

Window type	Highest side lobe, dB	Side lobe fall-off, dB/decade	Noise bandwidth*	Maximum amplitude error, dB
Rectangular	–13.4	–20	1.00	3.9
Hanning	–32	–60	1.50	1.4
Hamming	–43	–20	1.36	1.8
Kaiser-Bessel	–69	–20	1.80	1.0
Truncated Gaussian	–69	–20	1.90	0.9
Flattop	–93	0	3.70	<0.1

* Relative to line spacing.

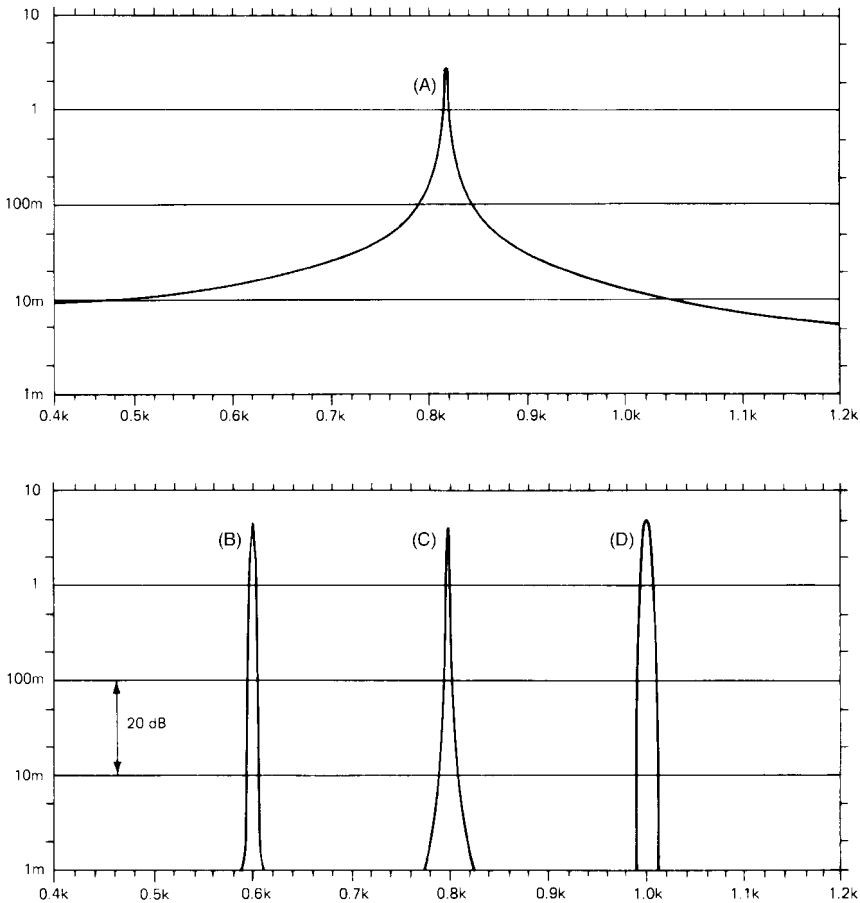

FIGURE 14.13 Comparison of worst-case filter characteristics for rectangular and other weighting functions for an 80-dB dynamic range. (*A*) Rectangular. (*B*) Kaiser-Bessel. (*C*) Hanning. (*D*) Flattop.

Data Windows for Analysis of Transient Signals. When using impulsive (e.g., hammer) excitation of structures for determining their frequency response characteristics (e.g., see Chap. 21), it is common to use the following special data windows:

1. A *short rectangular* window may be applied over the very short excitation impulse in order to exclude noise from the remaining portion of the record.
2. An *exponential* window can be applied where the response is very long (i.e., lightly damped structures) to reduce the signal to practically zero at the end of the record, and thus avoid discontinuities. The effect is the same as adding extra damping which is very precisely known and can be subtracted from the measurement results. A half-Hanning taper is often added to both the leading and trailing

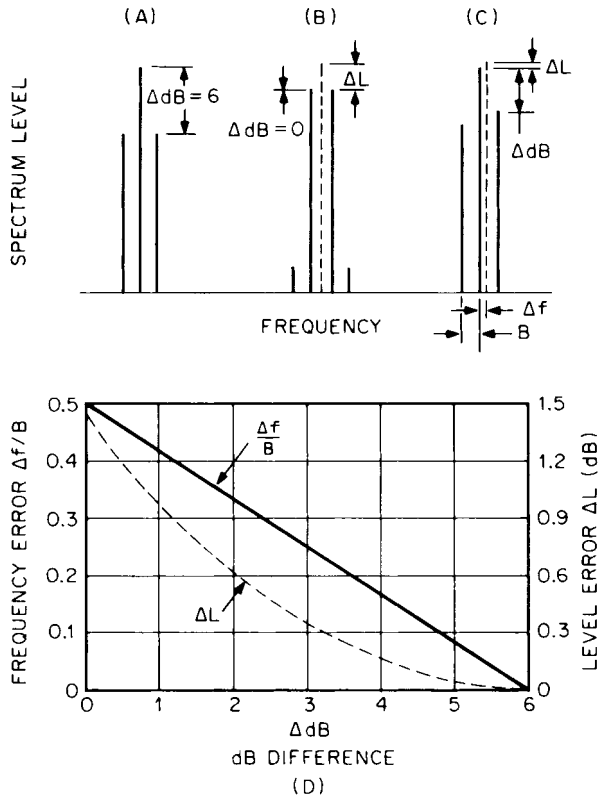

FIGURE 14.14 Picket fence corrections for Hanning weighting, where ΔL = level correction, dB; Δf = frequency correction; Hz; B = line spacing, Hz; ΔdB = difference in decibels between the two highest samples around a peak representing a discrete frequency component. Three examples are shown: (*A*) Actual frequency coincides with center line. (*B*) Actual frequency midway between two lines. (*C*) General situation. Note that the frequency correction $\Delta f/B$ is almost linear.

edges of a short rectangular window, and to the leading edge of an exponential window, to mitigate the effects of the discontinuities.

Zoom Analysis.[4] Zoom analysis is the term given to a spectrum analysis having increased resolution over a restricted part of the frequency range. The following are two techniques used to generate zoom analyses.

1. *Real-time zoom* (illustrated in the block diagram of Fig. 14.15) is a zoom process in which the entire signal is first modified to shift its frequency origin to the center of the zoom range. Then it is passed through a low-pass filter (usually a digital filter in real time) which has a passband corresponding to the original zoom-band (Fig. 14.16). Because of the low-pass filtration, the signal then can be resampled at a lower sampling rate without aliasing, and the resampled signal processed by an FFT

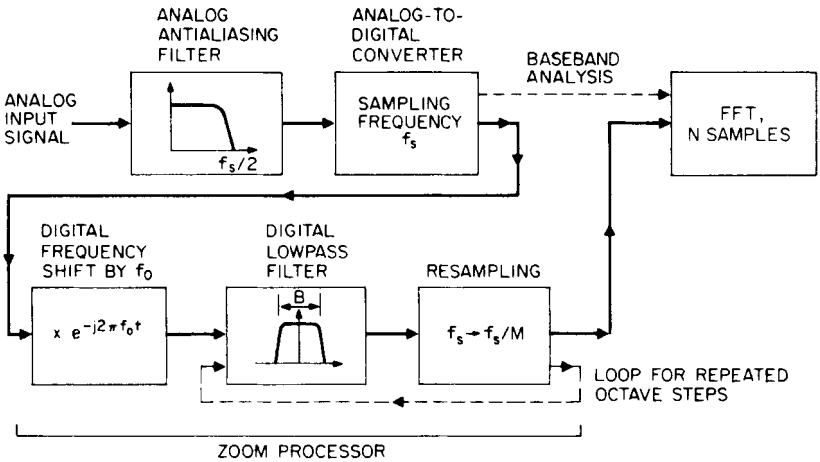

FIGURE 14.15 Block diagram for real-time zoom with bandwidth B centered on frequency f_0. M is the zoom factor and also the factor by which the sampling frequency is reduced.

transform. The original frequency shift is accomplished by multiplying the incoming signal by a unit vector (phasor) rotating at $-f_0$ (thereby subtracting f_0 from all frequencies in it), and the modified time signal is thus complex. This is one situation where the FFT transform of complex data is used. Figure 14.17 gives an example of the use of zoom analysis to show that what appears in a baseband analysis to be the second harmonic of shaft speed actually is dominated by twice the line frequency at 100 Hz and reveals that what appears to be a single-frequency component in a baseband spectrum actually comprises a family of uniformly spaced components, the second highest of which is the second harmonic of the shaft speed.

2. *Nondestructive zoom* is effectively a way of achieving a larger transform size without modification of the original data record. For a typical case, data are first captured in a 10K (i.e., 10,240-point) buffer. Ten 1K records obtained by taking every

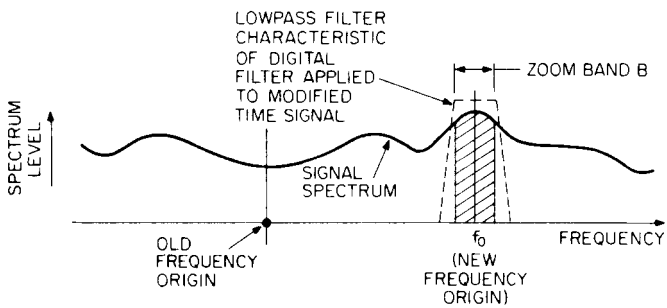

FIGURE 14.16 Principle of real-time zoom, using a low-pass filter to filter out the portion of the original signal in the zoom-band of width B. Prior to this, the frequency origin is shifted to frequency f_0 (the desired center frequency of the zoom-band) by multiplying the (digitized) time signal by $e^{-j2\pi f_0 t}$.

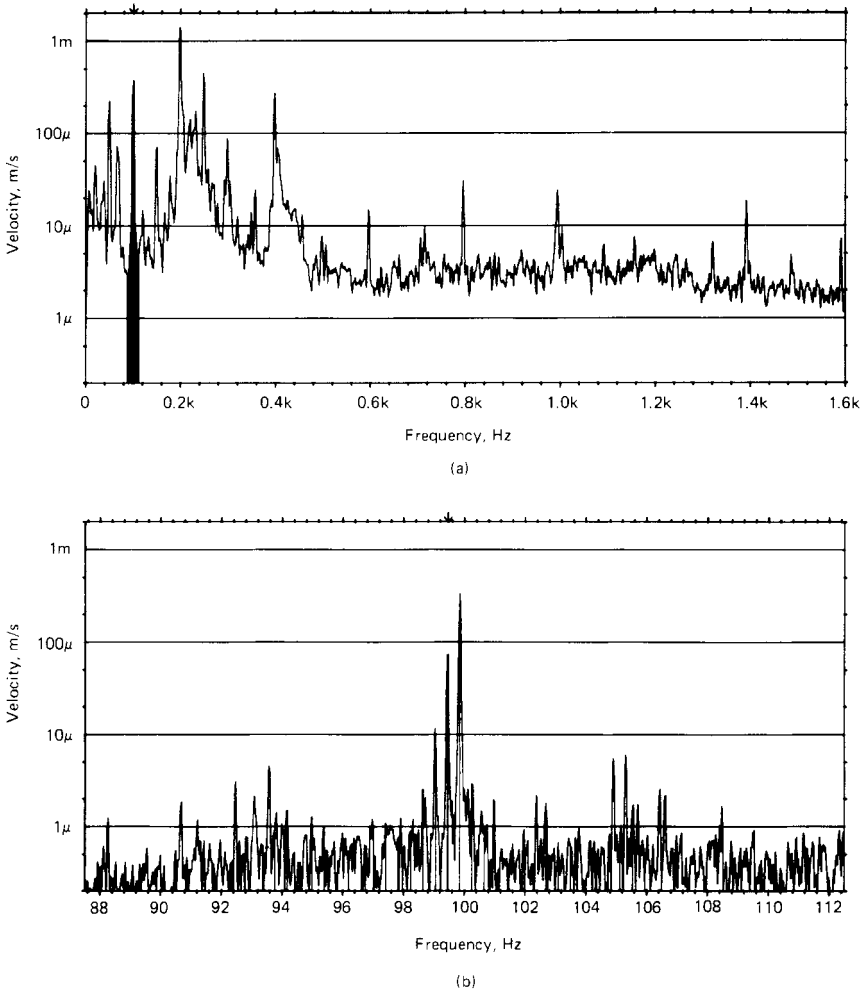

FIGURE 14.17 (*A*) Original baseband spectrum. (*B*) Shaded section of (*A*) zoomed by a factor of 64:1. Highest component at 100 Hz is twice the line frequency. Next highest component on the left is twice the shaft rotational speed.

10th sample are transformed using a 1K (1024-point) FFT transform. Even though this gives only 1024 frequency values per transform, the rest are generated by periodic repetition (because of aliasing resulting from the undersampling). After compensating the phase of the results for the small time shift of each of the undersampled records, the entire spectrum of the 10K record can in principle be obtained by addition. In practice, only a part of the whole spectrum is normally obtained at any one time in order to save on memory requirements, but the whole spectrum can be generated by repetitive operation on exactly the same data. As a result of the decreasing cost of memory and the increasing flexibility of transform size in analyzers, non-

destructive zoom has been effectively superseded by directly performing the larger-size transform.

Real-time zoom has the advantage that the zoom factor obtainable is virtually unlimited. A procedure is often employed (as illustrated in Fig. 14.15) whereby the signal samples are repeatedly circulated around a loop containing a low-pass filter which cuts off at one-half the previous maximum frequency, after which the sampling frequency is halved by dropping every second sample. Each circulation doubles the zoom factor and at the same time doubles the length of original signal required to fill the transform buffer. It is this time requirement which places a limit on the zoom factor, as well as on the stability of the signal itself. A zoom factor of 10 in a 400-line spectrum, for example, gives the equivalent of a 4000-line spectrum; a finer resolution is not required to analyze the vibration spectrum of a machine whose speed fluctuates by, say, 0.1 percent.

Real-time zoom suffers the disadvantage that the entire signal must be reprocessed to zoom in another band. This has two detrimental consequences:

1. For very narrow bandwidths (long record lengths), the analysis time is very long for each zoom analysis.
2. There is no certainty that exactly the same signal is processed each time.

On the other hand, nondestructive zoom (or a large transform) has the advantage that for zoom analysis in different bands, exactly the same data record is used. Thus it is known, for example, that there will be an exact integer relationship between the various harmonics of a periodic signal. This can be useful, as a typical example, in separating the various harmonics of shaft speed from those of line frequency, in induction motor vibrations. Furthermore, the long analysis time is required only once (to fill the data buffer); further zoom analyses on the same record are limited only by the calculation speed.

Nondestructive zoom suffers the disadvantage that the zoom factor is limited by the size of the memory buffer in the analyzer. Where the memory buffer is 10 times the normal transform size, for example, the zoom factor is equal to 10.

Thus, both types of zoom are advantageous for different purposes. Nondestructive zoom is probably best for diagnostic analysis of machine vibration signals, whereas real-time zoom gives more flexibility in frequency response measurements (system frequency response should not change even where the excitation signals change). Real-time zoom also gives the possibility of very large zoom factors when they are required.

In real-time zoom, it is only the preprocessing of the signal which has to be in real time; the actual FFT analysis of the signal, once it is stored in the transform buffer, does not have to be in real time.

ANALYSIS OF STATIONARY SIGNALS USING FFT

Equation (14.8) shows that for a single FFT transform, the product (*bandwidth* times *averaging time*) $BT_A = 1$, at least for rectangular weighting where B is equal to the line spacing Δf (Table 14.2). The same applies for any weighting function, the increased bandwidth being exactly compensated by a corresponding decrease in effective record length.[5]

For *stationary deterministic signals,* a single transform having a BT_A product equal to unity is theoretically adequate, although a small number of averages is

sometimes performed if the signal is not completely stable. Figure 14.18 illustrates the effect of averaging for a deterministic signal and demonstrates that the sinusoidal components are unaffected; the only effect is to smooth out the (nondeterministic) noise at the base of the spectrum (Fig. 14.18*B*).

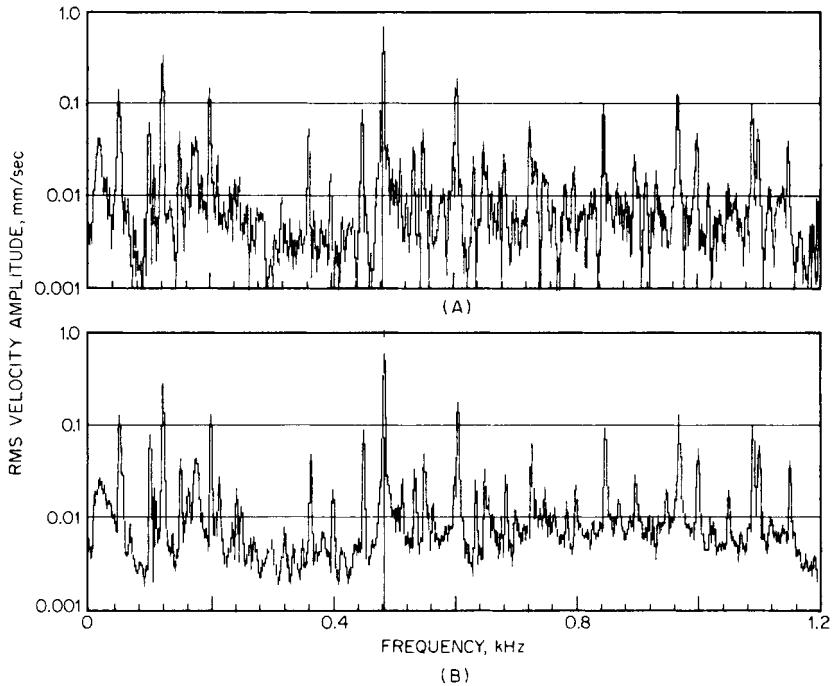

FIGURE 14.18 Effect of averaging with a stationary deterministic signal. (*A*) Instantaneous spectrum (average of 1). (*B*) The linear average of eight spectra.

For *stationary random signals,* the standard deviation of the result of averaging n independent spectra is given by the equivalent of Eq. (14.3), or

$$\varepsilon = \frac{1}{2\sqrt{n}} \tag{14.11}$$

Figure 14.19 illustrates (*A*) an instantaneous spectrum, (*B*) the average of eight spectra, and (*C*) the average of 128 spectra. The meaning of the standard error ε [Eq. (14.11)] is illustrated in (*B*) and (*C*). Statistically, there is a 68 percent probability that the actual error will be less than ε, a 95.5 percent probability that it will be less than 2ε, and a 99.7 percent probability that it will be less than 3ε.

For rectangular (flat) weighting, independent spectra are those from nonoverlapping time records; when other weighting functions are used, the situation is different. For example, Fig. 14.20*A* illustrates the overall (power) weighting obtained when Hanning windows are applied to contiguous records. Note that virtually half of the incoming signal is excluded from the analysis, whereas a 50 percent overlapping of

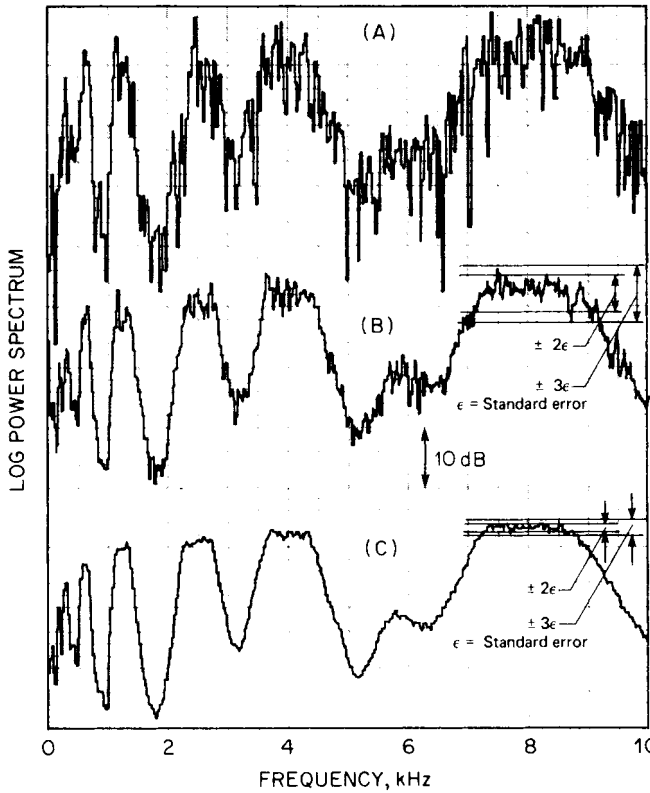

FIGURE 14.19 Effect of averaging with a stationary random signal. (*A*) Instantaneous spectrum. (*B*) Average of eight spectra. (*C*) Average of 128 spectra.

consecutive records regains most of the lost information. Thus, when using window functions similar to Hanning (as recommended for stationary signals), it is almost always advantageous to average the results from 50 percent overlapping records. A method for calculating the effective number of averages obtained in this way is given in Ref. 6; for 50 percent overlapping Hanning windows the error is very small in treating them as independent records.

Real-Time Analysis. An FFT analyzer is said to operate in real time when it is able to process all the incoming data, even though presentation of the results is delayed by an amount corresponding to the calculation time. This implies that the time taken to analyze a data record, T_a, is less than the time taken to collect the data transformed, T. It also implies that the analysis process should not interrupt the continuous recording of data, so that recording can continue in one part of the memory at the same time as analysis is being performed in another. T is inversely proportional to the selected frequency range, and the highest frequency range for which T_a is less than T is called the *real-time frequency.* This condition will ensure that all the incoming data are analyzed only when rectangular weighting is used. With Hanning weighting, for example, where 50 percent overlap analysis must be employed to ana-

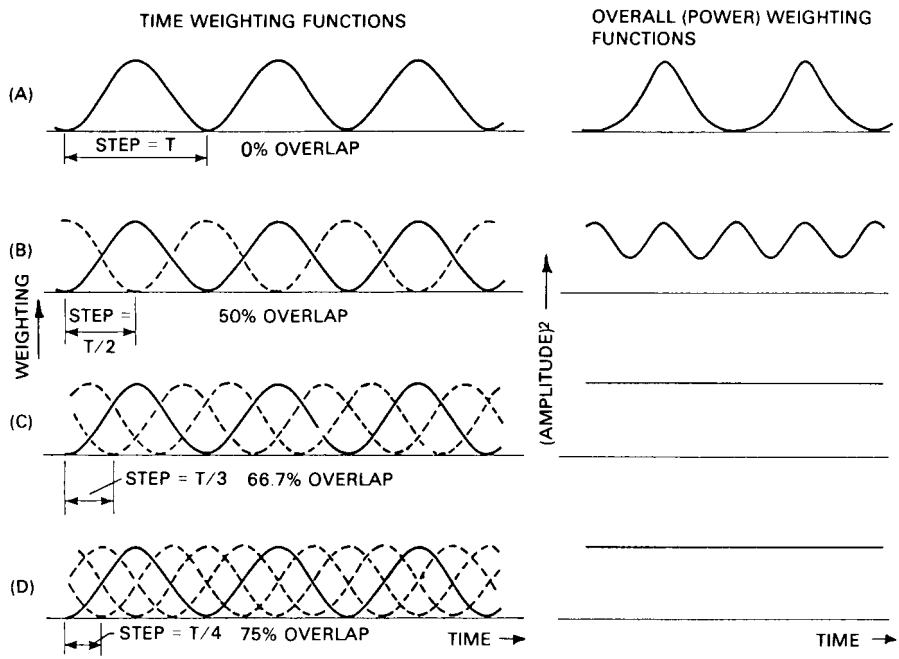

FIGURE 14.20 Overall weighting functions for spectrum averaging with overlapping Hanning windows. (*A*) Zero overlap (step length T). (*B*) 50 percent overlap (step length $T/2$). (*C*) 66.7 percent overlap (step length $T/3$). (*D*) 75 percent overlap (step length $T/4$). T is the record length for the FFT transform.

lyze all the data, the true real-time frequency will be halved, since twice as many transforms must be performed for the same length of data record. In yet another sense, the analysis is not truly real-time unless the overall weighting function is uniform. As illustrated in Fig. 14.20, the minimum overlap of Hanning windows to achieve this is two-thirds, which reduces the true real-time frequency to one-third of the commonly understood definition given above.

In practice, with stationary signals, there is no advantage to more than a 50 percent overlap, since (1) statistical reliability is not significantly improved and (2) all sections of the record are statistically equivalent, so that the overall weighting function is not important. It can be important for nonstationary signals, such as transients, as discussed below. For stationary signals, where any data missed are statistically no different from the data analyzed, the only advantages of real-time analysis are that (1) results with a given accuracy are obtained in the minimum possible time and (2) maximum information is extracted from a record of limited length.

FFT Analysis of Transients. Consider the use of FFT analysis when the entire transient fits into the transform size T without loss of high-frequency information. Figure 14.21 shows such an example where the duration of the transient is less than the analyzer record length of 2048 samples (2K) in a frequency range which does not exclude high-frequency information in the signal. Rectangular weighting should be

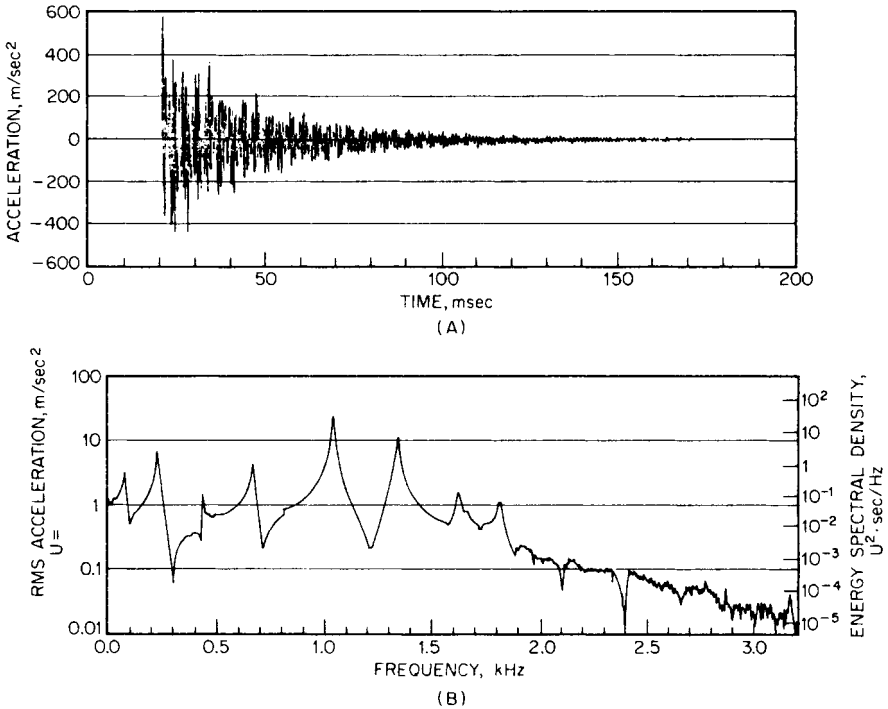

FIGURE 14.21 Example of an FFT analysis of a short transient signal. (*A*) Time signal of length 2048 samples (2K) corresponding to 250 milliseconds ($T = 250$ milliseconds). (*B*) 800-line FFT spectrum with bandwidth $B = \Delta f = 4$ Hz (rectangular weighting). Scaling on left is in rms units. Scaling on right is converted to energy spectral density (ESD) by multiplying mean-square values by T^2.

used in such a case, where the signal value is zero at each end, so that no discontinuity arises from making the record into a loop (an inherent property of the FFT process). Exponential weighting sometimes may be used to force the signal down to zero at the end of the record, but the frequency spectrum will then include the effects of the extra damping which this represents.

With rectangular weighting, the analysis bandwidth is equal to the line spacing $1/T$, which is always less than the effective signal bandwidth. Conversion of the results to energy spectral density, therefore, is valid in most practical situations. Some analyzers provide the results in terms of energy spectral density, but if the results are available only in terms of power (U^2), they must be multiplied by the time T corresponding to the record length to convert them to energy and divided by the bandwidth $1/T$ to convert them to energy spectral density, expressed in engineering units squared times seconds per hertz. Altogether, this represents a multiplication by T^2.

Where a transient is longer than the normal transform size T, it can be analyzed in one of the following ways:

1. *Zoom FFT* (see *Zoom Analysis,* above, for background information). A suitable zoom factor is chosen such that the transform length ($1/\Delta f$) is greater than

the duration of the transient. Thus in the case of nondestructive zoom, the entire transient can be recorded in the long memory and the entire narrow-band spectrum can be obtained by repetitive analysis in contiguous zoom-bands. In the case of real-time zoom, analysis in more than one zoom-band requires that the transient be recorded in an external medium and played back for each zoom analysis. Rectangular weighting should be used (thus $B = \Delta f$) and energy spectral density (as above) is always valid using a value of T corresponding to the zoom record length ($1/B$). The narrow bandwidth may give a restriction of dynamic range of the result. Figure 14.22 shows a typical energy spectrum, obtained by repetitive nondestructive zoom analysis; the same result would be obtained from a single large transform.

2. *Scan averaging.* When the entire transient is stored in digital form in a long memory (as for nondestructive zoom), it is possible to obtain its spectrum by scanning

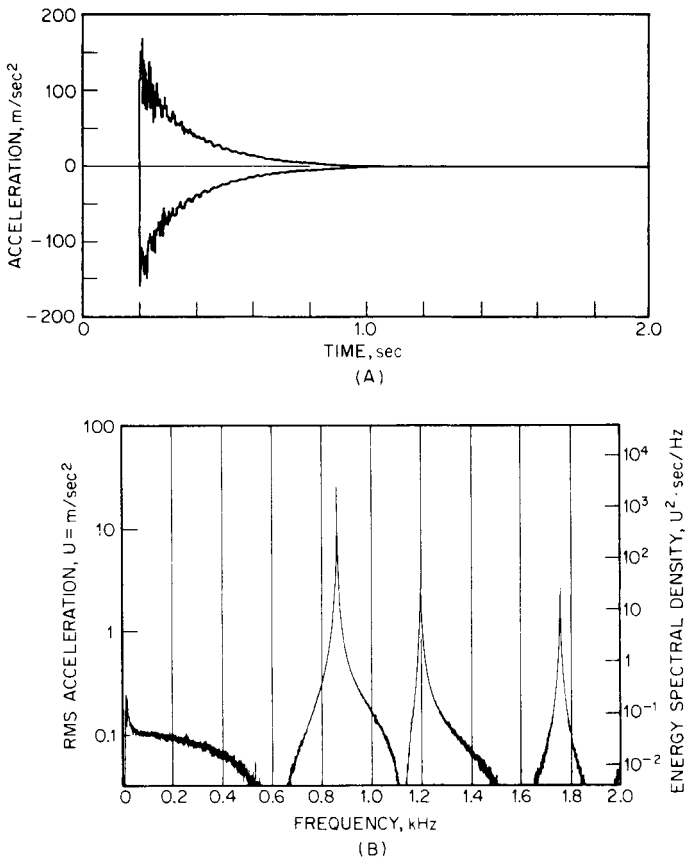

FIGURE 14.22 Analysis of a long transient signal using nondestructive zoom FFT. (*A*) Envelope of time signal of length 10,240 samples (10K) corresponding to 2 seconds ($T = 2$ seconds). (*B*) 4000-line composite zoom spectrum with bandwidth $B = \Delta f = 0.5$ Hz (flat weighting). Scaling on right is converted to energy spectral density (ESD).

a short time window (e.g., a Hanning window) of length T over the entire record; this is done in overlapping steps, and the results are averaged. As already demonstrated for stationary signals (Fig. 14.20), this procedure yields a result with uniform weighting for step lengths $T/3$ and $T/4$. The same applies to step lengths $T/5$, $T/6$, etc., but there is a slight difference with respect to the overall weighting function for the different step lengths. Figure 14.23 illustrates the overall time weighting function for different step lengths T/n (where n is an integer greater than 2) and shows the length of the uniform section (within which the entire transient should ideally be located) and the effective length T_{eff} by which power units should be multiplied to convert them to energy. For a conversion to energy spectral density to be valid, the width of spectrum peaks must be somewhat greater than the analysis bandwidth; this can be seen by inspection of the analysis results. For example, for the Hanning window, the bandwidth B is 1.5 times the line spacing Δf (see Table 14.2), and so spectrum peaks should have a 3-dB bandwidth of more than five lines.

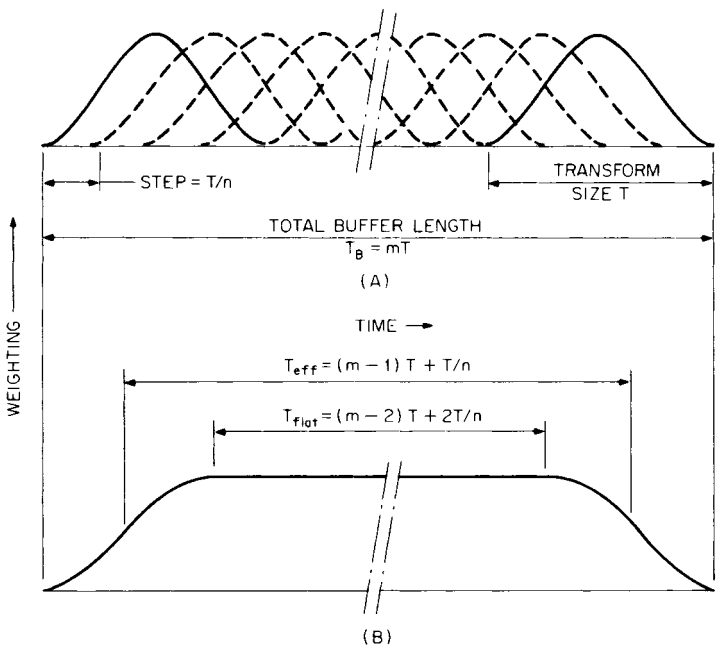

FIGURE 14.23 Overall weighting function for scan averaging of a transient. (*A*) Overlapping Hanning windows of length T with definition of parameters m and n. (*B*) Overall weighting function with indication of T_{eff} and T_{flat} in terms of T, m, and n. T_{eff} is the effective length of the time window for conversion of power to energy units. T_{flat} is the length of the section with uniform weighting within which the transient ideally should be located.

Even though the broader bandwidth obtained by scan averaging may result in a loss of spectrum detail, it provides considerable improvement in the dynamic range of the result. Figure 14.24 (using scan averaging) illustrates these points for the same signal as Fig. 14.22 (using zoom). The spectrum obtained by scan averaging generally has 12 dB more dynamic range than that obtained by zoom (with factor 10), but the

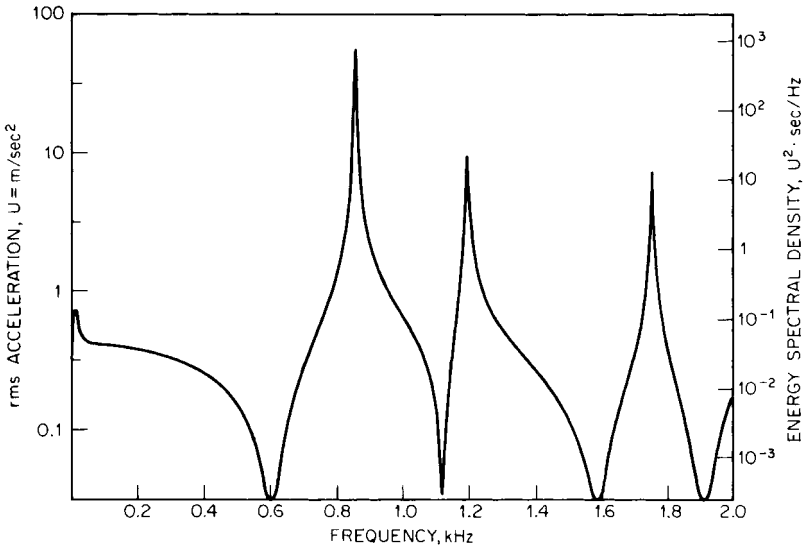

FIGURE 14.24 Analysis of a long transient by scan averaging (same signal as Fig. 14.22). The energy spectral density (ESD) scaling on the right can be compared with that in Fig. 14.22, although the peaks are not valid because of insufficient resolution.

level of peaks does not differ by this amount; this confirms that their resolution is not sufficient to allow scaling in terms of energy spectral density.

To obtain Fig. 14.24, scan averaging with a step length of $T/4$ was used (an overlap of successive records of 75 percent). Even though a step length of $T/3$ (overlap of 66.7 percent) is theoretically more efficient, $T/4$ is usually more convenient because the number of samples in T generally is a power of 2.

ANALYSIS OF NONSTATIONARY SIGNALS

A typical nonstationary signal results from measurements made during a machine run-up or coast-down (here, the primary cause of the nonstationary signal is a change in shaft speed). The signal can be analyzed by dividing it up into a series of short quasi-stationary time periods (often overlapping), in each of which the speed is roughly constant. The length of the time window used to select a portion of the continuous signal may have to be chosen so as to ensure this. The simplest way to analyze a nonstationary signal of this type is to use a tracking filter tuned to a specific harmonic of shaft speed and to record the results vs. rpm of the machine. If a phase meter is inserted between the filtered signal and the tracking signal, it is possible to record phase as well as amplitude against rpm to give what is called a *Bode plot.*[8]

Using an FFT analyzer, the behavior of several harmonics may be studied simultaneously. One way to do this, using an FFT analyzer having a long memory, is with a simple scan analysis; a short Hanning window is scanned through the record (as for a *scan average*), and successive instantaneous spectra (from each window position) are viewed on the display screen. The speed of the scan may be changed by varying the step length; this is one situation (in contrast to scan averaging) where very short

step lengths may be of advantage, for example, in slowing down the passage through a resonance.

A highly effective method of representing such a scan analysis is by a "waterfall," or "cascade," plot as shown in Fig. 14.25 (which represents a typical machine run-up). As indicated, the third dimension of such a three-dimensional plot can be either time or rpm; for a simple scan analysis it usually is time, but if the spectra are spaced at equal intervals of rpm, a number of advantages result. Harmonically related components (whose bases follow radial lines) then can be separated easily from constant-frequency components (e.g., related to line frequency or resonances) whose bases follow lines parallel with the rpm axis. Such a cascade plot, with rpm as the third axis, is sometimes referred to as a *Campbell diagram,* although strictly speaking a Campbell diagram has a vertical frequency axis, a horizontal rpm axis, and a signal amplitude represented as the diameter of a circle (or square) centered on the appropriate point in the diagram.

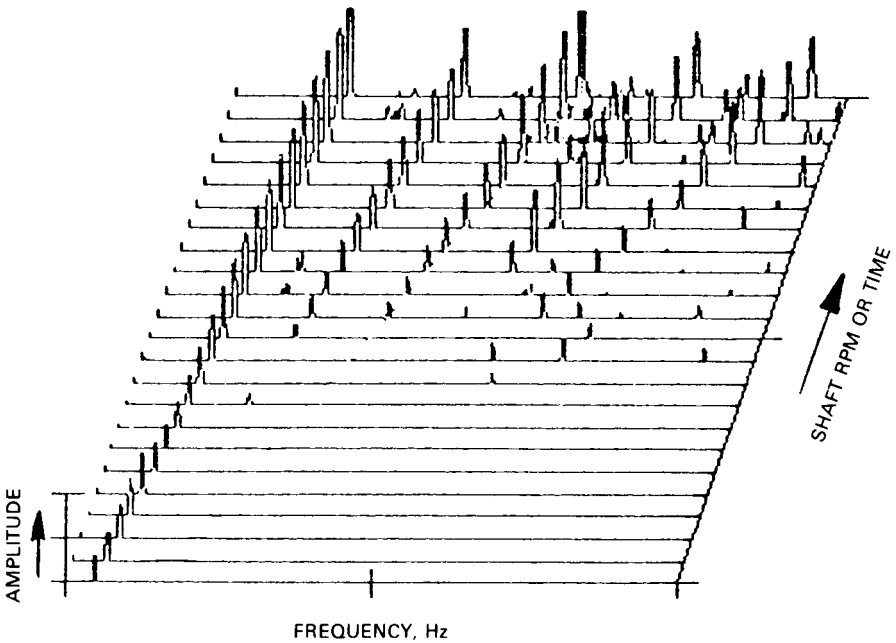

FIGURE 14.25 Three-dimensional spectral map or waterfall plot, showing how spectra change with shaft rpm or time.

Ideally, each of the spectra in a cascade plot such as Fig. 14.25 should be obtained with constant shaft speed at the respective rpm. This is sometimes possible, for example, during the very slow start-up of a large steam turbine, but usually each spectrum is a windowed section of a continuously varying signal with a small speed change within the window length. Consequently, the peak corresponding to each harmonic is not always localized in one analysis line; in particular, the higher harmonics are likely to be spread over progressively more lines. Thus, the height of each peak cannot be used directly as a measure of the strength of each component; it

would be necessary to integrate over the whole of a distributed peak to measure the total power contained in it.

A way of overcoming this problem is to use *tracking analysis,* where the sampling rate of the FFT analyzer is related directly to shaft speed. A frequency multiplier may be used to produce a sampling frequency signal (controlling the A/D converter of the analyzer) which is a specified multiple of the shaft speed.

Figure 14.26 illustrates the basic principles. Figure 14.26*B* shows a hypothetical signal produced by a rotating shaft during a run-up (in practice, the amplitude normally also would vary with shaft speed). Figure 14.26*A* shows the samples obtained by sampling the signal value at a constant sampling frequency (as for normal frequency analysis) and the spectrum resulting from FFT analysis of these samples. The spectral peak is seen to spread over a number of lines corresponding to the speed change along the time record. Figure 14.26*C* shows the samples obtained by sampling the signal a fixed number of times per shaft revolution (in this case, eight). The samples are indistinguishable from those obtained from normal analysis of a constant-frequency component, and thus the frequency spectrum is concentrated in one line.

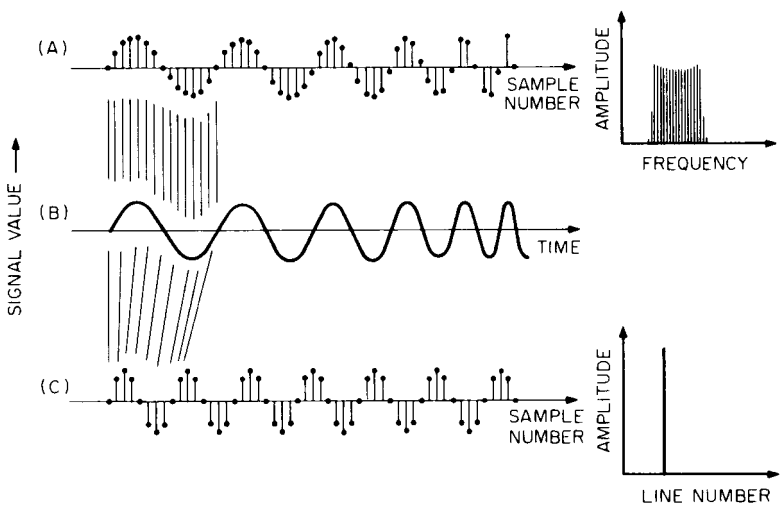

FIGURE 14.26 Analysis of a fundamental component which is increasing in frequency. (*A*) Data record resulting from a uniform sampling rate, and its spectrum, which spreads over a frequency band corresponding to the speed change. (*B*) The original time signal. (*C*) Data record resulting from sampling eight times per fundamental cycle, and its spectrum, which is concentrated in one analysis line.

A frequency multiplier, based on a phase-locked loop, suffers from the disadvantage of a finite response time, so that it cannot keep up if the speed is changing rapidly. A better alternative, offered by some analyzers, is based on digital resampling (interpolation) of each record in line with the simultaneously measured tachometer signal.

When the sampling frequency varies with shaft speed, however, special precautions must be taken to avoid problems with aliasing. One possibility is to use a tracking low-pass filter with a cutoff frequency suitably less than half the sampling

frequency. Because of the difficulty of obtaining a tracking filter having a very steep roll-off (e.g., 120 dB/octave), it is often simpler to choose one of a series of filters with a fixed cutoff frequency, depending on the current shaft speed. Such a series of filters (in, for example, a 2, 5, 10 sequence) often is available in the analyzer to determine the normal frequency ranges. Taking the case of a 400-line analyzer, for example, all 400 lines in the measured spectrum are valid when the sampling frequency is appropriate to the selected filter (Fig. 14.27*A*). If the sampling frequency is higher than the ideal for a given filter, the upper part of the spectrum is affected by the filter (Fig. 14.27*B*). If it is lower, the upper part of the spectrum may be contaminated by aliasing components (Fig. 14.27*C*). Nevertheless, by arranging for the selection of the optimum filter at all times (either manually or automatically), at least 60 percent of the measured spectrum (i.e., in this case 240 lines) is always valid. The analysis parameters can be selected so that the desired number of harmonics is contained within this range, based on the fact that the line number in the spectrum of a given component is equal to the number of periods it represents in the data record of length N samples. If, for example, the 30th harmonic is to be

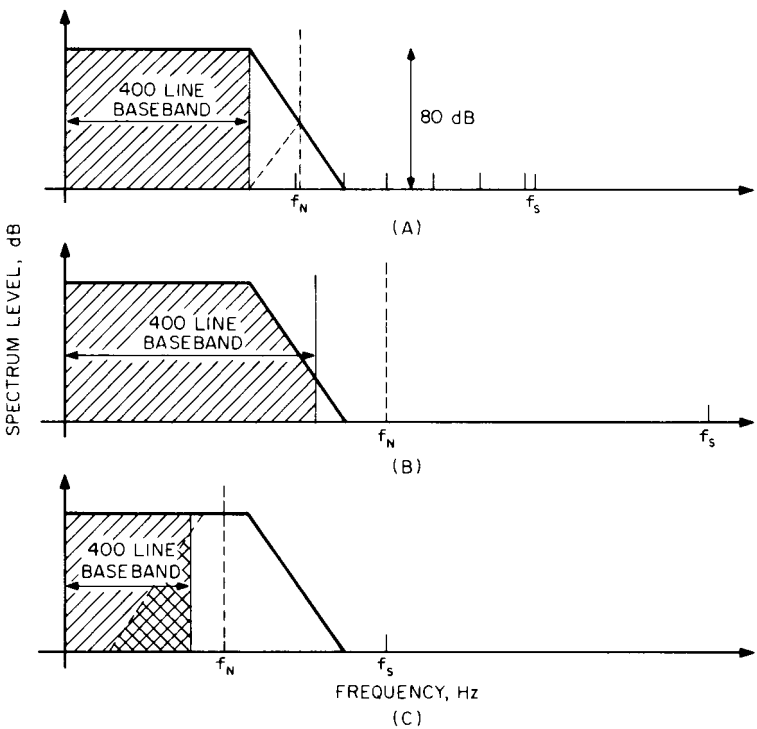

FIGURE 14.27 Effect of sampling frequency on the validity of spectral components, assuming an FFT analyzer with 400 lines and 80-dB dynamic range. f_s = sampling frequency. f_N = Nyquist folding frequency = $f_s/2$. (*A*) Normal situation with optimum choice of sampling frequency for the low-pass filter. (*B*) Situation with increased sampling frequency. The upper lines in the spectrum are influenced by the low-pass filter. (*C*) Situation with decreased sampling frequency. The upper lines in the spectrum are influenced by aliasing components folded around f_N (double cross-hatched area).

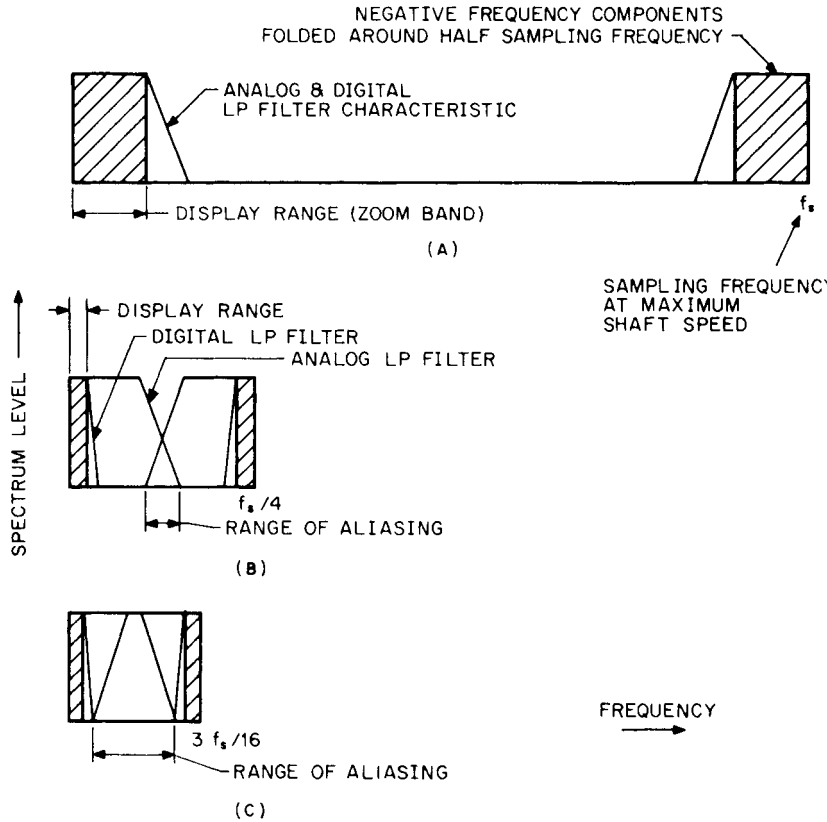

FIGURE 14.28 Use of a fixed low-pass filter to prevent aliasing when tracking with an FFT analyzer employing zoom to analyze in a lower-frequency band. For illustration purposes, the sampling frequency at maximum shaft speed has been made four times greater than that appropriate to the analog LP filter. The shaft speed range could be made proportionally greater by increasing this factor. (*A*) Situation at maximum shaft speed. All harmonics of interest must be contained in the display range. (*B*) Situation at one-fourth maximum shaft speed. The analog filter characteristics overlap, but are well separated from the display range. (*C*) Situation at three-sixteenths maximum shaft speed. The aliasing range almost intrudes on the display range.

located in line no. 240, the fundamental must be in line no. 8; there must be eight periods of the fundamental component along the data record. Where the data record contains 1024 samples (i.e., $N = 1024$), the sampling frequency must then be 128 times the shaft speed; thus a frequency multiplier with a multiplication factor of 128 should be used in this specific case.

For FFT analyzers with zoom, a simpler approach can be used, as illustrated in Fig. 14.28. An analog low-pass filter is applied to the signal with a cutoff frequency corresponding to the highest required harmonic at maximum shaft speed. However, a frequency multiplying factor is chosen so as to make the sampling frequency, say, 10 or 20 times this cutoff frequency (instead of the normal 2.56). The spectrum then

is obtained by zooming in a range corresponding to the highest required harmonic. As shown in Fig. 14.28, the shaft speed (and thus the sampling frequency) can then be varied over a wide range, without aliasing components affecting the measurement results. A somewhat similar procedure is used in conjunction with the digital resampling technique mentioned above. By using four times oversampling, a maximum speed range of 5.92:1 can be accommodated without changing the decimation rate (i.e., the proportion of samples retained after digital filtration), but an even wider range can be covered, at the expense of small "glitches" at the junctions, if the decimation rate is allowed to change.

Figure 14.29 shows the results of tracking FFT analysis on a large turbogenerator. It was made using nondestructive zoom with zoom factor 10. A frequency multiplying factor of 256 was used, giving 40 periods of the fundamental component in the 10K (10,240-point) memory of the FFT analyzer. The fundamental is thus located in line no. 40 of the 400-line zoom spectrum. Because the harmonics coincide exactly with analysis lines, rectangular weighting could have been used in place of the Hanning weighting actually used (all harmonics have exact integer numbers of periods along the record length); Hanning weighting can, however, be advantageous for nonsynchronous components such as constant-frequency components. Such a component at 150 Hz (initially coinciding with the third harmonic of shaft speed) is shown in Fig. 14.29. Constant-frequency components follow a hyperbolic locus in cascade plots employing order tracking.

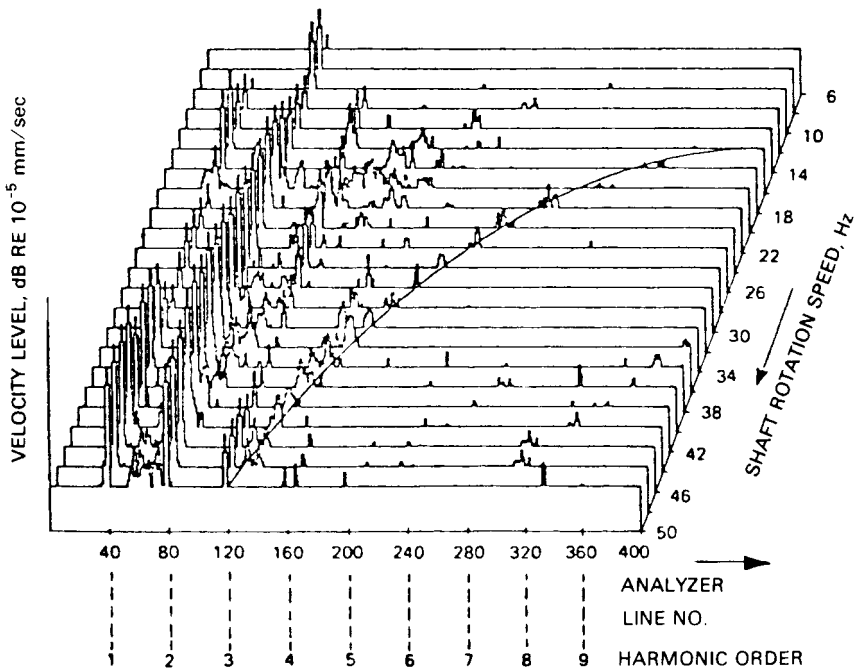

FIGURE 14.29 Tracking FFT analysis of the rundown of a large turbogenerator. The superimposed hyperbolic curve represents a fixed-frequency component at 150 Hz.

RELATED ANALYSIS TECHNIQUES

Signal analysis techniques other than those described above, which are useful as an adjunct to frequency analysis, include synchronous averaging, cepstrum analysis, and Hilbert transform techniques.

Synchronous Averaging (Signal Enhancement). *Synchronous averaging* is an averaging of digitized time records, the start of which is defined by a repetitive trigger signal. One example of such a trigger signal is a once-per-revolution synchronizing pulse from a rotating shaft. This process serves to enhance the repetitive part of the signal (whose period coincides with that of the trigger signal) with respect to nonsynchronous effects. That part of the signal which repeats each time adds directly, in proportion to the number of averages, *n.* The nonsynchronous components, both random noise and periodic signals with a different period, add like noise, with random phase; the amplitude increase is in proportion to $\sqrt{n}$. The overall improvement in the signal-to-noise rms ratio is thus $\sqrt{n}$, resulting in an improvement of $10 \log_{10} n$ dB, i.e., 10 dB for 10 averages, 20 dB for 100, 30 dB for 1000.

Figure 14.30 shows the application of synchronous averaging to vibration signals from similar gearboxes in good and faulty condition. Figure 14.30*A* shows the enhanced time signal (120 averages) for the gear on the output shaft. The signal is fairly uniform and gives evidence of periodicity corresponding to the tooth-meshing. Figure 14.30*B* is a similarly enhanced time signal for a faulty gear; a localized defect on the gear is revealed. By way of comparison, Fig. 14.30*C* shows a single time record, without enhancement, for the same signal as in Fig. 14.30*B;* neither the tooth-meshing effect nor the fault is readily seen.

For best results, synchronous averaging should be combined with tracking. Where there is no synchronization between the digital sampling and the (analog) trigger signal, an uncertainty of up to one sample spacing can occur between successive digitized records. This represents a phase change of 360° at the sampling frequency, and approximately 140° at the highest valid frequency component in the signal, even with perfectly stable speed. Where speed varies, an additional phase shift occurs; for example, a speed fluctuation of 0.1 percent would cause a shift of one sample spacing at the end of a typical 1024-sample record. The use of tracking analysis (generating the sampling frequency from the synchronizing signal) reduces both effects to a minimum.

Cepstrum Analysis. Originally the *cepstrum* was defined as the power spectrum of the logarithmic power spectrum.[9] A number of other terms commonly found in the cepstrum literature (and with an equivalent meaning in the cepstrum domain) are derived in an analogous way, e.g., *ceps*trum from *spec*trum, *quefr*ency from *frequ*ency, *rah*monic from *har*monic. The distinguishing feature of the cepstrum is not just that it is a spectrum of a spectrum, but rather that it is the spectrum of a spectrum on a logarithmic amplitude axis; by comparison, the autocorrelation function [see Eq. (22.21)] is the inverse Fourier transform of the power spectrum without logarithmic conversion.

Most commonly, the *power cepstrum* is defined as the inverse Fourier transform of the logarithmic power spectrum,[10] which differs primarily from the original definition in that the result of the second Fourier transformation is not modified by obtaining the amplitude squared at each quefrency; it is thus reversible back to the logarithmic spectrum. Another type of cepstrum, the *complex cepstrum,* discussed below, is reversible to a time signal.

Figure 14.31, the analysis of a vibration signal from a faulty bearing, shows the advantage of the power cepstrum over the autocorrelation function. In Fig. 14.31*A,* the same power spectrum is depicted on both linear and logarithmic amplitude axes;

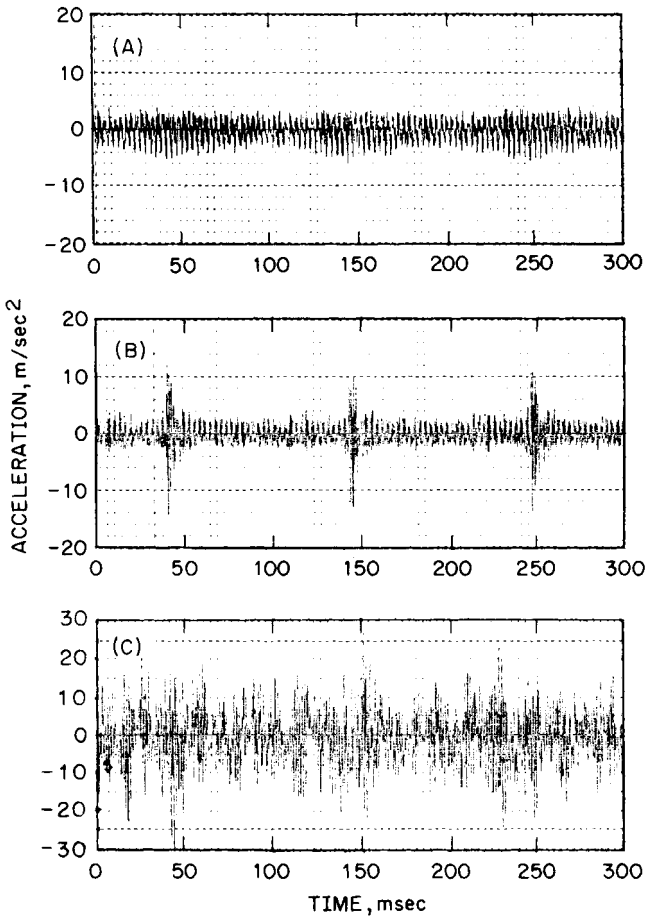

FIGURE 14.30 Use of signal enhancement in gear fault diagnosis. (*A*) Enhanced signal (120 averages) for a gear in normal condition. (*B*) Enhanced signal (120 averages) for a similar gear with a local fault. (*C*) Section of raw signal corresponding to (*B*).

in (*B*) and (*C*) the autocorrelation and cepstrum, respectively, are shown. In (*C*), the use of the logarithmic power spectrum reveals the existence of a family of harmonics which are concealed in the linear depiction. The presence of the family of harmonics is made evident by a corresponding series of rahmonics in the cepstrum (denoted ①, ②, etc.), but is not detected in the autocorrelation function. The quefrency axis of the cepstrum is a time axis, most closely related to the X axis of the autocorrelation function (i.e., time delay or periodic time rather than absolute time). The reciprocal of the quefrency of any component gives the equivalent *frequency spacing* in the spectrum, not the absolute frequency.

Most of the applications of the power cepstrum derive from its ability to detect a periodic structure in the spectrum, for example, families of uniformly spaced har-

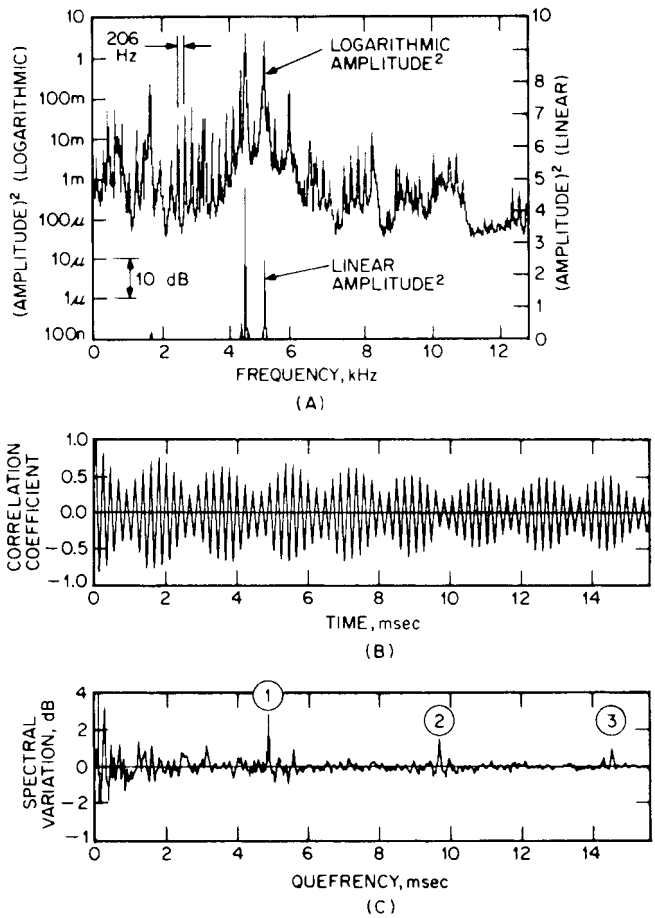

FIGURE 14.31 Effect of linear vs. logarithmic amplitude scale in power spectrum. (*A*) Power spectrum on linear scale (lower curve) and logarithmic scale (upper curve). (*B*) Autocorrelation function (obtained from linear representation). (*C*) Cepstrum (obtained from logarithmic representation)—①, ②, etc., are rahmonics corresponding to harmonic series in spectrum (4.85 milliseconds equivalent to 1/206 Hz). The harmonics result from a fault in a bearing.

monics and/or sidebands. The application of the cepstrum to the diagnosis of faults in gears and rolling element bearings is discussed in Chap. 16 and Ref. 11.

To obtain a distinct peak in the cepstrum, a reasonable number of the members of the corresponding harmonic or sideband family must be present (although the fundamental may be absent). These uniformly spaced components must be adequately resolved in the spectrum. As a guide, the spacing of components to be detected should be a minimum of eight lines in the original spectrum. For this reason, it is often advantageous to perform a cepstrum analysis on a spectrum obtained by *zoom FFT*. In this case it is desirable to use a slightly modified definition of the

cepstrum corresponding to the amplitude of the *analytic signal.*[11] (See the next section on *Hilbert Transform Techniques.*)

The *complex cepstrum*[10,12] (referred to above) is defined as the inverse Fourier transform of the complex logarithm of the complex spectrum. Despite its name, it is a real-valued function of time, differing from the power cepstrum primarily in that it uses phase as well as logarithmic amplitude information at each frequency in the spectrum. It is thus reversible to a time function (from which the complex spectrum is obtained by direct Fourier transformation).

Measured vibration signals generally represent a combination of source and transmission path effects; for example, internal forces in a machine (the source effect) act on a structure whose properties may be described by a frequency response function between the point of application and the measurement point (the transmission path effect). As shown in Refs. 10 and 12, the source and transmission path effects are convolved in the time signals, multiplicative in the spectra, and additive in the logarithmic spectra and in the cepstra (both power cepstra and complex cepstra). In the cepstra, they quite often separate into different regions, which in principle allows a separation of source and transmission path effects in an externally measured signal.[13]

Figure 14.32 shows an example of an internal cylinder pressure signal in a diesel engine, derived from an externally measured vibration acceleration signal making use of cepstrum techniques to generate the inverse filter.[14]

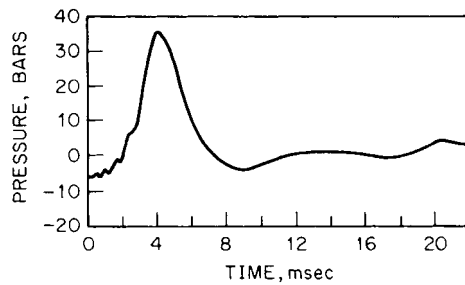

FIGURE 14.32 Diesel engine cylinder pressure signal, derived from an externally measured vibration-acceleration signal using cepstrum techniques. (*From R. H. Lyon and A. Ordubadi.*[14])

Reference 15 gives similar results for the tooth-mesh signal in a gearbox and also shows that a frequency response function derived by windowing in the cepstrum of an output signal compares favorably with a direct measurement (which requires measurement of both an input and an output signal).

Hilbert Transform Techniques. The *Hilbert transform* is the relationship between the real and imaginary parts of the Fourier transform of a one-sided signal.[16] An example is a causal signal such as the impulse response of a vibratory system (a *causal signal* is one whose value is zero for negative time). The real and imaginary parts of the frequency response (the Fourier transform of the impulse response) are related by the Hilbert transform; thus, only one part need be known—the other can be calculated.

Analogously, the time function obtained by an inverse Fourier transformation of a one-sided spectrum (positive frequencies only) is complex, but the imaginary part is the Hilbert transform of the real part. Such a complex time signal is known as an *analytic signal.*

An analytic signal can be thought of as a rotating vector (or phasor) described by the formula $A(t)e^{j\phi(t)}$ whose amplitude $A(t)$ and rotational speed $\omega(t) = d\phi(t)/dt$, in general, vary with time. Analytic signals are useful in vibration studies to describe modulated signals. For example, a *phase-coherent* signal [Eq. (22.3)] can be represented as the real part of an analytic signal, in which case the imaginary part can be obtained by a Hilbert transform. Therefore, from a measured time signal, $a(t)$, it is possible to obtain the amplitude and phase (or frequency) modulation components from the relationship

$$A(t)e^{j\phi(t)} = a(t) + j\tilde{a}(t) \tag{14.12}$$

where $\tilde{a}(t)$ is the Hilbert transform of $a(t)$.

The Hilbert transform may be evaluated directly from the equation

$$\tilde{a}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} a(\tau)\, \frac{1}{t-\tau}\, d\tau \tag{14.13}$$

but it can be more readily evaluated by a phase shift in the frequency domain, in particular in an FFT analyzer.[16] An alternative way of generating analytic signals using an FFT analyzer is by an inverse Fourier transformation of the equivalent one-sided spectrum formed from the spectrum of the real part only. The time signals resulting from the real-time zoom process (described above) automatically have the same amplitude function $A(t)$ as the equivalent bandpass-filtered analytic signal, since they are obtained from the positive frequency components only (Fig. 14.16). The frequency-shifting operation affects only the phase function $e^{j\phi(t)}$.

The major applications of Hilbert transform techniques in vibration studies involve either amplitude demodulation or phase demodulation.

Amplitude Demodulation. Figure 14.33 shows the analytic signal for the case of single-frequency amplitude modulation of a higher-frequency carrier component. The imaginary part is the Hilbert transform of the real part; this manifests itself as a 90° phase lag. The amplitude function is the envelope of both the real and imaginary parts and represents the modulating signal plus a dc offset. The phase function is a linear function of time (whose slope represents the speed of rotation, or frequency, of the carrier component); it is, however, shown modulo 2π, as is conventional.

One area of application of amplitude demodulation where it is advantageous to view the signal envelope rather than the time signal itself is in the interpretation of such oscillating time functions as autocorrelation and crosscorrelation functions (see Chap. 22). Figure 14.34[18] shows a typical case where peaks indicating time delays are difficult to identify in a crosscorrelation function as defined in Eq. (22.48), because of the oscillating nature of the basic function (Fig. 14.34*A*). The peaks are much more easily seen in the envelope or magnitude of the analytic signal (Fig. 14.34*B*). Another advantage of the analytic signal is that its magnitude can be displayed on a logarithmic axis; this allows low-level peaks to be detected and converts exponential decays to straight lines.[18]

Another area of application of amplitude demodulation is in *envelope analysis* (discussed in Chap. 13 in the section on *Envelope Detectors*). In particular, when the signal is to be bandpass-filtered before forming the envelope, this can be done by real-time zoom in the appropriate passband. Figure 14.35 shows an example from the same vibration source as was analyzed in Fig. 14.31. Figure 14.35*A* shows a typical envelope signal obtained from zooming in a 1600-Hz band centered at 3 kHz.

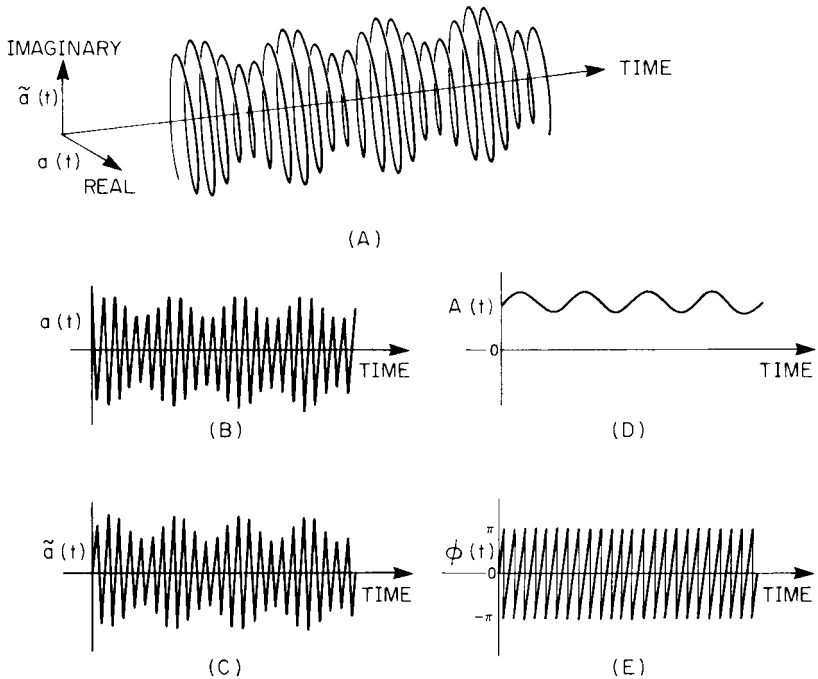

FIGURE 14.33 Analytic signal for simple amplitude modulation. (*A*) Analytic signal $a(t) + j\tilde{a}(t) = A(t)e^{j\phi(t)}$. (*B*) Real part $a(t)$. (*C*) Imaginary part $\tilde{a}(t)$. (*D*) Amplitude $A(t)$. (*E*) Phase $\phi(t)$.

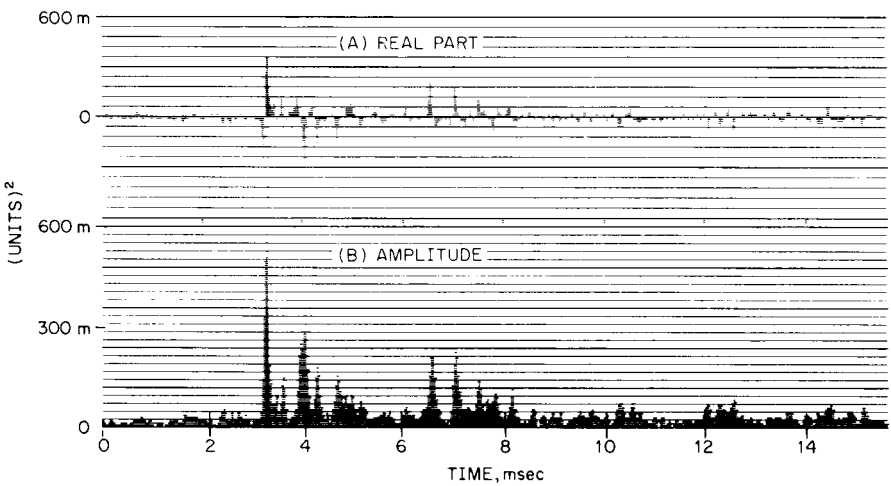

FIGURE 14.34 Example of a crosscorrelation function expressed as follows:[18] (*A*) The real part of an analytic signal, i.e., the normal definition [Eq. (22.48)]. (*B*) The amplitude of the analytic signal. The peaks corresponding to time delays are more easily seen in this representation. The signal was obtained by bandpass filtering (using FFT zoom) in the frequency range from 512 to 13,312 Hz.

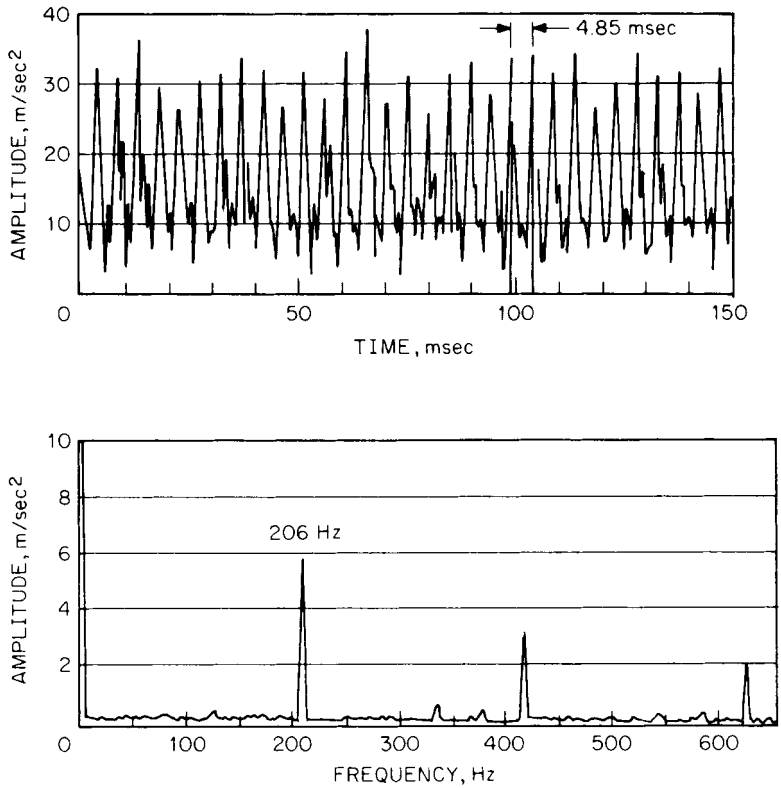

FIGURE 14.35 Envelope analysis using Hilbert transform techniques. (*A*) Typical envelope signal showing bursts with a period of 4.85 milliseconds from a fault in a ball bearing. (*B*) Average spectrum of the envelope signal showing corresponding harmonics of 206 Hz. Signal obtained by bandpass filtering (using FFT zoom) in the frequency range from 2200 to 3800 Hz (compare with Fig. 14.31*A*, which shows a baseband analysis of this same signal).

The spectrum of Fig. 14.31*A* shows that this frequency range is dominated by the harmonic family which results from a fault in a bearing. Consequently, the corresponding envelope signal (Fig. 14.35*A*) indicates a series of bursts with the same period, 4.85 milliseconds (compare with the cepstrum of Fig. 14.31*C*). Figure 14.35*B* shows the average spectrum of a number of such envelope signals; this gives a further indication that the dominant periodicity is 206 Hz.

Phase Demodulation. For a purely phase-modulated signal, the amplitude function $A(t)$ is constant and the phase function $\phi(t)$ is given by the sum of a carrier component of constant frequency f_c and the modulation signal $\phi_m(t)$. Thus

$$\phi(t) = 2\pi f_c t + \phi_m(t) \tag{14.14}$$

Real-time zoom analysis centered on frequency f_0 subtracts this frequency from all components in the signal; consequently, by zooming at the carrier frequency f_c, only the modulation signal $\phi_m(t)$ remains. In general it is possible to zoom exactly at the carrier frequency only when the latter is made to coincide exactly with an analysis line (for example, by employing order tracking). Otherwise, the small difference in frequency gives a residual slope to the phase signal.

Figure 14.36 shows an example of the application of this technique to the measurement of gear transmission error.[19] This can be obtained as the difference in torsional vibration (i.e., phase modulation) of the two gears in mesh, after appropriate compensation for the gear ratio (in this particular case the ratio is unity). The torsional vibrations were measured by demodulating the output signals from optical encoders attached to each shaft. The encoders give 16,000 pulses per revolution, but this was divided down to 4000 for the results shown here (and for the *zoom demodulation* technique even further decimation would be possible). The result obtained by zoom demodulation, including digital tracking, was produced by an advanced FFT analyzer, and is compared with a result obtained using a 100-MHz clock to time the intervals between pulses and thus measure phase modulation somewhat more directly. The two results are virtually identical, and are accurate to within a few arcseconds. Similar methods have been used to detect cracks in gears by amplitude and phase demodulation of the tooth-meshing signal.[20]

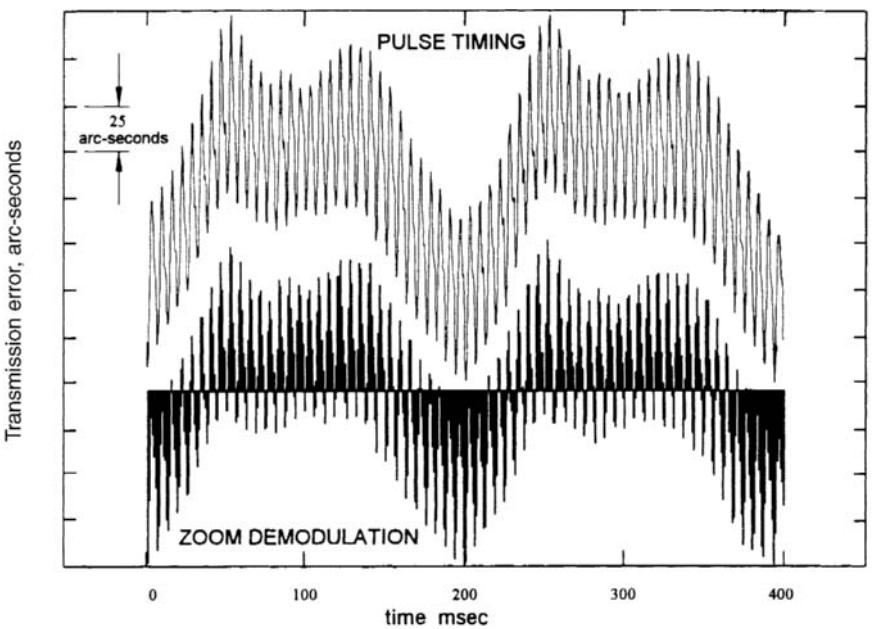

FIGURE 14.36 Gear dynamic transmission error measured using the zoom demodulation technique compared with direct measurement by timing the intervals between shaft encoder pulses.[19] Measurements were made with two 32-tooth gears, although the method is not limited to unity-ratio gears. Note the periodic repetition once per revolution of the gears (200 milliseconds) and the higher-frequency component corresponding to tooth-meshing.

REFERENCES

1. Cooley, J. W., and J. W. Tukey: *Math. Computing,* **19**(90):297 (1965).
2. Cooley, J. W., P. A. W. Lewis, and P. D. Welch: *J. Sound Vibration,* **12**(3):315 (1970).
3. Brigham, E. O.: "The Fast Fourier Transform," Prentice-Hall, Inc., Englewood Cliffs, N.J., 1974.
4. Thrane, N.: "Zoom-FFT," *Brüel & Kjaer Tech. Rev.,* (2) (1980).
5. Sloane, E. A.: *IEEE Trans. Audio Electroacoust.,* **AU-17**(2):133 (1969).
6. Welch, P. D.: *IEEE Trans. Audio Electroacoust.,* **AU-15**(2):70 (1967).
7. Randall, R. B.: "Frequency Analysis," Brüel & Kjaer, Naerum, Denmark, 1987.
8. Mitchell, J. S.: "An Introduction to Machinery Analysis and Monitoring," Penwell Publishing Company, Tulsa, Okla., 1981.
9. Bogert, B. P., M. J. R. Healy, and J. W. Tukey: In M. Rosenblatt (ed.), "Proceedings of the Symposium on Time Series Analysis," John Wiley & Sons, Inc., New York, 1963, pp. 209–243.
10. Childers, D. G., D. P. Skinner, and R. C. Kemerait: *Proc. IEEE,* **65**(10):1428 (1977).
11. Randall, R. B.: *Maintenance Management Int.,* **3:**183 (1982/1983).
12. Oppenheim, A. V., R. W. Schafer, and T. G. Stockham Jr.: *Proc. IEEE,* **56**(August):1264 (1968).
13. Gao, Y., and R. B. Randall: *Mechanical Systems and Signal Processing,* **10**(3):293–317, 319–340 (1996).
14. Lyon, R. H., and A. Ordubadi: *J. Mech. Des.,* **104**(Trans. ASME)(April):303 (1982).
15. DeJong, R. G., and J. E. Manning: "Gear Noise Analysis using Modern Signal Processing and Numerical Modeling Techniques," *SAE Paper* No. 840478, 1984.
16. Papoulis, A.: "The Fourier Integral and Its Applications," McGraw-Hill Book Company, Inc., New York, 1962.
17. Thrane, N.: *Brüel & Kjaer Tech. Rev.,* (3) (1984).
18. Herlufsen, H.: *Brüel & Kjaer Tech. Rev.,* (1 and 2) (1984).
19. Sweeney, P. J., and R. B. Randall: *Proc. I. Mech. E., Part C, J. Mech. Eng. Sc.,* **210**(C3):201–213 (1996).
20. McFadden, P.: *J. Vib. Acoust. Stress & Rel. Des.,* **108**(Trans. ASME)(April):165 (1986).