

FREQUENCY ANALYSIS

by

R.B. Randall, B. Tech., B.A.

Revision September 1987

3rd edition 1st print

ISBN 87 87355 07 8

FOREWORD

In the years since the First and Second Editions of this book were published a great many developments have occurred; in particular there has been a huge swing from analog to digital analysis techniques, even for day-to-day analysis. The main reason for this development has been the increase in power and reduction in cost of FFT (Fast Fourier Transform) analyzers, but digital filters have also increasingly taken over some of the application areas formerly covered by analog filters.

At the same time it has been realised that the primary distinction is not between digital and analog analysis, but rather between the FFT process, where the signal is analyzed in blocks, and analysis using filters, where the signal is processed continuously. Digital filters are a special case of the latter, but otherwise have very similar properties to analog filters. For this reason the layout of the book has been modified considerably, with digital filters generally being treated along with analog filters, and FFT techniques treated separately.

The basic theory has been expanded to include a discussion of the Hilbert transform, and in particular how a model of frequency analysis, based on an analytic signal, relates to the model promulgated earlier, which is based on a sum of positive and negative frequency phasors. In particular in the case of modulated signals (involving amplitude, phase and frequency modulation) the approach using an analytic signal is useful, because its amplitude represents the amplitude modulation signal (plus DC offset) while its phase represents the phase (or frequency) modulation signal superimposed on the carrier component.

The analysis of stationary signals, using filters and FFT techniques, respectively, is treated in Chapters 3 and 4. Despite rearrangement, much of the material is similar to that in the earlier editions, although the discussion of FFT techniques has been expanded somewhat. At the same time, obsolete material on "Time Compression" and "High Speed" analysis has been reduced considerably, or dropped.

Similarly, Chapter 5, on the analysis of short transients, contains much material from the earlier editions, in particular with respect to analog analysis, but the section on FFT techniques has been expanded considerably, in line with the overwhelming preponderance of these methods today.

The subject of Chapter 6, the analysis of more slowly varying non-stationary signals, was treated very briefly in the earlier editions, but the treatment here has been expanded greatly. Only FFT analysis is covered, as this is by far the most practical way of applying the "time windows" involved. This chapter also contains new material on "order tracking" where the sampling frequency of the FFT analyzer is tied directly to machine shaft speed, and the x-axis becomes one of harmonic order, rather than frequency.

Chapters 7 and 8 also represent considerable expansions of material treated briefly in the earlier editions. Chapter 7 covers dual channel FFT analysis, while Chapter 8 covers Cepstrum analysis. In the intervening years there have been considerable developments in both fields, and for this reason there is somewhat more emphasis on examples of application in these two chapters compared with earlier chapters.

Overall, it is hoped that the new arrangement will provide more direct access to the somewhat larger body of material.

R.B. Randall

CONTENTS

1. INTRODUCTION	11
2. THEORETICAL FREQUENCY ANALYSIS	16
2.1. BASIC CONCEPTS	16
Complex Notation	16
Delta Functions	19
2.2. FOURIER ANALYSIS	19
Fourier Series	20
Fourier Transform	25
Sampled Time Functions	26
Discrete Fourier Transform	28
Spectral Density	32
Power vs. Energy	32
2.3. BANDWIDTH	33
2.4. SIGNAL TYPES	39
Stationary Deterministic Signals	41
Stationary Random Signals	42
Pseudo-random Signals	44
Transient Signals	45
Non-Stationary Signals	48
2.5. CONVOLUTION	49
Convolution with a Delta Function	51
Convolution of Spectral Functions	52
Convolution Theorem	53
2.6. THE HILBERT TRANSFORM	58
Analytic Signals	62
Modulation	70

3. FILTER ANALYSIS OF STATIONARY SIGNALS	78
3.1. FILTERS	81
Practical Filter Characteristic	83
Filter Response Time	85
3.2. DETECTORS	87
Squaring	88
Averaging	90
Square Root Extraction	97
Logarithmic Conversion	99
Detector Response	99
3.3. RECORDERS	101
AC Recording	101
DC Recording	103
3.4. ANALOG ANALYZER TYPES	104
Discrete Stepped Filter Analyzers	104
Sweeping Filter Analyzers	106
Real-time Parallel Analyzers	107
Time Compression Analyzers	109
3.5. DIGITAL FILTERING	109
The Z-transform	112
The General Multiple-Pole Filter	114
The Brüel & Kjær Digital Frequency Analyzer Type 2131	115
1/12-octave Analysis	120
Filter Characteristics	122
3.6. PRACTICAL ANALYSIS OF STATIONARY SIGNALS	125
Stepped Filter Analysis	125
Averaging Time T_A	126
DC or AC Recording	127
Writing Speed	127
Paper Speed	127
Swept Filter Analysis	130
Constant vs. Constant Percentage Bandwidth	130
Linear vs. Logarithmic Frequency Scale	131
Choice of Bandwidth	132
Averaging Time T_A	133
Filter Dwell Time T_D	134
Writing Speed W	136
Paper Speed P	137
Summary	138
Calibration	139
Digital filter analysis	143
Linear vs. Exponential Averaging	143
Averaging Time T_A	144

4. FAST FOURIER TRANSFORM (FFT)	146
4.1. THE FFT ALGORITHM	147
4.2. THE FFT FOR REAL-VALUED SERIES	152
4.3. LIMITATIONS OF FFT PROCESSING	153
Aliasing	156
Window Effects	157
Picket Fence Effect	161
4.4. ZOOM FFT	165
Real-time Zoom	166
Non-destructive Zoom	169
Comparison of Zoom Techniques	171
4.5. PRACTICAL ANALYSIS OF STATIONARY SIGNALS	172
Analysis Parameters	172
Calibration and Scaling	173
Averaging	174
Real-Time Analysis	176
4.6 HILBERT TRANSFORM TECHNIQUES	177
Amplitude Demodulation	178
Phase Demodulation	182
5. TRANSIENT ANALYSIS	185
5.1. FFT TECHNIQUES	185
Short transients	185
Longer transients	187
5.2. FILTER ANALYSIS TECHNIQUES	192
Digital Filter Analysis of Transients	193
Sequential Filter Analysis	197
Method A	197
Method B	199
Method C	202
Examples	203
6. NON-STATIONARY SIGNALS	211
6.1. SCAN ANALYSIS	211
Choice of analysis parameters	212
Cyclic signals	213
Spectral Variations with Machine Speed	215
6.2. ORDER TRACKING ANALYSIS	219
Optimum Choice of Internal Filter	222
Zoom tracking	222

7. DUAL CHANNEL ANALYSIS	227
7.1. CROSS SPECTRUM	228
Definitions and Calculation Procedures	228
Applications	230
7.2. COHERENCE	230
Definition and Properties	230
Applications	234
7.3. FREQUENCY RESPONSE FUNCTIONS	239
Definitions and Calculation Procedures	239
Effects of Noise on H_1 and H_2	241
Effects of Leakage on H_1 and H_2	245
Applications of Frequency Response Functions	250
7.4. CORRELATION FUNCTIONS	252
Definitions and Calculation Procedures	252
Properties and Applications of Correlation Functions	256
Autocorrelation Function	256
Cross Correlation Function	259
7.5. IMPULSE RESPONSE FUNCTIONS	263
Definitions and Calculation Procedures	263
Properties and Applications	265
8. CEPSTRUM ANALYSIS	271
8.1. DEFINITIONS AND CALCULATION PROCEDURES	272
8.2. PROPERTIES OF THE CEPSTRUM	275
8.3. APPLICATIONS OF THE POWER CEPSTRUM	282
Echo Detection and Removal	282
Properties of a reflecting surface	283
Speech Analysis	284
Machine diagnostics	288
8.4. PRACTICAL CONSIDERATIONS WITH THE POWER CEPSTRUM	291
8.5. THE COMPLEX CEPSTRUM	297
8.6. APPLICATIONS OF THE COMPLEX CEPSTRUM	299
Echo Removal	299
Deconvolution	303
APPENDIX A — Mathematical Details	305
APPENDIX B — Mean Square Error for Narrow Band Random Noise	312

APPENDIX C — Response of Physical Systems	318
REFERENCES	330

1. INTRODUCTION

The object of frequency analysis is to break down a complex signal into its components at various frequencies, and it can be shown that this is possible for all practical signals (Ref. 1.1). The word “components” can be interpreted in several different ways, however, and it is important to clarify the differences between them.

Mathematicians and theoretically inclined engineers tend to interpret “components” as the results of a Fourier Analysis, while practical engineers often think in terms of measurements made with filters tuned to different frequencies. It is one of the purposes of this book to clarify the relationships between these two approaches. Another is to give a simple pictorial interpretation of Fourier analysis which greatly facilitates the understanding of fundamental relationships, in particular the connection between mathematical theory and practical analysis by both analog and digital means.

The approach taken is to consider frequency components as vectors (or phasors) rotating in a complex plane rather than as sums of sines and/or cosines. A typical (co-) sinusoidal component is represented initially as the sum of two contrarotating vectors (representing positive and negative frequency, respectively). This gives a meaning to the mathematically useful concept of negative frequency, and a two-sided frequency domain gives a valuable symmetry with the time domain. It also allows for the simple pictorial representation of Fourier analysis which is developed and used extensively in this book.

Later, in connection with a discussion of the Hilbert transform, an alternative representation is developed, based on the projection on the real axis of a single rotating vector. This gives a one-sided frequency spectrum, but results in a complex time signal (analytic signal), whose real and imaginary components are related by a Hilbert transform. This representation is very useful in the analysis of modulated signals (amplitude, phase, and frequency modulation).

In both approaches a considerable simplification of the mathematics arises from the fact that a vector (having two components e.g. amplitude and phase) can be represented as a single complex variable and because differentiation

and integration of such complex variables are so simple. It is assumed that the reader is familiar with the basic relationships between complex exponentials and sinusoids, but a brief résumé of the most important relationships is given at the start of Chapter 2.

Chapter 2 continues with a discussion of Fourier analysis, first the various forms taken by the Fourier transform, then the important practical consideration of bandwidth. This leads into a discussion of the different types of signal encountered in practice, and how they appear in spectral and other representations. Next, the very important subject of convolution is treated in some detail. It is shown how the output of a linear physical system is obtained by convolving the input signal with the impulse response of the system and how this rather complicated operation transforms by the Fourier transform to a multiplication (which in turn becomes an addition in the normal logarithmic representation of spectra). Other applications of this so-called "Convolution Theorem" are also given. The chapter closes with an introduction to the Hilbert transform, how it can be calculated, and its application to the analysis of modulated signals.

Chapter 3 is entitled "Filter Analysis of Stationary Signals" and covers both traditional analysis with analog filters and detectors, and analysis with digital filters (and detectors) which have many similarities with their analog counterparts. The chapter starts with a discussion of the function and properties of the basic elements of an analyzer based on filters, viz., the filter, the detector (for measuring the power in the filter output signal) and the graphic recorder often used to record the spectrum on paper. Any one of these elements can determine the speed with which the analysis can be made. The chapter goes on to describe the various types of analyzers available, encompassing those based on both analog and digital filters. Finally, the practical analysis of stationary signals is discussed in detail, and formulae and procedures developed for the optimum choice of analyzer and analysis parameters such as filter bandwidth, averaging time, and recorder paper speed and writing speed. This is done for stepped filter analysis (e.g. 1/3-octave and 1/1-octave filters), then sweeping narrow-band filters, and finally for real-time analysis using digital filters. Calibration of the results is also clarified. These procedures allow determination of results with a known degree of accuracy in the shortest possible time.

Chapter 4 introduces another digital analysis method, the so-called Fast Fourier Transform (FFT), which since its introduction in 1965 has revolutionized signal analysis. It is first shown how the FFT procedure greatly speeds up the calculation of the DFT (Discrete Fourier Transform) introduced in Chapter 2, and how the results obtained using the DFT differ from those obtained by direct Fourier transform. Precautions have to be taken to avoid problems with the three "pitfalls" of the DFT, viz., "aliasing", "leakage", and the "picket fence effect". The principles of "Zoom" analysis are explained and compared for the two major techniques used to obtain increased resolution in a limited frequency band. Finally, procedures are given for the practical analysis of stationary

signals, including normal spectrum analysis, and analysis of demodulated signals obtained using Hilbert transform techniques. In Chapter 4 the discussion is limited to simple spectrum analysis of a single channel signal, for direct comparison with the methods of the previous chapter. It is seen that the FFT is best adapted to constant bandwidth analysis on a linear frequency scale.

Chapter 5 discusses the analysis of transient signals, firstly using FFT techniques, and then using filters. Analysis using analog filters has now largely been superseded by digital techniques (FFT and digital filters) but is included for the sake of completeness. Even though many analyzers give results in RMS or mean square (i.e. "power") values, the conversion of the results to "energy spectral density" (ESD), or "band energy level" is described for all analysis techniques.

Chapter 6 covers the analysis of non-stationary signals such as speech, or machine vibration signals obtained during run-ups and run-downs in speed. The approach used employs time windows which are moved along the signal in overlapping steps so that the signal in each window position is either quasi-stationary, or an isolated event. The results are usually displayed in a 3-dimensional spectral map, with time or shaft RPM as the third axis. For cyclic signals, such as those from reciprocating machines, the third axis may be crank angle (over 360° or 720°) and the possibility of averaging over several cycles is explained. Order analysis (where the frequency axis is changed to one of harmonic order) is also described. Because the analyzer sampling frequency is then variable, being linked to shaft speed, problems may be encountered with aliasing, and a range of measures to avoid this are described.

Chapter 7 is a brief introduction to multiple channel FFT analysis, where the signals are analyzed two at a time. These techniques are usually employed to determine the response properties of physical systems (e.g. structural, acoustical, electrical systems) rather than from an interest in the signals themselves. Dual channel analysis involves the calculation of a number of functions not previously defined, viz., cross spectrum, coherence, and frequency response functions, in the frequency domain, and autocorrelation, cross correlation, and impulse response functions in the time domain. The definitions, properties, calculation procedures and major applications of these functions are described and illustrated with examples.

Chapter 8 is a discussion of Cepstrum analysis, which has been included because it can readily be carried out in modern FFT analyzers, and because it throws light on the material of earlier chapters, both with respect to signal analysis, and system analysis. The cepstrum is the inverse Fourier transform of a logarithmic spectrum and often extracts valuable information from the latter in the same way that normal Fourier analysis extracts information from a time signal. There are two types of cepstrum, the "power cepstrum" and the "complex cepstrum". The former is obtained from the power spectrum, while the

latter is derived from a complex spectrum and involves both log amplitude and phase information at each frequency. For this reason the process of calculating the complex cepstrum is reversible back to a time signal. Some applications of the cepstrum are based on its simple ability to detect periodic structures in the spectrum (e.g. harmonics, sidebands and echoes); others are based on the fact that source and transmission path effects, convolved together in the time domain, are additive and often separable in the cepstrum, thus allowing deconvolution by selective editing.

For the somewhat more novel topics of the later chapters, some examples of applications have been given, but for simple frequency analysis not much has been said about applications; it being assumed that a reader already had a need for frequency analysis, and only required information on how to obtain the results. Even so, it is perhaps desirable to give a brief survey of the major reasons for the widespread use of frequency analysis.

Without a doubt the major reason is that many physical and biological systems only respond to a limited range of frequencies. For example, the human ear only perceives sound within the range 20 Hz — 20 kHz, while for estimating the effects of "whole body vibration" the range is rather more restricted (typically 1 — 80 Hz). Quite often physical systems, be they electrical circuits or civil engineering structures, only have strong resonances or excitations within a limited frequency range, and hence the study of their behaviour when subjected to an external excitation can be restricted to this frequency range, thus greatly simplifying the problem. As mentioned previously, the calculation of the response of a linear system reduces in the frequency domain to a simple multiplication. An input at a given frequency gives an output only at that same frequency, and thus frequency analysis permits treating each frequency individually without regard to what is happening at other frequencies.

Sometimes frequency analysis is used to make a picture clearer. Quite often the "noisiness" of a signal is contained mainly in its phase spectrum, the power spectrum being relatively stable, and quite often it is the power spectrum which gives most information with respect to "average" effects.

The frequency content of a signal can often be used diagnostically; for example in tracing the source of a noise problem to a particular machine, or of a mechanical problem to a particular rotating component.

On the other hand it must be kept in mind that frequency analysis is not a universal panacea. In enhancing periodic events it tends to smear out individual ones; for example, one obtains information about toothmeshing frequencies in a gearbox but tends to lose information about individual gear teeth.

Even so, there are many problems where frequency analysis reigns supreme, and it is hoped that the information in this book will help to avoid some of the pitfalls which can be encountered, so that the data obtained by frequency analysis are correct (or at least have a known degree of accuracy) before they are applied to the solution of a practical problem.

2. THEORETICAL FREQUENCY ANALYSIS

2.1. BASIC CONCEPTS

2.1.1. Complex Notation

As mentioned in the Introduction, the approach to Fourier analysis used here is based on the representation of frequency components as rotating vectors (strictly speaking, phasors), these in turn being represented mathematically as complex exponentials. A short résumé is given here of some of the most important relationships which will be used later in the text.

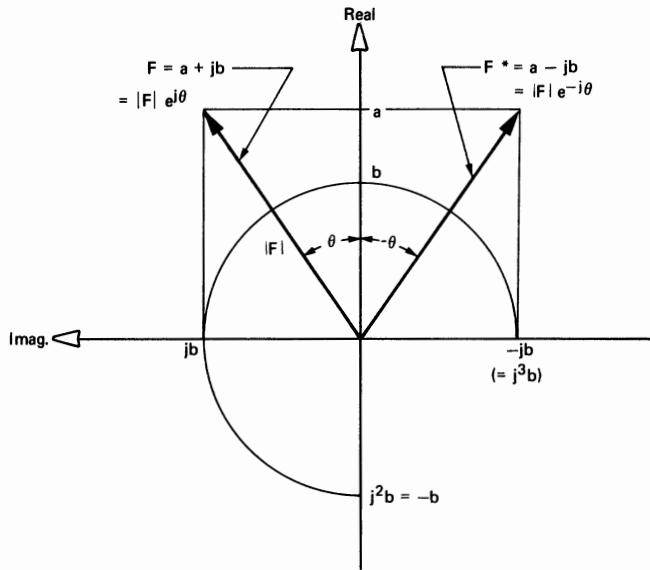
Fig. 2.1 represents a two-dimensional vector F in the so-called “complex plane”. This has a “real” component a directed along the “Real Axis” and an “imaginary” component jb directed along the “Imaginary Axis”. The vector as a whole is represented as the sum of these, viz.:

$$F = a + jb \quad (2.1)$$

Note that b alone, as a real number, would lie along the Real Axis, but that multiplication by j has the effect of rotating it through $\pi/2$ radians. Accordingly, a further multiplication by j would result in a further rotation through $\pi/2$ so that the vector of length b would then lie along the negative Real Axis as shown in Fig. 2.1. Hence, multiplication by j^2 corresponds to a multiplication by -1 and j can thus be interpreted as $\sqrt{-1}$.

Note that the complex plane shown here is turned through 90° as compared with the conventional representation with the Real Axis horizontal. This is done purely to simplify interpretation of the Real Axis as being in the plane of the paper rather than at right angles to it in 3-dimensional diagrams (e.g. Fig. 2.4).

In many cases it is desirable to represent F in terms of its amplitude $|F|$ and phase angle θ instead of its real and imaginary components, and from Fig. 2.1 it can be seen that the relationships between these two sets of coordinates are:



761196

Fig. 2.1. Complex notation for a 2-dimensional vector

$$a = |F| \cos \theta$$

$$b = |F| \sin \theta$$

$$|F| = \sqrt{a^2 + b^2} \quad (2.2)$$

$$\theta = \tan^{-1}\left(\frac{b}{a}\right)$$

From Equations (2.1) and (2.2) it follows that:

$$F = |F| (\cos \theta + j \sin \theta) \quad (2.3)$$

and since it is well-known (Euler's relation) that:

$$\cos \theta + j \sin \theta = e^{j\theta} \quad (2.4)$$

the most concise way of representing F in terms of its amplitude and phase is as the complex exponential

$$F = |F| e^{j\theta} \quad (2.5)$$

As a point of terminology, Fig.2.1 also illustrates what is meant by F^* , the “complex conjugate” of F . This is seen to be the mirror image of F around the real axis. Thus, the real parts and amplitudes have the same sign while the imaginary parts and phase angles have opposite sign. The absolute values of the equivalent components are the same.

Normally in this book we will be considering uniformly rotating vectors, i.e. vectors whose amplitude $|F|$ is a constant and whose phase angle θ is a linearly varying function of time

$$\text{i.e. } \theta = \omega t + \phi$$

where ω is a constant angular frequency (in radians/s) and ϕ is the “initial” phase angle at time zero.

Normally, the frequency will be expressed as circular frequency f in revolutions/s (Hertz) rather than ω in radians/s and thus

$$\theta = 2\pi ft + \phi \quad (2.6)$$

It follows from the above that $e^{j\phi}$ is a unit vector (amplitude = 1) with angular orientation ϕ , and $e^{j\theta}$ (where θ is as defined in (2.6)) is a unit vector rotating at frequency f Hz and with angular orientation ϕ at time zero.

The more general case where $|F|$ and θ are variable functions of time is discussed in Section 2.6.

Vector (i.e. phasor) multiplication is simplest when the vectors are expressed in the form of Equation (2.5) and for two vectors $F_1 (= |F_1| e^{j\theta_1})$ and $F_2 (= |F_2| e^{j\theta_2})$ is obtained simply as:

$$\begin{aligned} F_1 \cdot F_2 &= |F_1| e^{j\theta_1} \cdot |F_2| e^{j\theta_2} \\ &= |F_1| \cdot |F_2| e^{j(\theta_1 + \theta_2)} \end{aligned} \quad (2.7)$$

i.e. the amplitude of the product is equal to the product of the two amplitudes while the phase is equal to the sum of the phases.

In particular, multiplication by a fixed unit vector $e^{j\phi}$ has no effect on the amplitude but adds ϕ to the phase angle (i.e. rotates the vector through an angle ϕ) while multiplication by the rotating unit vector $e^{j2\pi ft}$ causes a vector to rotate at frequency f .

2.1.2. Delta Functions [‡]

Another mathematical concept of which considerable use will be made is the Dirac delta function, also known as the “unit impulse”. A typical delta function, located on an *X*-axis at $x = x_0$ may be represented as $\delta(x - x_0)$. It has the property that its value is zero everywhere except at $x = x_0$, where its value is infinite. It has the further property, however, that the result of integrating over any range of the *X*-axis which includes x_0 is a unit area. It can be considered as the limiting case of any pulse with unit area whose length is made infinitely short at the same time as its height is made infinitely large, while retaining the unit area. The unit delta function can be weighted by a scaling factor (with or without physical dimensions) so that the result of integrating over it gives the value of the weighting. The delta function provides a means of treating functions which are infinitely narrowly localised on an axis at the same time as other functions which are distributed along the axis. A typical case is that of a single discrete frequency component which is to be represented in terms of its “power spectral density” (see Section 2.2.5). Because of the infinitely narrow localisation of a discrete frequency component on a frequency axis, its spectral **density** (power per unit frequency), will be infinitely high, but since it represents a certain finite power, it can be considered as a delta function weighted by this value of power.

2.2. FOURIER ANALYSIS

The mathematical basis of frequency analysis is the Fourier Transform which takes different forms depending on the type of signal analyzed. All have in common that the signal is assumed to be composed of a number (perhaps an infinite number) of (co-)sinusoidal components at various frequencies, each having a given amplitude and initial phase. A typical (co-)sinusoidal component with amplitude A , initial phase ϕ and circular frequency f [§] is illustrated in Fig. 2.2(a).

The representation of Fig. 2.2(a) has the disadvantage that both time and phase angle are represented along the X-axis and can thus be easily confused. Furthermore, the mathematically useful concept of negative frequency is either meaningless or in any case unclear.

Fig. 2.2(b) illustrates another representation of the same sinusoidal component, this time as the vector sum of two contra-rotating vectors, each with amplitude $A/2$. One has initial phase angle ϕ and rotates with frequency f while the other has initial phase $-\phi$ and rotates with frequency $-f$. The concept of

[‡] Ref. 1.1 contains a mathematically rigorous discussion of delta functions and their relationship to Fourier Analysis

[§] $f = \frac{1}{T}$ where T is the periodic time

negative frequency now has the clear physical interpretation of indicating a negative rate of change of phase angle, and is necessary in order to indicate rotation in opposite directions.

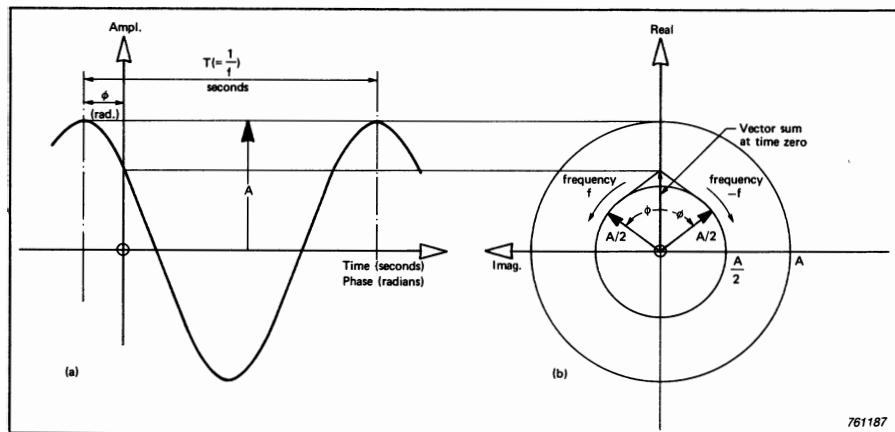


Fig. 2.2. (a) Typical sinusoidal component $A \cos(2\pi ft + \phi)$
 (b) Representation as sum of contra-rotating vectors

Fig. 2.2(b) represents the position of the vectors at time zero, but it can be seen that as they rotate with time, the vector sum will always be real (the imaginary parts cancel out) and will trace out the sinusoidal curve illustrated in Fig. 2.2(a).

The equivalence of the two forms is contained in the mathematical identity:

$$A \cos \theta = \frac{A}{2} (e^{j\theta} + e^{-j\theta}) \quad (2.8)$$

where $\theta = (2\pi ft + \phi)$ as in Equation (2.6)

2.2.1. Fourier Series

The application of this to Fourier analysis can be understood by considering the case of Fourier series expansion of a periodic function.

If $g(t)$ is a periodic function i.e.

$$g(t) = g(t + nT) \quad (2.9)$$

where T is the periodic time
and n is any integer

then it can be shown (Ref. 2.1) that it can be represented as a sum of sinusoidal components (or equivalently rotating vectors) at equally spaced frequencies kf_1 where f_1 ($= 1/T$) is the reciprocal of the periodic time and k is an integer (including zero and negative integers). The k th component is obtained from the integral

$$G(f_k) = \frac{1}{T} \int_{-T/2}^{T/2} g(t) e^{-j2\pi f_k t} dt \quad (2.10)$$

where $f_k = kf_1$

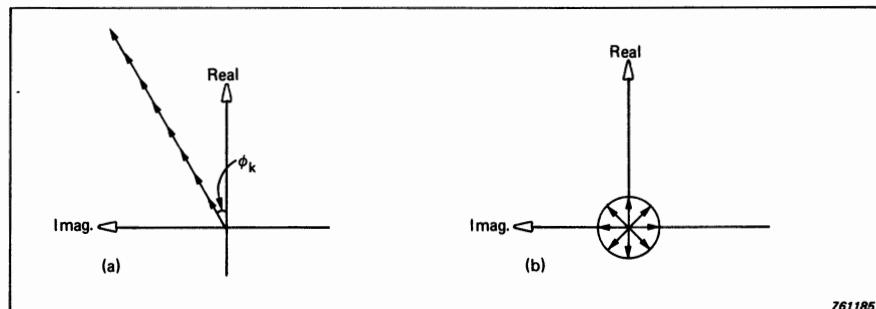
(i.e. the k th harmonic of f_1)

It is worthwhile examining in detail what this integral achieves. If the signal $g(t)$ contains a component rotating at a frequency of f_k , then multiplication by the unit vector $e^{-j2\pi f_k t}$ (which rotates at $-f_k$) annuls the rotation of the signal component such that it integrates with time to a finite value. (Fig. 2.3(a)).

All components at other frequencies will still rotate even after multiplication by $e^{-j2\pi f_k t}$ and will thus integrate to zero over the periodic time. (Fig. 2.3(b)).

It is of interest here that the effect of multiplying the signal by $e^{-j2\pi f_k t}$ is in fact to shift the frequency origin to frequency f_k and thus all the original harmonics will still be harmonics of the modified time signal.

Thus Equation (2.10) has the effect of extracting from $g(t)$ the components it contains which rotate at each frequency f_k . At the same time, as illustrated in



*Fig. 2.3. (a) Integration of a non-rotating vector to a finite value
(b) Integration of a rotating vector to zero*

Fig. 2.3(a) it "freezes" the phase angle of each as that existing at time zero (when $e^{-j2\pi f_k t} = 1$). The actual position of each vector at any other time t can thus be obtained by multiplying its initial value $G(f_k)$ by the oppositely rotating unit vector $e^{j2\pi f_k t}$ and the total signal $g(t)$ will thus be the (vector) sum of all these vectors in their instantaneous positions, i.e.

$$g(t) = \sum_{k=-\infty}^{\infty} G(f_k) e^{j2\pi f_k t} \quad (2.11)$$

The series of (complex) values $G(f_k)$ are known as the spectrum components of $g(t)$ and since there is an amplitude and phase (or equivalently real and imaginary part) associated with each one, a full representation requires three dimensions. Fig. 2.4 illustrates such a 3-dimensional representation of a typical spectrum.

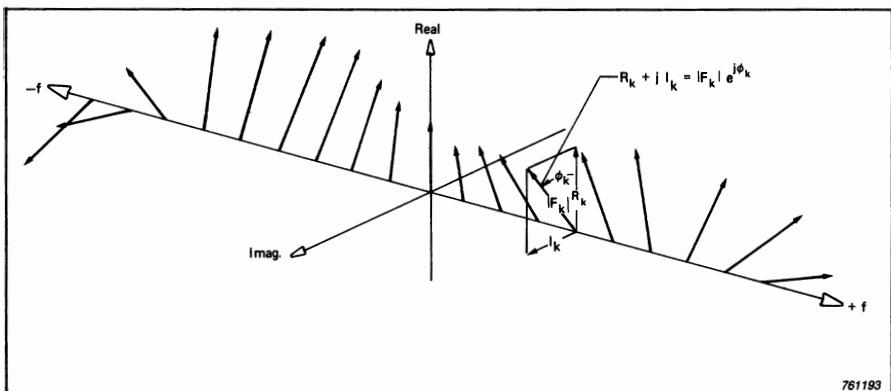


Fig. 2.4. 3-dimensional representation of the spectrum of a periodic function

Several important remarks can be made at this stage. One is that a signal which is periodic in the time domain has a spectrum which is discrete and in which all components fall at frequencies which are integral multiples of the fundamental frequency f_1 . The reason for this can easily be understood. The time for one rotation of the vector at the fundamental frequency f_1 corresponds to one period time T . Since all the other vectors rotate at speeds which are integer multiples of f_1 , they will all rotate an integer number of turns during this time and will all have returned to their starting positions. Thus, after time T the whole process will begin to repeat itself exactly and the function will clearly be periodic with period T .

Another important observation is that if the function $g(t)$ is real-valued (as it generally will be in the case of physical signals) then each component at frequency f_k must be matched by a component at $-f_k$ which has equal amplitude but opposite phase (or equivalently equal real part and opposite imaginary part). In this way the imaginary parts at all frequencies will always cancel and the resultant will always be real. This is obvious for the single sinusoidal component of Fig. 2.2 and applies to each component in the more general case. Mathematically, it can be said that the spectrum of a real-valued function is “conjugate even” i.e.

$$G(f_k) = G^*(-f_k) \quad (2.12)$$

where G^* is the complex conjugate of G .

One immediate result of this is that since the series of imaginary parts (or equivalently phase angles) is antisymmetric around zero frequency, then the zero frequency (or DC) component has zero (or $\pm\pi$) phase angle and is always real, which is intuitively obvious. Fig. 2.4 has been drawn to represent such a conjugate even spectrum.

A very important relationship concerns the distribution with frequency of the power content of the signal. The instantaneous power [§] of the time signal $g(t)$ is equal to $[g(t)]^2$ and the mean power over one period (and thus over any number of periods) is given by integrating the instantaneous value over one period and dividing by the periodic time.

Thus,

$$P_{\text{mean}} = \frac{1}{T} \int_0^T \{ g(t) \}^2 dt \quad (2.13)$$

For the typical sinusoidal component $A_k \cos(2\pi f_k t + \phi_k)$ this results in

$$\begin{aligned} P_{\text{mean}} &= \frac{1}{T} \int_0^T A_k^2 \cos^2(2\pi f_k t + \phi_k) dt \\ &= \frac{A_k^2}{T} \int_0^T \frac{1}{2} - \frac{1}{2} \cos 2(2\pi f_k t + \phi_k) dt \\ &= \frac{A_k^2}{2} \end{aligned}$$

since the sinusoidal part integrates to zero over the two periods of frequency $2f_k$ within time T .

This is the well-known result for the mean square value of a sinusoid of peak amplitude A_k and results in a root mean square (RMS) value of $A_k/\sqrt{2}$ (i.e. 0.707 A_k).

[§] The word “power” is here used in the sense of a squared variable (independent of the units of that variable) and is thus only related to physical power by a dimensioned scaling constant

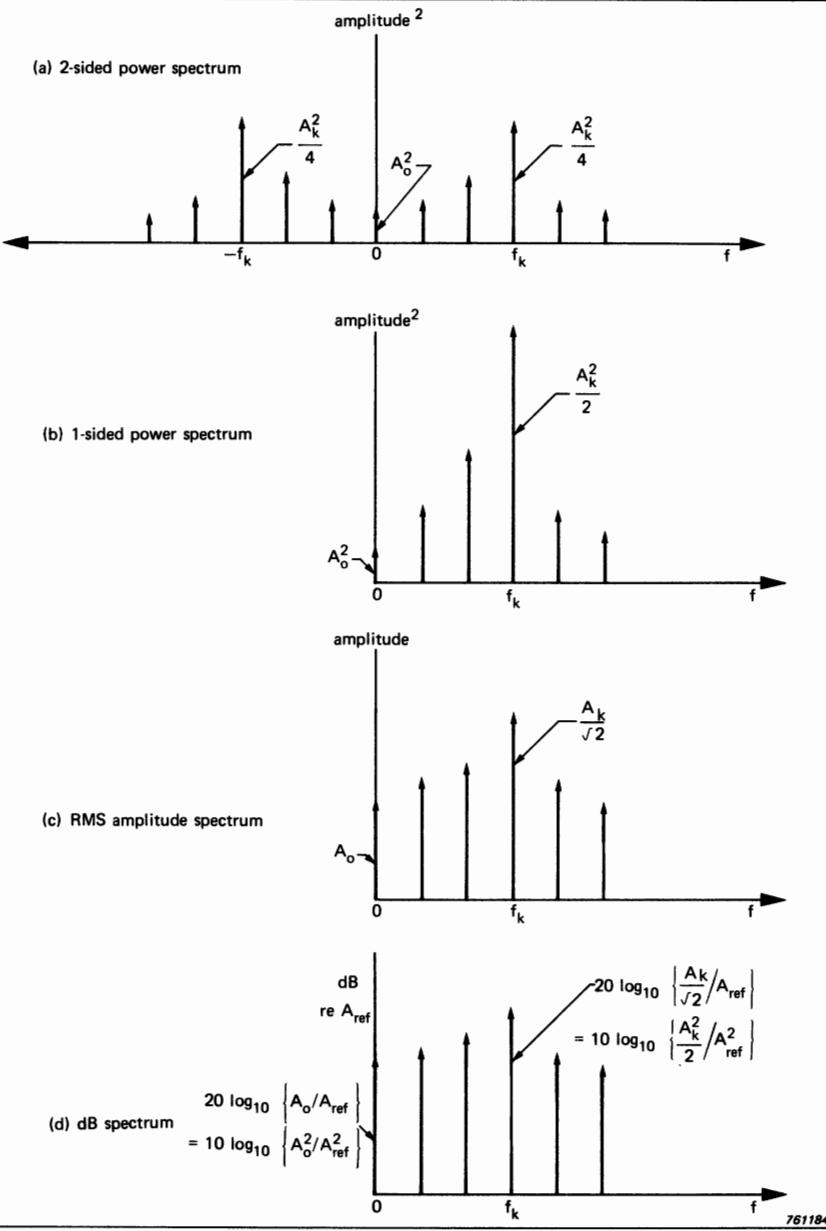


Fig. 2.5. Various spectrum representations

It will be found that the power content at each frequency is given directly by the square of the amplitude of the Fourier series component. We have seen (Fig. 2.2) that except for the DC component the amplitude of $G(f_k)$ is $A_k/2$, where A_k is the peak amplitude of the k th sinusoid, and thus the square of this is $A_k^2/4$. Since the amplitude spectrum is even, the negative frequency component (from $G(-f_k)$) will also be $A_k^2/4$, and thus the total power associated with frequency f_k will be $A_k^2/2$, the same as obtained in the time domain. The total power can thus be obtained either by integrating the squared instantaneous signal amplitude with time (and dividing by this time) or by summing the squared amplitudes of all the frequency components. This is one manifestation of the so-called "Parseval's Theorem" (Ref. 2.1).

The spectrum of squared amplitudes (all real) is known as the "power spectrum", and this is often the most useful part of the whole spectrum. However, since the initial phase information is lost, it is not possible to resynthesize the original time signal from the power spectrum.

Fig. 2.5(a) illustrates the 2-sided power spectrum corresponding to Fig. 2.4, and Fig. 2.5(b) the 1-sided power spectrum obtained by adding the negative frequency components to their positive counterparts (thus doubling them). This is the normal representation of the power spectrum and that corresponding to measurements with practical filters, since the latter pass positive and negative frequency contributions equally. Fig. 2.5(c) shows the spectrum of RMS values (the square roots of the values in 2.5(b)) while Fig. 2.5(d) shows the dB amplitude spectrum, which is defined alternatively as $10 \log_{10}$ (Mean square) or $20 \log_{10}$ (RMS) which of course gives the same result in a particular case. The values inside the brackets must be ratios with respect to a specified reference level. Note that the DC component is the same in both the 1-sided and 2-sided representations, since the total power at zero frequency is contained in this one component.

2.2.2. Fourier Transform

All the above results apply to periodic signals but it is possible to extend Equation (2.10) to a more general case by letting $T \rightarrow \infty$, in which case the spacing $1/T$ between the harmonics tends to zero and $G(f)$ becomes a continuous function of f . It can be shown (Ref. 2.1) that Equation (2.10) tends to

$$G(f) = \int_{-\infty}^{\infty} g(t) e^{-j2\pi ft} dt \quad (2.14)$$

and Equation (2.11) becomes

$$g(t) = \int_{-\infty}^{\infty} G(f) e^{j2\pi ft} df \quad (2.15)$$

Equation (2.14) is known as the “forward” transform and Equation (2.15) as the “inverse” transform while together they form the “Fourier Transform Pair”. It can be seen that they are almost symmetrical. The only difference is the sign of the exponent of e . The most important thing about the symmetry is that results which apply to transformation in one direction generally also apply to transformation in the other direction. Fig. 2.6 compares the Fourier Integral Transform with Fourier Series and other degenerate forms which are to be discussed.

2.2.3. Sampled Time Functions

Another form of the Fourier Transform pair applies to sampled time functions, i.e. functions which are represented by a “Time-Series”, a sequence of values at discrete equi-spaced points in time. This case is becoming very important with the increase in digital processing of data.

It can be seen that this is a situation which is the reverse of the Fourier Series case (as illustrated in Fig. 2.6(b)) and because of the symmetry of the Fourier transform pair, it happens that the spectrum becomes periodic, with a period equal to the sampling frequency f_s (the reciprocal of the time interval Δt between samples). This situation is represented in Fig. 2.6(c), and the reason for the periodicity will become clear in the following discussion of the Discrete Fourier Transform.

The particular form the Fourier Transform takes for sampled time functions is as follows:

$$G(f) = \sum_{n=-\infty}^{\infty} g(t_n) e^{-j2\pi f t_n} \quad (2.16)$$

$$g(t_n) = \frac{1}{f_s} \int_{-f_s/2}^{f_s/2} G(f) e^{j2\pi f t_n} df \quad (2.17)$$

where $t_n = n \Delta t$, i.e. the time corresponding to the n th time sample.

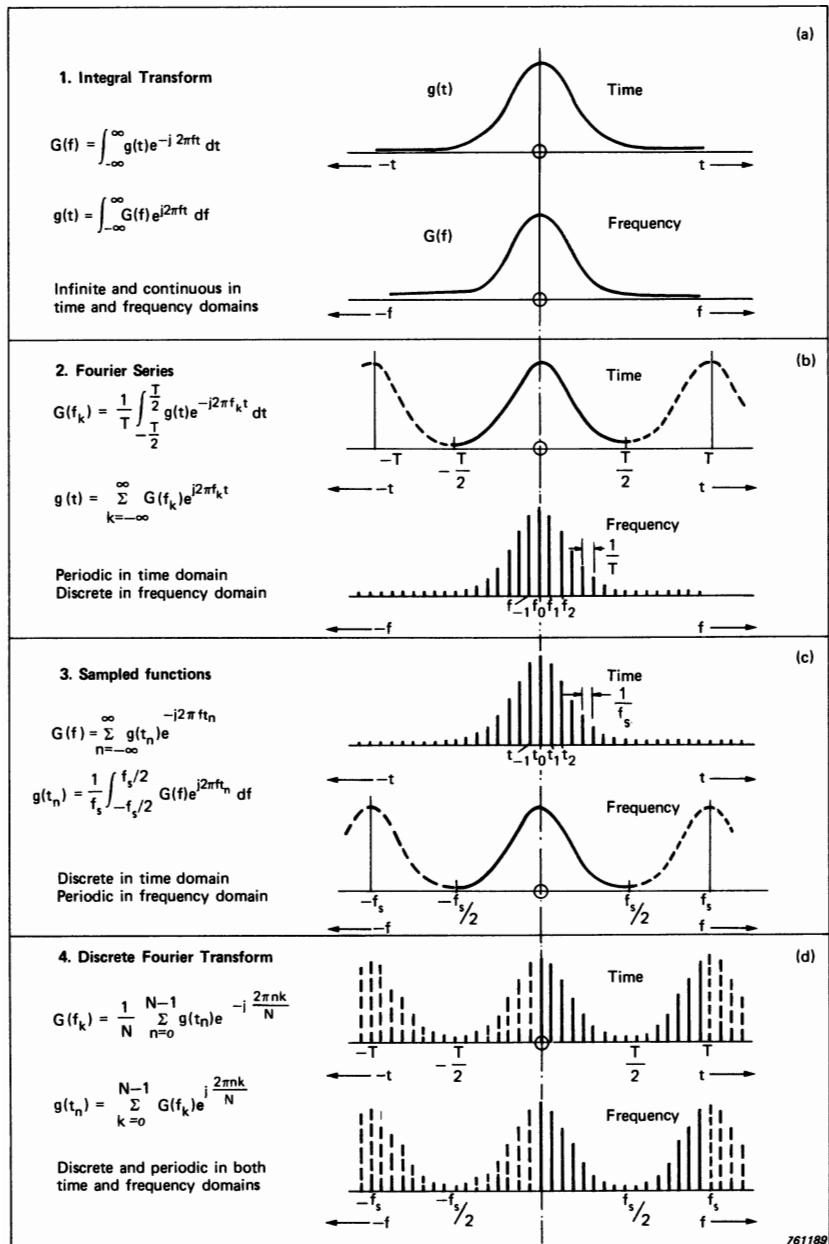


Fig. 2.6. Various forms of the Fourier transform

2.2.4. Discrete Fourier Transform

The final possibility occurs when the functions are sampled in both time and frequency domains as illustrated in Fig. 2.6(d). Because of the sampling, it is evident that both time signal and frequency spectrum are implicitly periodic in this case, and this periodicity (or “circularity”) leads to some interesting effects which will be discussed later.

The forward transform now takes the form[†]

$$G(k) = \frac{1}{N} \sum_{n=0}^{N-1} g(n) e^{-j \frac{2\pi kn}{N}} \quad (2.18)$$

and the inverse transform takes the form[†]

$$g(n) = \sum_{k=0}^{N-1} G(k) e^{j \frac{2\pi kn}{N}} \quad (2.19)$$

Because the infinite continuous integrals of Equations (2.14) and (2.15) have been replaced by finite sums, the above transform pair, known as the “Discrete Fourier Transform” or DFT, is much better adapted to digital computations. Even so, it can be seen that in order to obtain N frequency components from N time samples (or vice versa) requires N^2 complex multiplications. A calculation procedure, known as the “Fast Fourier Transform” or FFT algorithm, which obtains the same result with a number of complex multiplications of the order of $N \log_2 N$, is discussed in detail in Chapter 4. The reduction factor in computation time is thus of the order of $N/\log_2 N$, which for the typical case of $N = 1024$ (2^{10}) is more than 100.

However, while leaving the FFT till later, it is useful at this stage to look at the properties of the DFT, since this gives an insight into many fundamental concepts, e.g. sampling theory. One way of interpreting Equation (2.18) is as the following matrix equation:

$$\overline{G}_k = \frac{1}{N} \mathbf{A} \overline{g}_n \quad (2.20)$$

[†] Note that for convenience the time and frequency functions have in this case not been made symmetrical about the origin, but because of the periodicity of each, the second half also represents the negative half period to the left of the origin. Note also that parameter k refers to frequency f_k and n to time t_n .

where \overline{G}_k is a column array representing the N complex frequency components,

$1/N$ is a simple scaling factor,

\mathbf{A} is a square matrix of unit vectors and

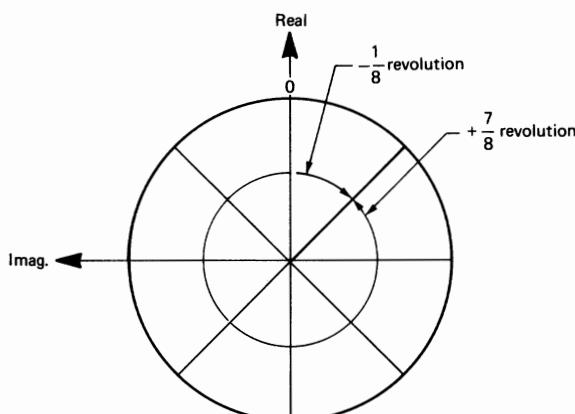
\overline{g}_n is a column array representing the N time samples.

For the particular case of $N = 8$, the Equation (2.20) may be visualized as follows:

$$\begin{bmatrix} G_0 \\ G_1 \\ G_2 \\ G_3 \\ G_4 \\ G_5 \\ G_6 \\ G_7 \end{bmatrix} = \frac{1}{8} \begin{bmatrix} \uparrow & \uparrow \\ \uparrow & \nearrow & \rightarrow & \uparrow & \swarrow & \downarrow & \leftarrow & \nearrow \\ \uparrow & \rightarrow & \downarrow & \leftarrow & \uparrow & \rightarrow & \downarrow & \leftarrow \\ \uparrow & \swarrow & \leftarrow & \uparrow & \nearrow & \downarrow & \rightarrow & \swarrow \\ \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \downarrow \\ \uparrow & \nearrow & \swarrow & \rightarrow & \uparrow & \swarrow & \uparrow & \nearrow \\ \uparrow & \leftarrow & \downarrow & \rightarrow & \uparrow & \downarrow & \leftarrow & \rightarrow \\ \uparrow & \uparrow \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \\ g_6 \\ g_7 \end{bmatrix} \quad (2.20 \text{ a})$$

Each element in the square matrix represents a unit vector $e^{-j2\pi kn/N}$ with a certain angular orientation, and multiplication by this vector results in a rotation through the angle depicted. Each row in the matrix represents a different value of frequency ($k = 0, 1, 2 \dots 7$) while each column represents a different point in time ($n = 0, 1, 2 \dots 7$).

For either k or n equal zero the angle is always zero and thus multiplication is by unity. The first row of the matrix ($k = 0$) represents zero frequency, and since



770136

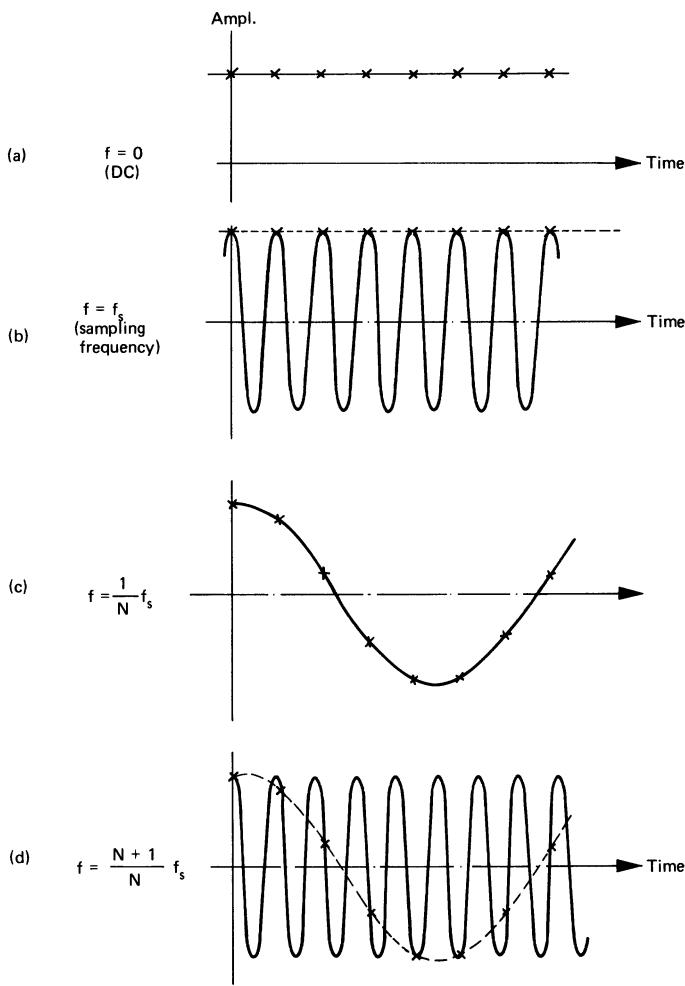
Fig. 2.7. Equivalence of positive and negative rotations for sampled functions

all elements equal one, calculation of G_0 involves a simple addition of all the time samples (g_n) followed by division by 8 ($= N$). As would be expected this results in the DC component. The second row ($k = 1$) represents the lowest non-zero frequency and it can be seen that for increasing values of n the angle changes by $2\pi/N$ i.e. $1/N$ th of a revolution. (Note that for the forward transform the negative sign of the exponent actually gives a rotation in the negative direction as explained previously). For $k = 2$ the rotational frequency is $2/N$ ths of a revolution per time sample and so on up to the last row which represents $(N-1)/N$ (in this case $7/8$) revolution per time sample. Note that this can be more easily interpreted as a rotation, in the opposite direction, of $1/N$ th of a revolution per time sample and thus equally well represents the frequency $-2\pi/N$ per time sample (Fig. 2.7). In fact all the frequencies above $k = N/2$ (in this case 4) are more easily interpreted as negative frequencies and this is perhaps the easiest way of understanding Shannon's Sampling Theorem[†]. We have seen (Equation 2.12) that the negative frequency components of a real-valued time function are determined with respect to the positive frequency components and thus if there were any frequency components above the Nyquist frequency (half the sampling frequency i.e. half a revolution per time sample) then these would become inextricably mixed with the required negative frequency components, thus introducing an error. The periodicity of the spectrum for sampled time functions also becomes clear from examination of Equation (2.20a). The first row of \mathbf{A} could equally well represent the sampling frequency f_s (one revolution per time sample) or $2f_s$, and so on, and thus the \mathbf{A} matrix could equally represent the frequencies $k = 8$ to 15 , 16 to 23 etc. (Fig. 2.8). Since the rotating vectors are only sampled at discrete points in time, all information is lost about how many complete revolutions may occur between samples. However, restriction of frequency content to less than half the sampling frequency removes the ambiguity.

The misinterpretation of high frequencies (above half the sampling frequency) as lower frequencies, as illustrated in Fig. 2.8, is termed "aliasing", and this is obviously one of the pitfalls to be avoided when digitizing continuous signals. It may help in understanding aliasing, to consider two practical cases with which most people are familiar.

- (1) The cartwheels in western films often appear to run backwards (i.e. negative frequency) or too slowly forwards because of the sampling involved in filming.
- (2) The stroboscope is in fact an aliasing device which is designed to represent high frequencies as low ones (even zero frequency when the picture is frozen).

[†] Shannon's Sampling Theorem states that a sampled time signal must not contain components at frequencies above half the sampling rate (the so-called Nyquist frequency)



761195

Fig. 2.8. Illustration of “aliasing”

(a) Zero frequency or DC component

(b) Component at sampling frequency f_s interpreted as DC

(c) Frequency component at $(1/N)f_s$

(d) Frequency component at $[(N+1)/N]f_s$

2.2.5. Spectral Density

Despite the similarity between the various forms of the Fourier transform, it is important at the same time to point out the differences between them. For example, with those forms where the spectrum is a continuous function of frequency (Equations (2.14), (2.16)) the spectral components have the dimensions of "spectral density". In particular, the amplitude squared spectrum typically has the units of energy per unit frequency (e.g. Volt² seconds/Hz) and must be integrated over a finite bandwidth to give a finite energy. The term "power spectral density" (PSD) is used for stationary random functions, whose spectrum is continuous, but which have a finite and (statistically) constant power. To apply this concept to stationary deterministic signals (with discrete spectra) involves the use of delta functions as described in Section 2.1 and it is more common to represent their spectra as "power spectra" scaled directly in power units (cf. Fig. 2.5).[†]

2.2.6. Power vs. Energy

A further difference in dimensions becomes evident when comparing Equation (2.14) with Equation (2.10). Equation (2.14) is intended to apply to transients (with finite total energy) and if for example it is applied to a tone burst, then the value of the integral increases directly as the length of the tone burst increases. In contrast, Equation (2.10) applies to periodic (stationary) signals, and the division by T normalises the result to a (virtually) constant value independent of the length taken into account even if T is allowed to extend over several periods. Consequently, the spectrum which results from squaring the amplitudes of the components obtained from Equation (2.14) has the dimensions of "energy spectral density". In cases where a PSD analyzer is used to analyze an impulse which has been recorded on a loop (tape loop or recirculating digital memory) the power spectral density units obtained must be multiplied by the loop length (i.e. repetition time) to obtain the correct results in terms of energy spectral density. If the power spectrum has been measured, then the results must also be divided by the analyzer bandwidth to obtain the spectral density. It is a precondition that the analyzer bandwidth must be less than the bandwidth of the function being analyzed for such a result to be correct (see Section 2.4.4 for a more detailed discussion of this).

[†] Strictly speaking, none of the forms of the Fourier transform shown in Fig. 2.6 applies to stationary non-periodic signals, and for example it would be necessary to normalise Equation (2.14), by dividing by a factor proportional to the long-term limit of the integral (e.g. T for deterministic signals, \sqrt{T} for random signals) before taking the limit as $T \rightarrow \infty$. In practice, this mathematical finesse is not necessary, because all practical signals are of finite duration and can thus be treated as transients (to which Equation (2.14) applies) even when the results are interpreted in terms of the equivalent infinitely long signal, which is a mathematical abstraction. The various signal types mentioned here are discussed in detail in Section 2.4.

2.3. BANDWIDTH

So far the concept of "bandwidth" has not been explained, since the results of a mathematical Fourier analysis have infinitely narrow bandwidth (df). This is never possible in practice, for reasons which will become obvious, and so the concept of bandwidth must be introduced.

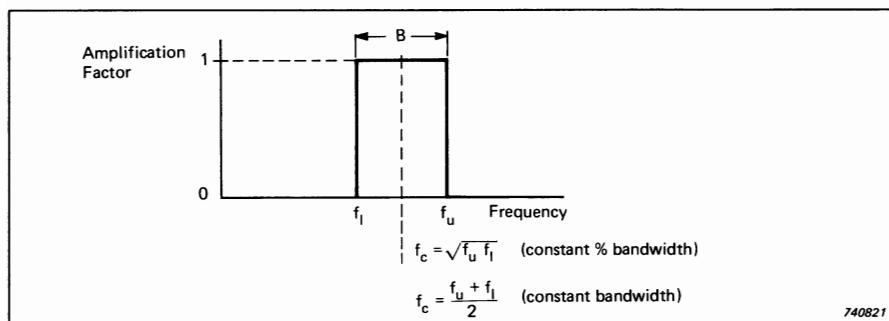


Fig. 2.9. Ideal filter

The term originates from the use of bandpass filters, which have the property of passing only that part of the total power whose frequency lies within a finite range (the bandwidth). The concept can be understood from consideration of the so-called "ideal filter" whose power transmission characteristics are illustrated in Fig. 2.9. This transmits, at full power, all components lying within its passband of width B and attenuates completely all components at other frequencies.

The concept of bandwidth can also be extended to mean the degree of frequency uncertainty associated with a measurement. This applies directly to the case of the ideal filter, in the sense that the frequency of a transmitted component can only be said to lie somewhere in the bandwidth. Practical filters have a characteristic which differs from that of an ideal filter as illustrated in Fig. 2.10. These differences are discussed in more detail in the next chapter, but at this stage it is useful to consider the meaning of bandwidth in this case, since it is no longer immediately obvious. The so-called "effective noise bandwidth" is defined as being the width of ideal filter with the same reference transmission level which transmits the same power from a white noise source (which has a PSD which is constant with frequency). It can be obtained by integrating the area under the power transmission curve (the shaded area in Fig. 2.10) and dividing by the reference level.

The bandwidth associated with a measurement is not necessarily determined by the bandwidth of the filter used (in particular where none is used such as in

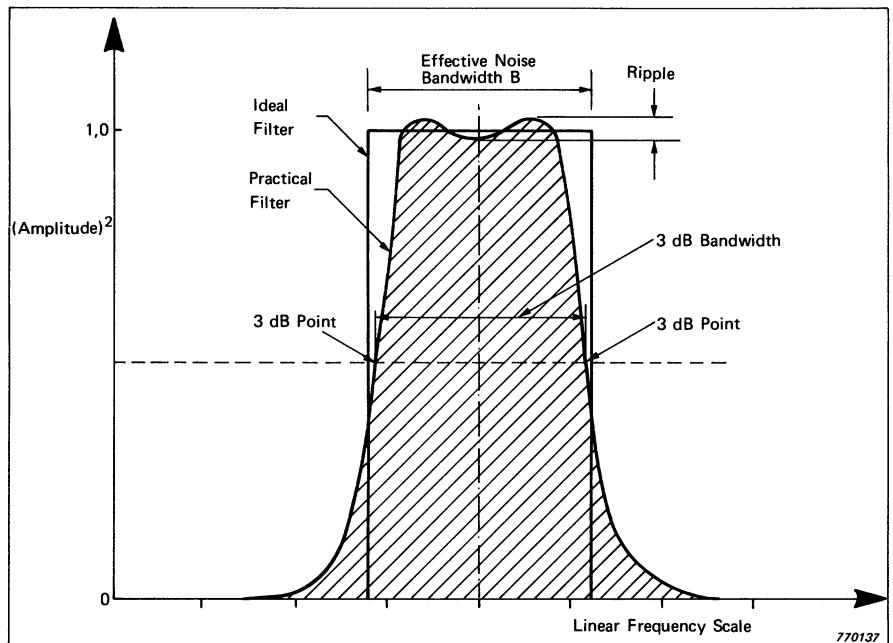


Fig. 2.10. Practical vs. ideal filter

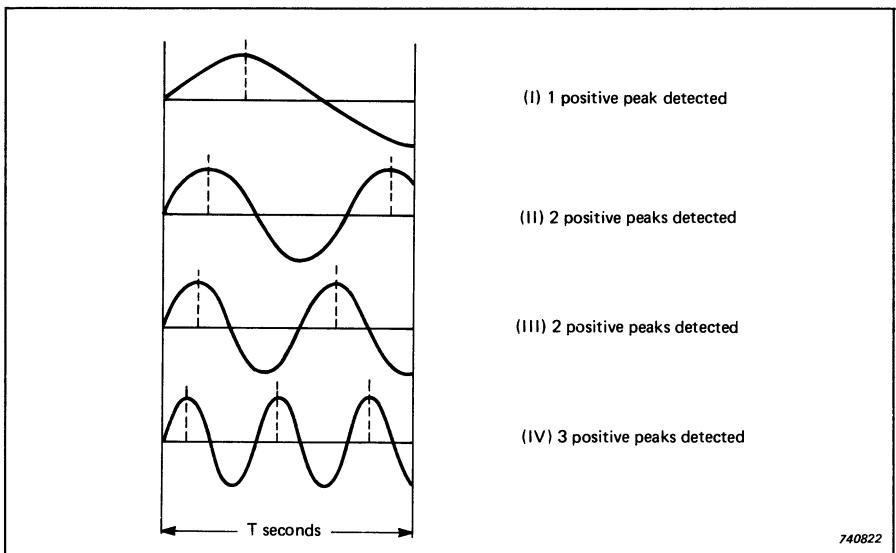


Fig. 2.11. Measurement of frequency by counting peaks

digital calculations). The other factor which plays a role is the effective length of the record on which the measurement is made, and in fact a record length of T limits the bandwidth to a minimum of $1/T$.

This simple relationship may be demonstrated from the following model: suppose that it is necessary to measure the frequency of an unknown sine-like wave, and that the duration of the measurement is fixed at T seconds. With what bandwidth can the measurement be made? One way of making the measurement is to count the number of positive peaks of the sine wave and divide the number by T . This method is illustrated in Fig. 2.11.

Let us now examine which frequencies are measured for the examples of Fig. 2.11. In case (I), one peak is counted, and hence the frequency is $1/T$ Hz. In cases (II) and (III), 2 peaks are detected, and hence the frequency of both is $2/T$

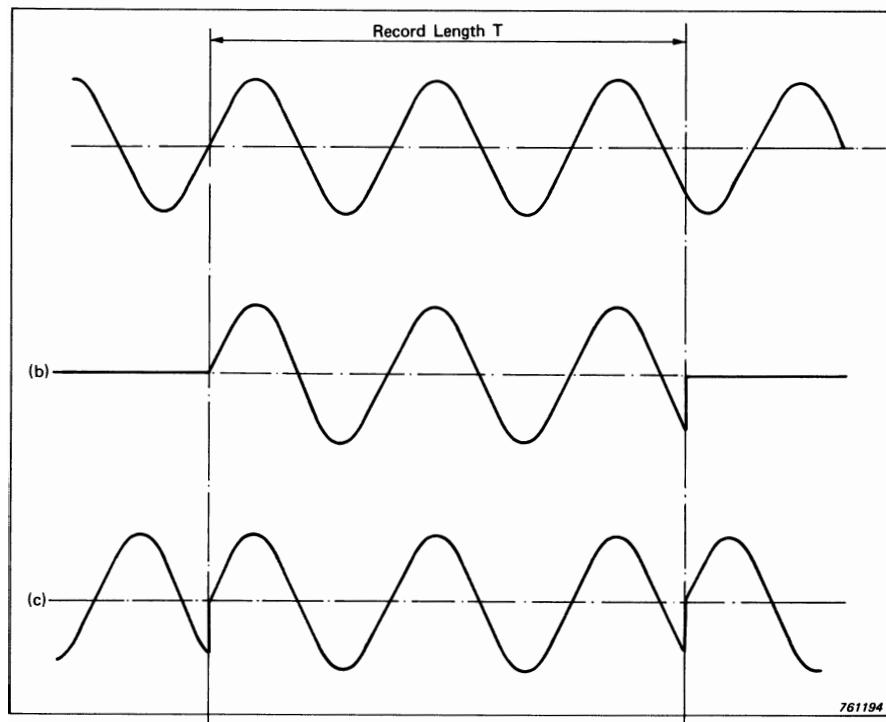


Fig. 2.12. Three possibilities for the signal of Fig. 2.11(IV)
(a) Continuous sine wave
(b) Tone-burst of length T
(c) Loop of length T

Hz. In case (IV), 3 peaks give a frequency of $3/T$ Hz. Hence, because we have limited the measurement period to T seconds, it is not possible to measure a sine wave frequency to better than $1/T$, i.e. the best bandwidth which can be obtained is $1/T$ Hz.

It might be argued that the frequency could be measured more accurately by extrapolating the sinewave to an integer number of periods and measuring this time exactly, but it must not be forgotten that it is not certain that the signal continues as a sinewave. Two other possibilities are illustrated in Fig. 2.12 for the signal of Fig. 2.11 (IV). In one case the signal is a tone burst of length T , while in the other case it is a periodic signal of length T . Even though these have different spectra, this can only be determined by using a bandwidth less than $1/T$. Thus the frequency uncertainty, or bandwidth, is equal to $1/T$.

The concept of a filter characteristic being associated with a record length of T can also be illustrated using the rotating vector analogy introduced previously. The discussion of periodic signals in Section 2.2.1 considered only the case where the frequency components all fell at multiples of f_1 , the fundamental frequency. Even after multiplication by $e^{-j2\pi f_k t}$, the rotation of any component would still be at an integer multiple of f_1 and would thus result in an integral number of rotations over the period time T . As illustrated in Fig. 2.3(b) this would always result in an integration to zero for components other than that with frequency f_k . If the signal contained a continuous range of frequencies, however, then the integration of the rotating vector over time T would not always be exactly zero. For example, if there is a component with a frequency of $f_k + \Delta f$, then after multiplication by $e^{-j2\pi f_k t}$, it will continue to rotate at Δf . If Δf is only a small percentage of f_1 , then the total rotation over time T would

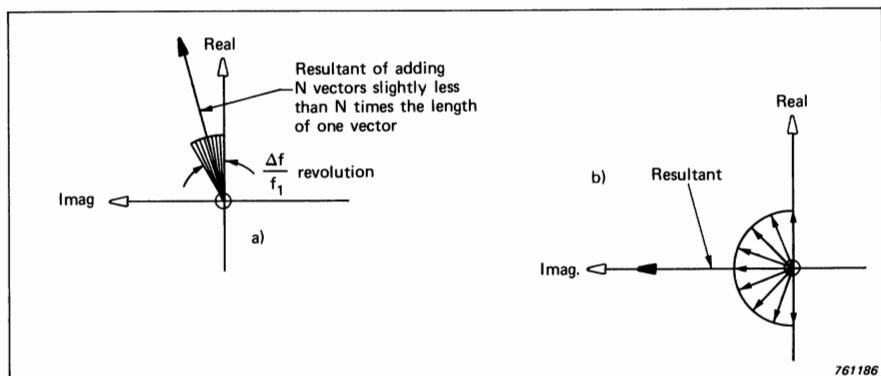


Fig. 2.13. (a) Resultant for $\Delta f \ll f_1$
 (b) Resultant for $\Delta f = f_1/2$

only be small part of a revolution, and the average would still be almost the same as with zero rotation (see Fig. 2.13(a)). Note, however, that the phase of the resultant is turned through a small angle.

When $\Delta f = f_1/2$ (i.e. the frequency lies halfway between f_k and f_{k+1}) then the vector will rotate through half a revolution in time T and the resultant is illustrated in Fig. 2.13(b).

The length of the resultant can be determined by integrating that component of each vector which is aligned with the resultant and in Appendix A it is shown how this results in

$$A_{\text{result}} = \frac{2}{\pi} A \quad (2.21)$$

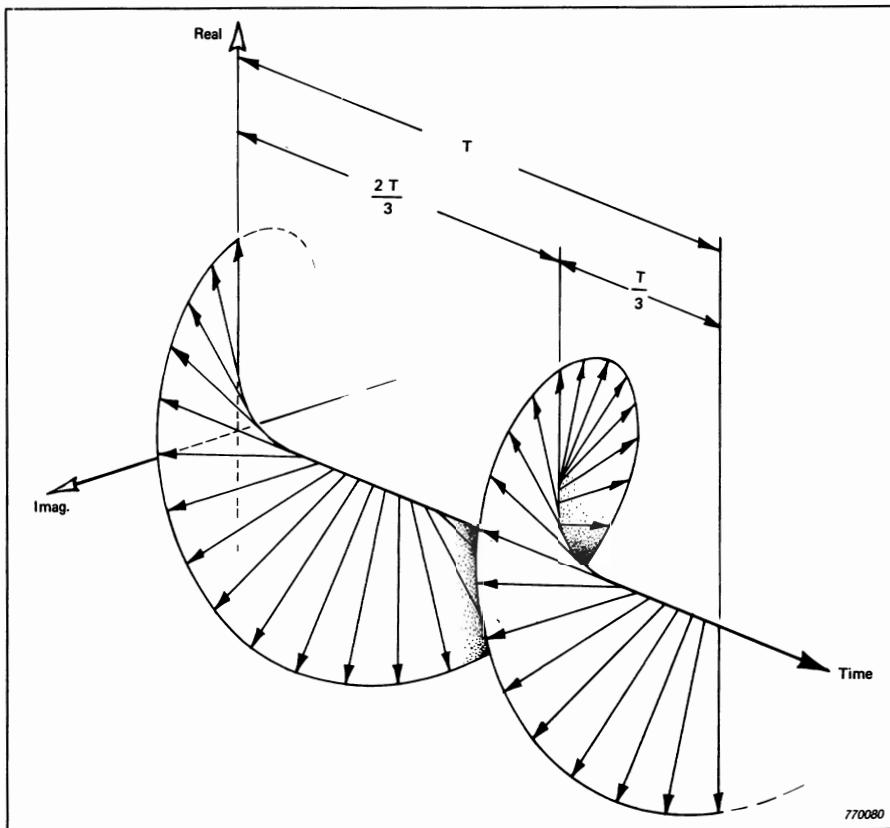


Fig. 2.14. Effect of $\Delta f = 3f_1/2$

This is approximately 3,9 dB less than A . In this case, the phase of the resultant is obviously turned through 90° .

Continuing on in the same way, one finds that the resultant at $\Delta f = f_1$ is of course zero, but for example with $\Delta f = 1,5 f_1$ the situation is as illustrated in three dimensions in Fig. 2.14 (where the axis normal to the complex plane now represents time rather than frequency).

In this case, the vector rotates through 1,5 revolutions in time T , where the resultant of the first revolution is zero and only the last half revolution gives a resultant. The length of the resultant will now evidently be $1/3$ of that for $\Delta f = f_1/2$ (since only $1/3$ of the total number of vectors contribute to it).

$$\text{i.e. } A_{\text{result}} = \frac{2}{3\pi} A \text{ (approx. 13,4 dB less than } A) \quad (2.22)$$

If the relative amplitude of the resultant is plotted against Δf it will in fact be found to trace the well-known $|\sin x/x|$ curve as illustrated in Fig. 2.15.

This can be considered as a filter characteristic with which the original signal is filtered when the record length is limited to T . Note that the power transmission characteristic is thus equal to $\sin^2 x/x^2$ (where $x = \pi \Delta f / f$) and as shown in Appendix A this has an effective noise bandwidth given by

$$B_{\text{eff}} = \frac{1}{T} \quad (2.23)$$

as previously stated.

Before leaving this analogy it is worth noting that the phase shift of the resultant is due to the fact that the time signal was not taken as symmetrical

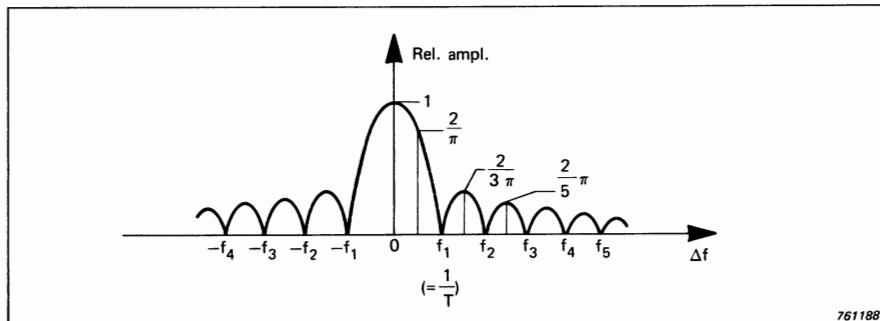
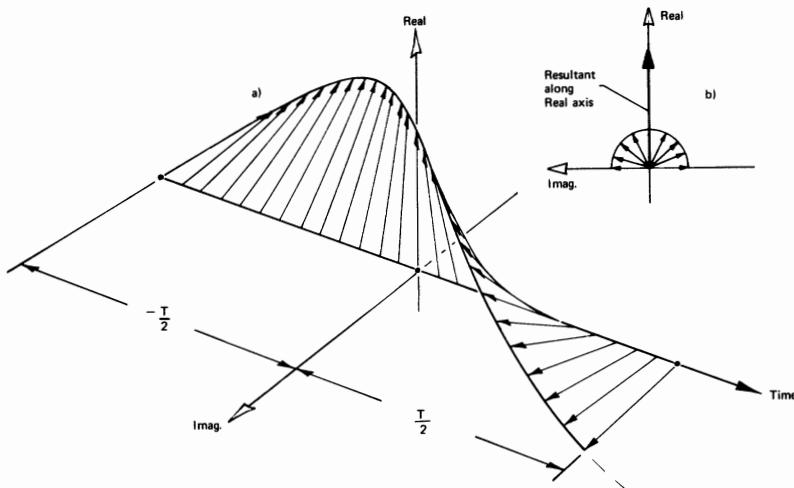


Fig. 2.15. Effective "filter characteristic" for record length T



*Fig. 2.16. Effect of symmetry about time zero
 (a) 3-dimensional representation
 (b) View along time axis*

770079

about zero time. If the integration period is chosen to be from $-T/2$ to $T/2$ and the original signal symmetrical (i.e. a cosine) the resultant will in all cases be directed along the Real Axis and no phase shift will be introduced. This is illustrated for the case of $\Delta f = f_1/2$ in Fig. 2.16, which should thus be compared with Fig. 2.13(b).

Quite a lot of space has been devoted to these fundamental relationships between record length (or "time window") and the resulting spectrum. The results can be obtained much more efficiently using the "Convolution Theorem" of Section 2.5.3, but it is thought that this elementary approach gives more physical insight, and helps in the interpretation of convolution.

2.4. SIGNAL TYPES

Before discussing convolution it is as well to examine the various types of signal which are encountered in practice. The type of signal to be analyzed has an influence on the type of analysis to be carried out and also on the choice of analysis parameters. Fig. 2.17 indicates the basic divisions into different signal types.

The most fundamental division is into stationary and non-stationary signals. A rigorous definition of stationary random functions is given in Ref. 2.2, but for

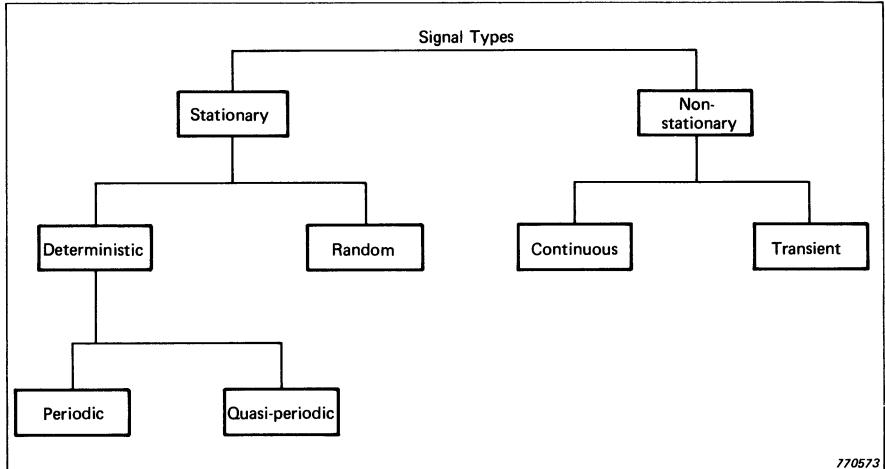


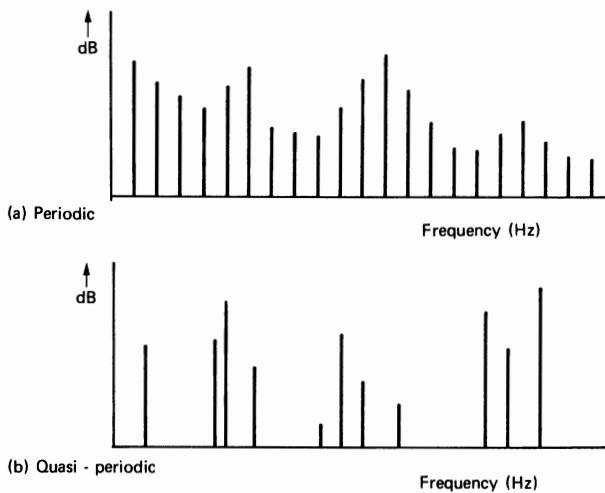
Fig. 2.17. Division into different signal types

practical purposes it is sufficient to interpret stationary functions as being those whose average properties do not vary with time and are thus independent of the particular sample record used to determine them. This applies to both deterministic and random signals, but in particular in the latter case it is important to realise that the results obtained from different records are not necessarily identical, just equally valid.

The instantaneous value of a stationary deterministic signal is predictable at all points in time, while with stationary random signals it is only the statistical properties such as mean values, variances etc., which are known.

Non-stationary signals may be roughly divided into continuous non-stationary signals (of which a good example is speech) and transient signals which may be defined as those which start and finish at zero. Of course in practice even signals such as speech must start and finish at some time, but the difference is perhaps more fundamentally that a transient is treated and analyzed as a whole, whereas a continuous non-stationary signal, such as speech, will normally be analyzed in short sections, each of which will often be quasi-stationary.

It is interesting to look at each of these signal types in more detail, so as to see how the differences show up in various representations.



740812

Fig. 2.18. Typical periodic and quasi-periodic spectra

2.4.1. Stationary Deterministic Signals

Stationary deterministic signals are made up entirely of sinusoidal components at discrete frequencies. In periodic signals, as we have seen, all these discrete frequencies are multiples of some fundamental frequency, the reciprocal of the periodic time. In quasi-periodic signals, the frequencies of the various sinusoids are not harmonically related. If carried to extremes, this means that the ratio between at least two frequencies must be an irrational number such as $\sqrt{2}$, but in practice it can be seen that quasi-periodic signals will typically arise from mixtures of two or more independent sets of harmonics, such as from an aircraft turbine engine with two independently rotating shafts.

Fig. 2.18(a) and (b) shows how the power spectra of typical periodic and quasi-periodic signals appear. The approach to frequency analyzing them is basically the same. The filter bandwidth should be selected so as to separate the most closely spaced components and in that case there will only be one sinusoid in the filter passband at one time. If this condition is satisfied, the transmitted power is independent of the bandwidth. It might be argued that the closest spacing will only be known after analysis, and in some cases it may be necessary to use a sort of trial-and-error process, but in many cases the likely location of frequency components will be known in advance, e.g. as harmonics of a machine rotational speed, or mains frequency.

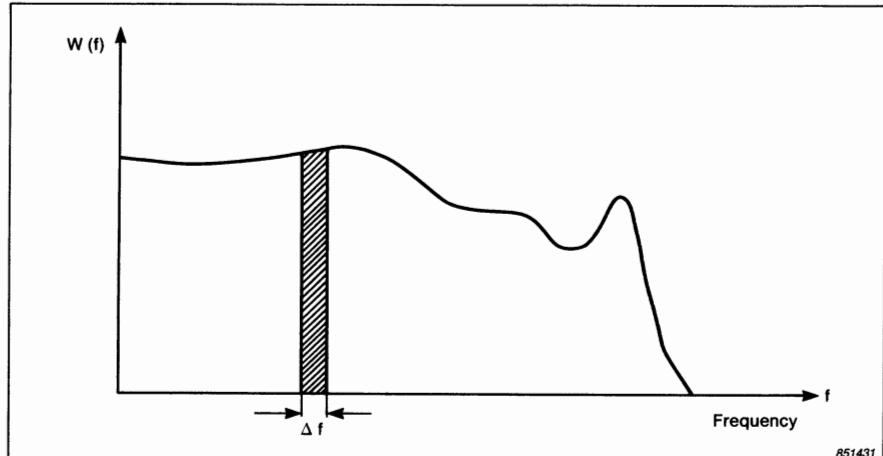


Fig. 2.19. Continuous spectrum of a stationary random signal

Normally, a constant bandwidth analysis on a linear frequency scale will be most appropriate to the analysis of deterministic signals, since harmonically related components will then be equally separated and resolved.

Occasionally, it is not possible to separate closely spaced components, and then it is important to realise that the signal passed by the filter will have a beat frequency equal to the difference between the most closely spaced components. The signal must then be analyzed over several periods of the beat frequency in order to be considered as stationary. This will occasionally be a more demanding requirement than that required in the analysis of random signals.

2.4.2. Stationary Random Signals

In contrast to deterministic signals, random signals have a spectrum which is continuously distributed with frequency, as shown in Fig. 2.19. Accordingly, the power transmitted by a filter varies with the bandwidth, and for a relatively flat spectrum is directly proportional to it. As mentioned in Section 2.2.5 it is possible to remove this influence of filter bandwidth by dividing the transmitted power by the bandwidth, thus normalising the result to a "power spectral density". The requirement that the spectrum is relatively flat will be satisfied if the filter bandwidth is chosen to be narrower (e.g. one-third of the width) of any peaks in the spectrum being measured. Once again a trial-and-error process may sometimes be required to determine whether this condition is satisfied, but in many cases it will be known in advance from the physical conditions. For example, peaks in the spectra of random signals often arise from filtration of a

broadband signal by a physical system whose frequency response is characterised by a number of resonances, each with a certain Q-factor (i.e. amplification factor). The range of Q-factors will often be determined by choice of materials etc., and since there is an inverse relationship between Q-factor and percentage bandwidth, the latter will often be known roughly in advance.

Even though the power spectrum of a random function may be well-defined, the phase spectrum is random, and thus normally not of interest in the analysis of single signals. In the analysis of multiple channel signals the phase relationships between two or more spectra will often be meaningful.

Even though their instantaneous value cannot be predicted, random signals may be characterised by their probability density curves. The meaning of this concept can be understood by reference to Fig. 2.20. The probability density $p(x)$ at some level x is defined as the probability that the signal value lies between x and $x + \Delta x$, divided by the interval width Δx (thus giving a density). Thus if $P(x)$ represents the total probability that the signal value is less than x , then

$$p(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x + \Delta x) - P(x)}{\Delta x} \quad (2.24)$$

Referring to Fig. 2.20, it will be seen that

$$P(x + \Delta x) - P(x) = \lim_{T \rightarrow \infty} \frac{\sum \Delta t_n}{T} \quad (2.25)$$

where each Δt_n represents one of the time intervals in T where the signal lies between x and $x + \Delta x$.

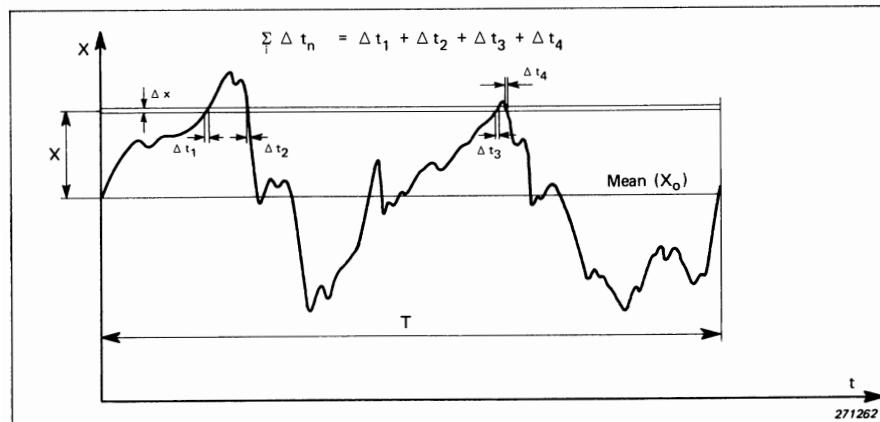


Fig. 2.20. Sketch illustrating the concept of probability density

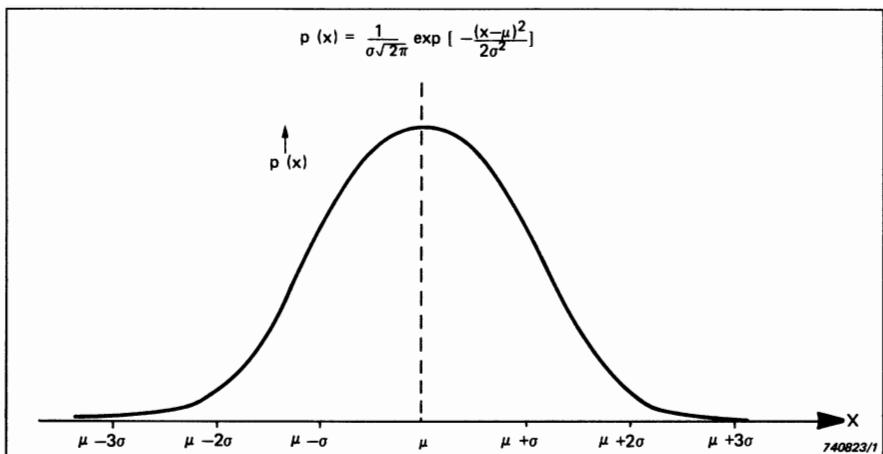


Fig. 2.21. Gaussian distribution

Gaussian random signals, which can be used as a model for many random signals encountered in practice, have a probability density curve of Gaussian shape, as illustrated in Fig. 2.21.

The equation of this curve is

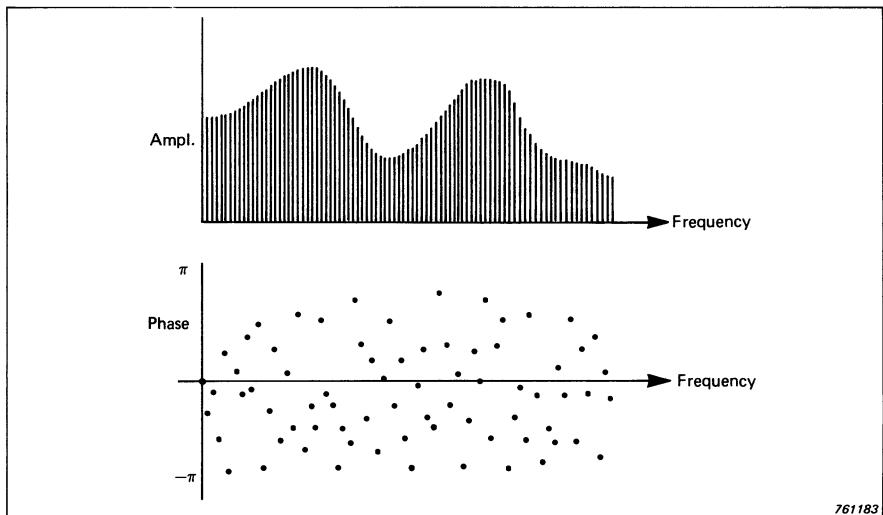
$$p(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} \quad (2.26)$$

This may appear somewhat formidable, but it is in fact just an e^{-x^2} curve centered on the mean value μ and scaled in the following way:

1. In the x -direction it is scaled in terms of σ , the standard deviation from the mean μ . For zero mean, σ is also the RMS level of the signal, and σ^2 the variance or power.
2. In the y -direction it is scaled so that the total integral under the curve for all x is 1, which of course is the probability that x can have any value between $\pm \infty$.

2.4.3. Pseudo-random Signals

Pseudo-random signals are a particular type of periodic signal sometimes used to simulate random signals. Even though periodic, the periodic time T is very long and thus the spectrum line spacing ($1/T$) very close (Fig. 2.22). Phase



761183

Fig. 2.22. Amplitude and phase spectra for a pseudo-random signal

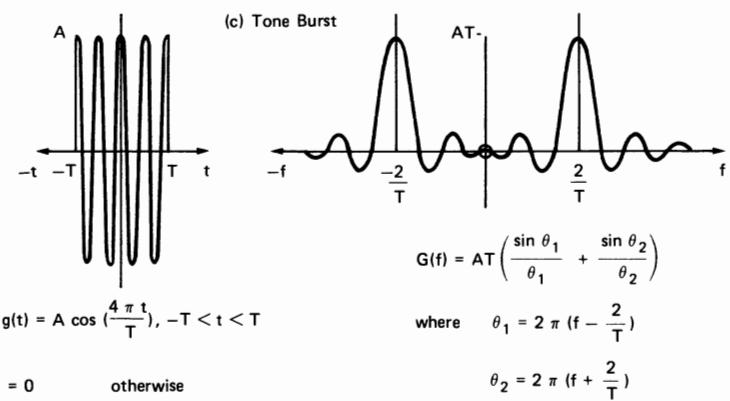
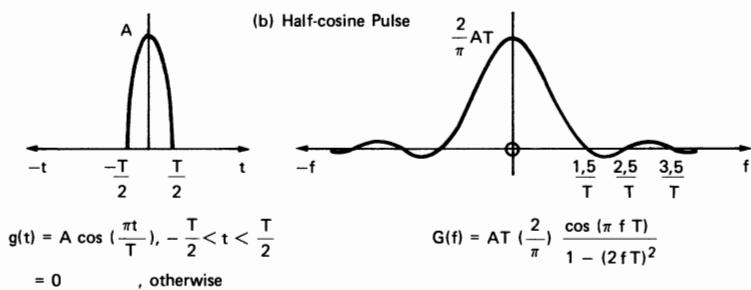
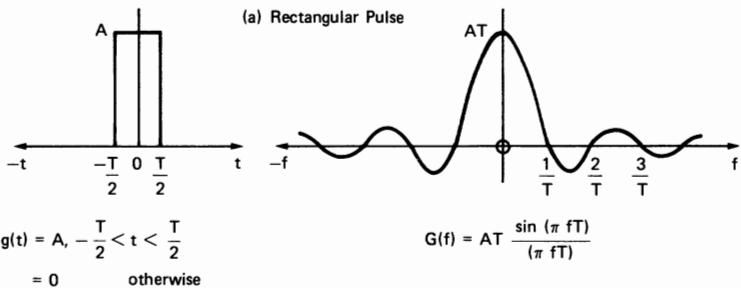
relationships between adjacent spectral lines are to all intents and purposes random, so that provided the bandwidth of any resonance peaks spans over a large number of spectral lines then the effect of applying it as input to a linear physical system will be very similar to that of a truly random signal. The probability density of such pseudo-random signals may be made very close to Gaussian.

On the other hand, a pseudo-random signal can be reproduced exactly, and this may be of benefit in the standardisation of testing. It must be remembered though, that the signal is periodic, and thus for example non-linearities are always excited in exactly the same way; one does not obtain the best linear approximation as one does with a truly random signal.

2.4.4. Transient Signals

As mentioned previously, a transient may be considered as a signal which commences and finishes at zero. Fig. 2.23 shows the following three typical examples:

1. A rectangular pulse
2. A half cosine pulse
3. A tone burst



770155

Fig. 2.23. Various transients and their spectra

(a) Rectangular pulse

(b) Half cosine pulse

(c) Tone-burst

together with their spectra as derived using Equation (2.14). It is emphasized that the squares of these spectrum amplitudes have units of **energy** spectral density, as explained in Section 2.2.6. Thus the integral of the squared spectrum amplitude over all frequency gives the total energy of the transient. This can also be obtained by integrating the instantaneous power (i.e. amplitude squared) over all time.

$$\text{i.e. } \int_{-\infty}^{\infty} |G(f)|^2 df = \int_{-\infty}^{\infty} |g(t)|^2 dt \quad (2.27)$$

This is a more general form of Parseval's theorem which was referred to in Section 2.2.1 (See also Ref. 2.1).

All the above examples are real even functions which transform to real even functions in the other domain. In the more general case, description of the overall spectrum requires that the phase spectrum be shown as well. As an example, Fig. 2.24 shows the amplitude and phase spectra for a rectangular pulse starting at zero time, and this can be compared with Fig. 2.23(a).

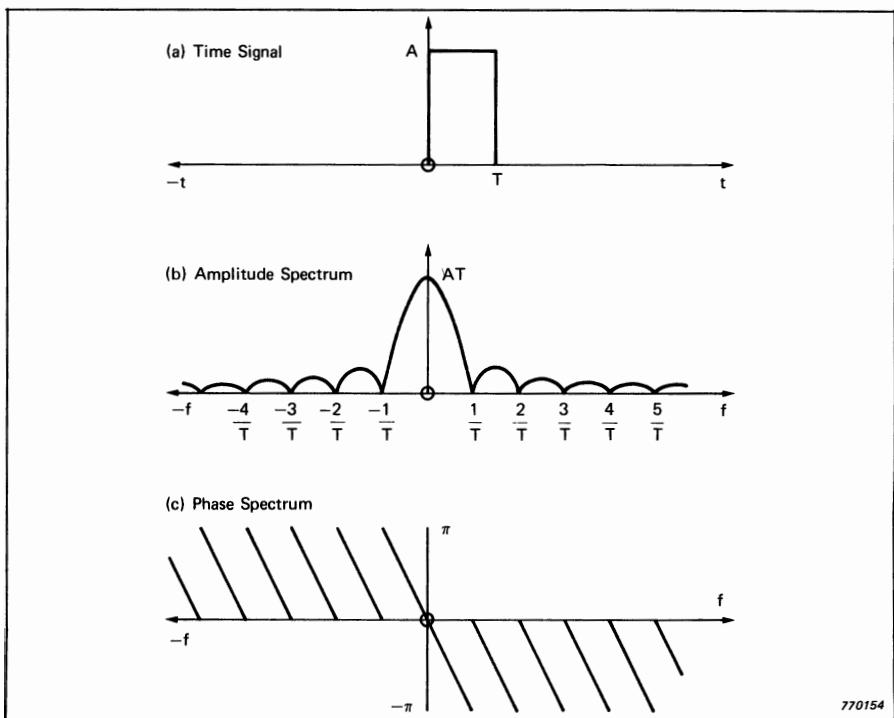


Fig. 2.24. Rectangular pulse starting from zero

Analysis of a transient is often performed by treating it either directly or implicitly as though it were repeated periodically with repetition time T . Because of the artificially introduced periodicity, the measured spectrum becomes a line spectrum with line spacing $1/T$, but the individual lines can be considered as samples of the true continuous spectrum, (see Section 2.5.3) and the line spacing can be made arbitrarily small by increasing T .

Of course the spectrum of the periodic signal will normally be represented on an amplitude rather than a spectral density scale, but the scaling of the results is relatively simple. The measured power of each spectral line is the average power of the periodic signal in a frequency bandwidth corresponding to the line spacing. The energy in the same frequency bandwidth for one repetition of the transient is evidently obtained by multiplying this mean power by the repetition time T . Finally, the energy spectral density at each frequency is obtained by dividing by the bandwidth $1/T$, so that the overall result is a multiplication by T^2 . It should be noted that the result obtained will even so be independent of T (provided the latter is longer than the transient) since for example a doubling of the repetition time will result in a quartering (-6 dB) of the measured power of a spectral line at a given frequency, partly because the same energy is spread over twice the time, and partly because one spectral line is replaced by two. A doubling of the repetition time will however permit measurement with half the bandwidth, and thus allow a more detailed resolution of the spectrum. This and other practical considerations are discussed in more detail in Chapter 5.

2.4.5. Non-Stationary Signals

Although the term “non-stationary” covers all signals which do not satisfy the requirements for stationary signals, the majority of useful non-stationary signals are such that they can be divided up into short quasi-stationary sections. For example, a continuous train of speech can be divided up into short individual sounds; vowels, consonants etc.

The process of dividing up such a continuous signal into short sections is called “time windowing” because the total signal can be considered to be viewed through a window which only transmits the portion of interest. The simplest way of applying such a window is to cut off the signal at each end. It will be seen that this can be considered as a multiplication by a rectangular weighting function of length T , which gives uniform weighting along the selected sample. However, the effect of this on a typical frequency spectrum component was discussed in Section 2.3 where it was shown that the original spectrum has effectively been filtered by a filter characteristic corresponding to the Fourier transform of the rectangular weighting function. In Section 4.3.2 it is shown that smooth non-uniform weighting functions, such as the Hanning weighting function, have a more desirable filtering effect and are often prefera-

ble, and details are given as to their selection. Fig. 2.25 shows how such a "window" can be moved along a speech record to permit successive analysis of the various components of a particular utterance. After multiplication of the original signal by such a weighting curve, the result is a transient which may be analyzed in the same way as transients in general. The question of scaling is, however, the reverse of that discussed in Section 2.4.4, since results will normally be interpreted in terms of the power of the equivalent stationary signal represented by the windowed segment. Details of this are given in the appropriate sections on practical analysis.

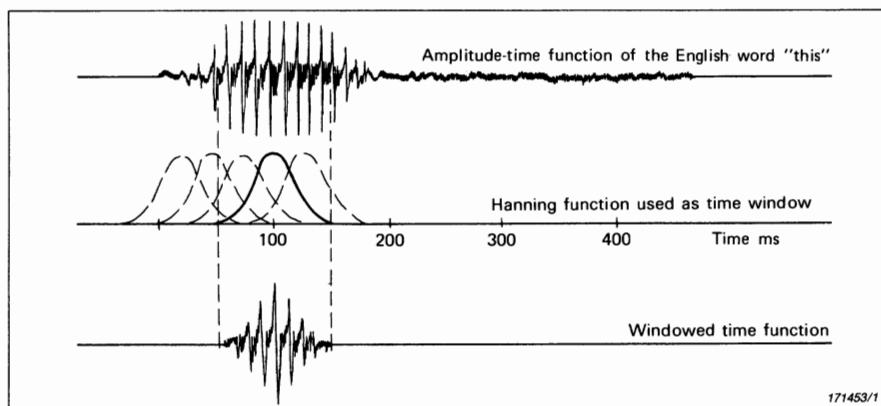


Fig. 2.25. Use of a time window in speech analysis

2.5. CONVOLUTION

It was mentioned in the Introduction that one of the most important properties of the Fourier Transform is that it transforms a convolution into a multiplication. It is the intention here to examine this statement in more detail, and at the same time to point out the advantages which accrue from it, as well as using the theorem to give a better theoretical background to some of the statements made in earlier sections.

First, it is necessary to define convolution and also to give examples of its application so that it acquires a physical meaning.

The convolution of two time functions $f(t)$ and $h(t)$ is defined mathematically as:

$$g(t) = \int_{-\infty}^{\infty} f(\tau) h(t-\tau) d\tau \quad (2.28)$$

For convenience, this is often represented symbolically as:

$$g(t) = f(t) * h(t) \quad (2.28(a))$$

where the star means “convolved with”.

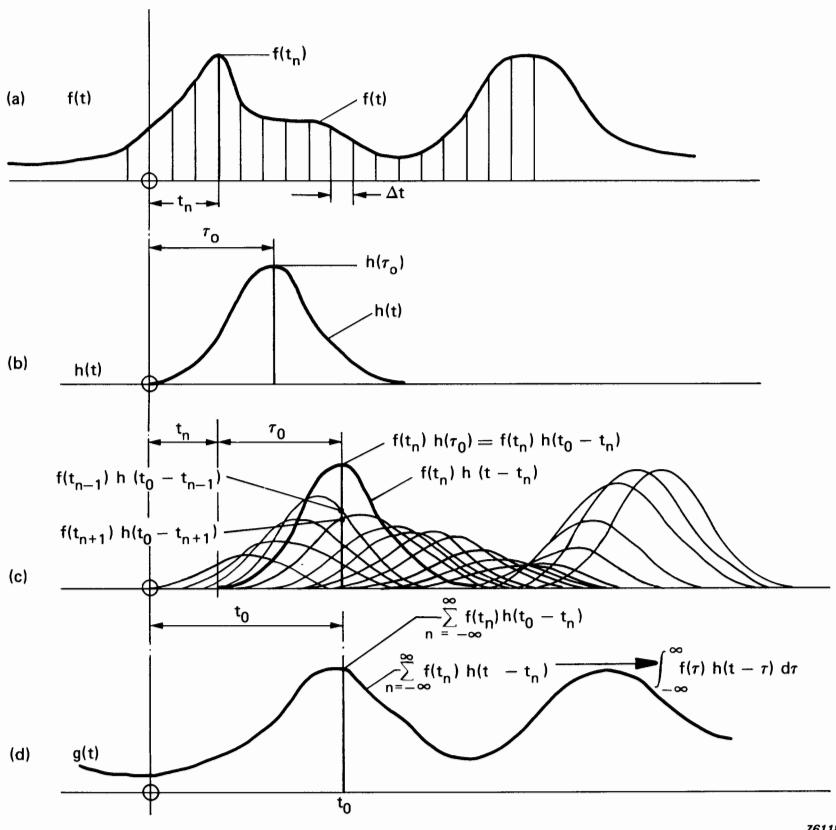
One major application of this relationship is to the case where $f(t)$ represents an input signal to a physical system and $h(t)$ the impulse response of the system. $g(t)$ will then be the output of the system. A discussion of Fig. 2.26 should help to illustrate why this is so. Fig. 2.26(a) represents the time signal $f(t)$ and Fig. 2.26(b) the impulse response $h(t)$ of a physical system to which it is applied. The assumption is made that each point in $f(t)$ can be considered as an impulse (delta function) weighted by the value of $f(t)$ at that point. Each such impulse excites an impulse response, the scaling of which is proportional to the level of $f(t)$ and whose time origin coincides with the impulse. The output signal at time t , $g(t)$, consists of the sum of these scaled impulse responses each delayed by the appropriate time interval from the time of excitation up to the time of measurement. Because each point of the response curve consists of a sum of components which have been excited at different times, it is necessary to integrate over a dummy time variable τ . It is simplest at first to consider $f(\tau)$ as a series of impulses at discrete times t_n (with time increment Δt) and then let Δt tend to zero in a final limiting process. Fig. 2.26(a) shows a typical impulse $f(t_n)$, and in Fig. 2.26(c) the impulse response from this alone is shown as a heavy line. It will be seen that this response has t_n as its origin, and for example the peak value of the response occurs at time $t_o = t_n + \tau_o$ where τ_o is depicted in Fig. 2.26(b). The value of this peak is evidently $f(t_n) \cdot h(\tau_o)$, i.e. $f(t_n) \cdot h(t_o - t_n)$, and thus at any other time t the value of the response (from $f(t_n)$ alone) is equal to $f(t_n) \cdot h(t - t_n)$ as illustrated. The response at time t_o from the signal applied at time t_n might be termed $g_n(t_o)$ and thus the total response at time t_o is equal to the sum of all the responses excited at the various times.

$$\text{i.e. } g(t_o) = \sum_{n=-\infty}^{\infty} g_n(t_o) = \sum_{n=-\infty}^{\infty} f(t_n) h(t_o - t_n)$$

More generally, at time t the response

$$g(t) = \sum_{n=-\infty}^{\infty} f(t_n) h(t - t_n) \quad (2.29)$$

It will be seen that in the limit as $\Delta t \rightarrow 0$ Equation (2.29) tends to Equation (2.28) as shown in Fig. 2.26(d).



761181

Fig. 2.26. Convolution of two time functions

2.5.1. Convolution with a Delta Function

A special case which arises in various situations involves convolution with a delta function, which in fact is found to be relatively simple. Fig. 2.27, which is comparable with Fig. 2.26, illustrates a typical situation where $h(t)$ is a unit delta function $\delta(t - \tau_0)$ with a delay time of τ_0 (Fig. 2.27(b)). In Fig. 2.27(c) it is shown how each discrete impulse in the original time function $f(t_n)$ now generates a single impulse in the response, delayed by τ_0 . The overall effect is to delay the whole signal by τ_0 , but otherwise to leave it unchanged. If the delta function is weighted with a scaling factor then the entire response is also weighted by that scaling factor.

In general it can be said that the effect of convolving a function with a delta function is to shift its origin to the delta function.

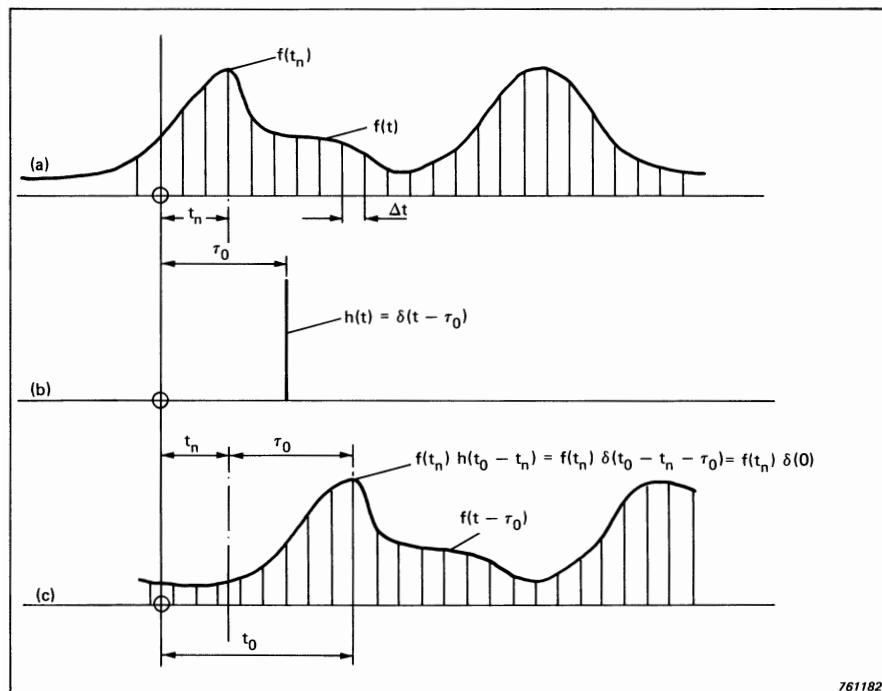


Fig. 2.27. Convolution with a delta function

761182

2.5.2. Convolution of Spectral Functions

So far only the convolution of two real-valued functions has been discussed, but it is possible for two complex functions to be convolved, for example two frequency spectra.

The convolution equation is still the same, viz.

$$F(f) * H(f) = \int_{-\infty}^{\infty} F(\phi) H(f - \phi) d\phi \quad (2.30)$$

but since $F(f)$ and $H(f)$ are complex variables, the multiplication is now a complex multiplication (i.e. multiplication of amplitudes and addition of phases) and the integration represents a complex or vector addition.

2.5.3. Convolution Theorem

The Convolution Theorem states that the Fourier Transform (either forward or inverse) transforms a convolution into a multiplication and vice versa. For example, for the case represented by Equation (2.28)

$$\begin{aligned} \text{if } G(f) &= \mathcal{F}\{g(t)\} && \text{i.e. the forward Fourier transform of } g(t) \\ F(f) &= \mathcal{F}\{f(t)\} \\ \text{and } H(f) &= \mathcal{F}\{h(t)\} && (2.31) \\ \text{and if } g(t) &= f(t) * h(t) \\ \text{then } G(f) &= F(f) \cdot H(f) \end{aligned}$$

A proof of this is given in Appendix A for a forward Fourier transform, but it can be appreciated that because of the symmetry of Equations (2.14) and (2.15) the same will apply to the inverse transform.

The benefits of this are immediately apparent when interpreted in terms of the excitation and response of a physical system. $H(f)$, the forward Fourier transform of the impulse response is known as the frequency response function. The spectrum of the output is obtained very simply by multiplying the input spectrum by the frequency response function at each frequency. The equivalent convolution in the time domain is evidently a much more complicated procedure.

Another example of a convolution transforming to a product is represented by the case referred to in Section 2.4.4., viz. that periodic repetition of a transient results in a sampling of its spectrum in the frequency domain. This is illustrated in Fig. 2.28. Fig. 2.28(a) shows the original transient and its spectrum. Fig. 2.28(b) shows a train of unit impulses with a spacing of T , and its spectrum which is another train of impulses with spacing $1/T$ (Ref. 2.1, 2.3). Fig. 2.28(c) shows the periodically repeated transient which can be considered as the convolution of the original transient with the impulse train (see Section 2.5.1). The Convolution Theorem indicates that the result of this convolution will be a multiplication of the respective spectra, thus giving a train of delta functions at frequency intervals of $1/T$ and weighted by the original spectrum level at the corresponding frequency. This corresponds to a sampling of the spectrum at intervals of $1/T$ as assumed in Section 2.4.4.

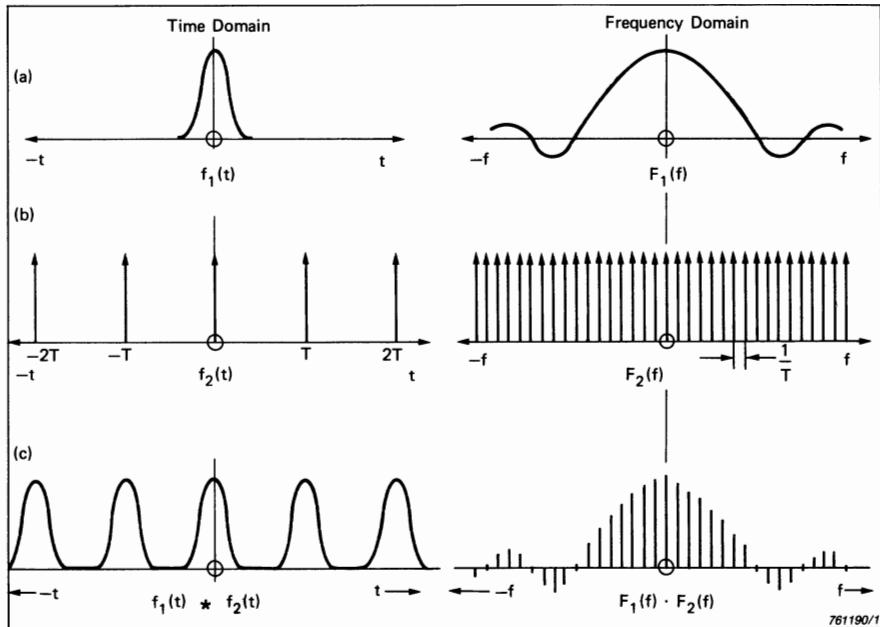


Fig. 2.28. Periodic repetition of an impulse

It is perhaps also useful to give an example of the other version of the Convolution Theorem viz. that a multiplication in the time domain transforms to a convolution in the frequency domain.

Fig. 2.29 illustrates the case discussed in Section 2.3, viz. the effect of a truncation to a record length of T . Fig. 2.29(a) represents an endless cosine function in both time and frequency domains. Note that the arrows at $\pm f_0$ represent delta functions which should of course be infinitely high. For comparison purposes Fig. 2.29(b) shows a sine function, whose spectrum only differs by virtue of the initial phase angles of the positive and negative frequency components. Fig. 2.29(c) shows a rectangular "time window" of length T , evenly divided about zero time, and its frequency spectrum. As shown in Appendix A the latter is a $(\sin x/x)$ function with zeroes at multiples of $1/T$. The third column of Fig. 2.29 shows the power spectrum which is to be discussed later.

Restricting the length of either the cosine or sine function to T is the same as multiplying it by the rectangular time window and this is illustrated in Fig. 2.29(d) and (e). This corresponds in the frequency domain to a convolution of the respective frequency spectra. This can be done separately for the positive and negative frequency delta functions and consists, as shown previously, in replacing each of them by the convolving function $F_3(f)$. The result is then obtained as the sum of the positive and negative frequency contributions. It will be appreciated that it is primarily in the vicinity of zero frequency (where the two $(\sin x/x)$ functions have the same magnitude) that there is a significant interaction. In the vicinity of either f_o or $-f_o$ the result is virtually a $(\sin x/x)$ filter characteristic centred on $\pm f_o$, as derived heuristically in Section 2.3. In the vicinity of zero frequency it is perhaps of interest to note that because the two $(\sin x/x)$ functions have the same phase in Fig. 2.29(d), they reinforce and give a DC component, while in Fig. 2.29(e) they are opposed and give zero DC component. This can also be seen from the respective time functions, where in Fig. 2.29(d) it has 5 positive lobes against 4 negative, while that in Fig. 2.29(e) has $4\frac{1}{2}$ positive and negative lobes and is thus balanced.

It is interesting to investigate what effect the convolution of the complex spectra has on the power spectra, since, as mentioned previously, one is often most interested in the power spectrum. The third column of Fig. 2.29 shows the power spectra corresponding to the various complex spectra, and this reveals that although the power spectra for the sine and cosine are the same (Fig. 2.29(a) and (b)) the power spectra for the convolved spectra are different (Fig. 2.29(d) and (e)). Thus, the power spectrum of the convolved function is not equal to the convolution of the power spectra. However, Fig. 2.30 compares the latter with the two other alternatives, and it is found that it lies between them. The convolution of the two power spectra has a value at zero frequency which is double the PSD of the contribution of the positive frequency component alone (i.e. +3 dB). The power spectrum of the cosine case (Fig. 2.29(a)) is four times larger at zero frequency (i.e. +6 dB) since the linear amplitudes add rather than the squared amplitudes. The convolution of the power spectra in fact represents the "average" case, where the phase angle between the positive and negative frequency components of the sinusoid is 90° , and this is incidentally the value which will be approached if the power spectrum is obtained by averaging over several records of length T taken at random along the sinusoid.

Thus in this sense it can be said that the power spectrum of the convolution is equal to the convolution of the power spectra.

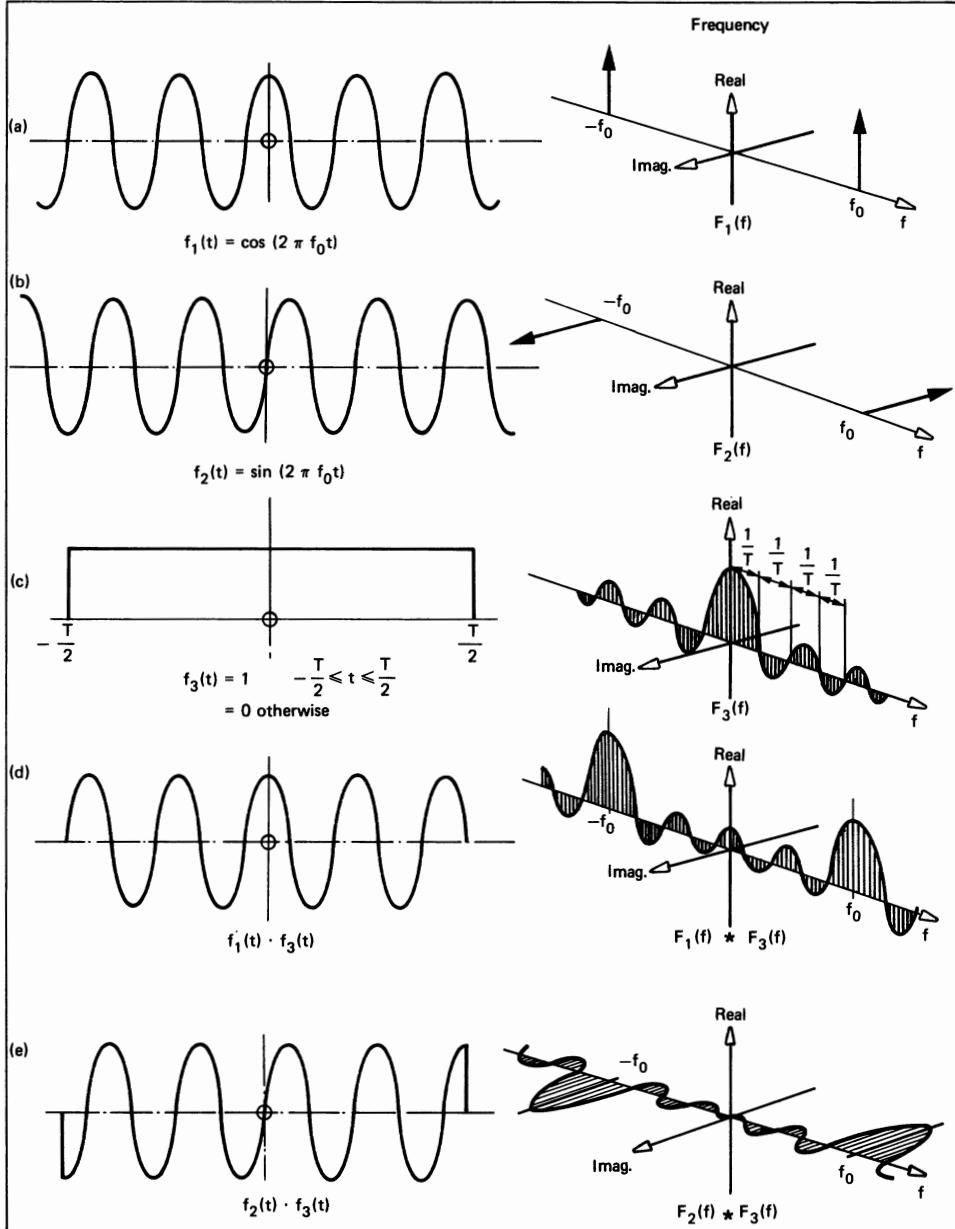
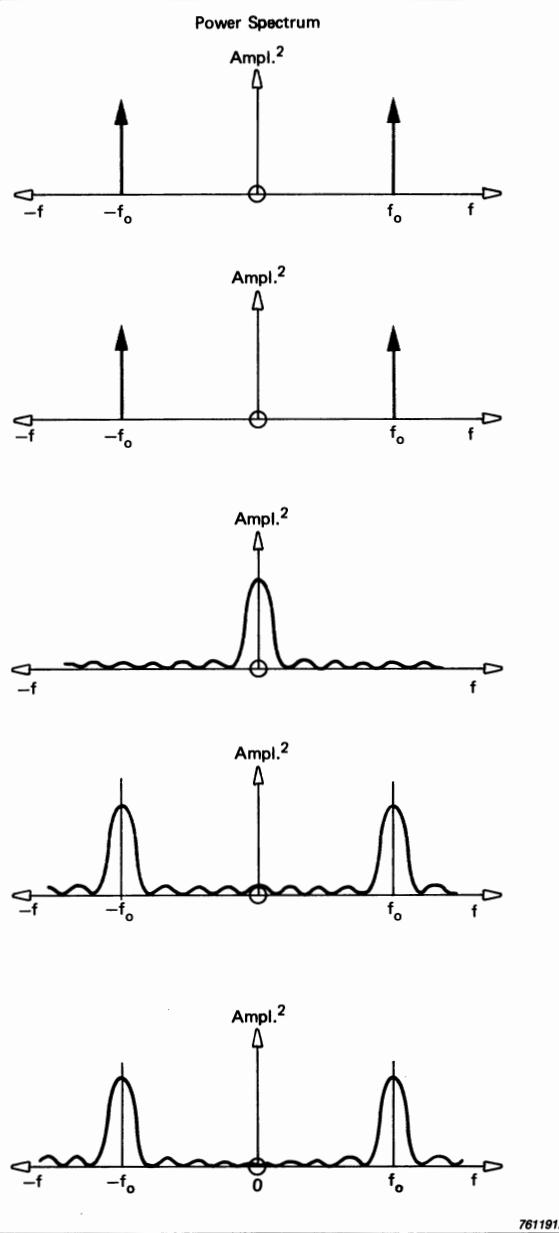


Fig. 2.29. Restriction of sinusoidal signals to a length of T



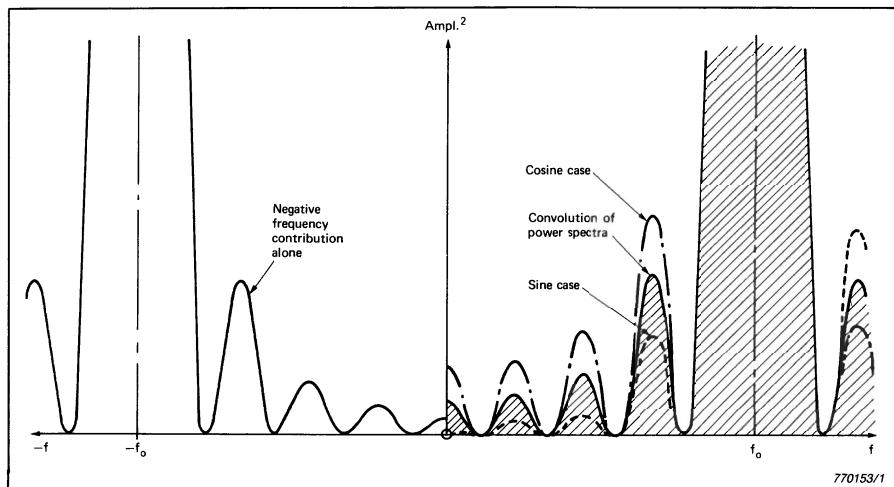


Fig. 2.30. Detail of power spectra in the vicinity of zero frequency

2.6. THE HILBERT TRANSFORM

Another relationship of which considerable use will be made in this book is the Hilbert transform. The Hilbert Transform expresses the relationship between the real and imaginary components of the Fourier transform of a one-sided signal.

That there should be a fixed relationship will become evident from the following discussion, which uses the example of a causal signal to illustrate the general case. A causal time signal is one which is equal to zero for negative time, as illustrated in Fig.2.31(a).

$$\text{i.e.} \quad a(t) = 0, \quad t < 0 \quad (2.32)$$

A typical example would be the impulse response of a causal system, meaning that there can be no output before the input is applied at time zero.

As illustrated in Fig.2.31, a causal signal can be divided into even and odd components which are not independent of each other. In order for the negative time components to cancel each other, the positive time components must be identical. Another way of expressing this relationship makes use of the fact that the even and odd components are related by the "sign" function, i.e.

$$a(t) = a_e(t) + a_o(t) \quad (2.33)$$

$$a_e(t) = a_o(t) \cdot sgn(t) \quad (2.34)$$

and

$$a_o(t) = a_e(t) \cdot \text{sgn}(t) \quad (2.35)$$

To relate these even and odd components of the time signal to the real and imaginary components of the Fourier transform it is necessary to make use of some relationships which have not previously been stated explicitly.

Recalling the Fourier transform equations:

$$G(f) = \int_{-\infty}^{\infty} g(t) e^{-j2\pi ft} dt \quad (2.14)$$

and

$$g(t) = \int_{-\infty}^{\infty} G(f) e^{j2\pi ft} df \quad (2.15)$$

it will be seen that the only difference between the forward and inverse transforms is the sign of the exponent. This can be compensated for by reversing the sign of the independent variable. Thus, for example, the effect of a forward Fourier transform on a time signal $g(t)$ is the same as an inverse transform on the same time signal reversed end for end, $g(-t)$.

More generally,

$$g(t) \xrightarrow{\mathcal{F}} G(f) \xrightarrow{\mathcal{F}} g(-t) \xrightarrow{\mathcal{F}} G(-f) \xrightarrow{\mathcal{F}} g(t) \quad (2.36)$$

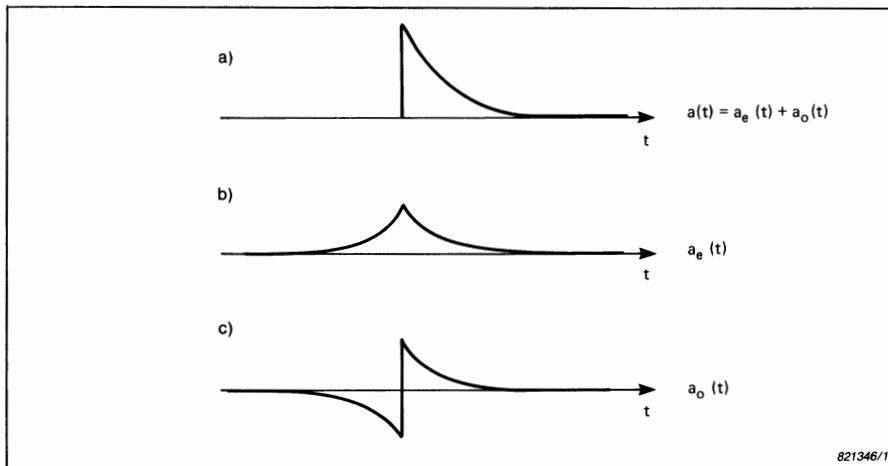


Fig. 2.31. Division of a causal signal into even and odd components

We also need the more general equivalent of Eqn.(2.12), which for a real valued time signal $g(t)$, states that

$$G(f) = G^*(-f) \quad (2.37)$$

i.e. $R(f) = R(-f)$ (2.38)

and $I(f) = -I(-f)$ (2.39)

where $G(f) = R(f) + jI(f)$ in terms of its real and imaginary components. The spectrum is thus conjugate even (real parts even, imaginary parts odd).

From Eqn.(2.36) it follows that for an even function, where $g(t) = g(-t)$, then $G(f) = G(-f)$. Hence, making use of Eqn.(2.37)

$$R(f) + jI(f) = R(f) - jI(f)$$

from which $I(f) = 0$, and the spectrum is real.

Conversely, for an odd function, where $g(t) = -g(-t)$ and $G(f) = -G(-f)$, then $R(f) + jI(f) = -R(f) + jI(f)$ from which $R(f) = 0$, and the spectrum is imaginary. These results, plus similarly derived relationships are summarised in Table 2.1.

Time Signal	Spectrum
real and even	real and even
real and odd	imag and odd
imag and even	imag and even
imag and odd	real and odd
real	conjugate even
conjugate even	real

T01045GB0

Table 2.1. Fourier Transform relationships

Coming back to Eqns.(2.33) and (2.34) and making use of Table 2.1. we find that if

$$\mathcal{F}\{a(t)\} = A(f) = A_R(f) + jA_I(f) \quad (2.40)$$

then $A_R(f) = \mathcal{F}\{a_e(t)\}$ (2.41)

and $jA_I(f) = \mathcal{F}\{a_o(t)\}$ (2.42)

Moreover, $A_R(f) = \mathcal{F}\{a_o(t) \cdot sgn(t)\}$ (2.43)

from which it can be seen that $A_R(f)$ and $A_I(f)$, the real and imaginary components of the Fourier transform, $A(f)$, of the causal time signal, $a(t)$, are not independent since both can be calculated from the odd part $a_o(t)$. The relationship between $A_R(f)$ and $A_I(f)$ is known as the Hilbert transform (actually $A_R(f)$ is the Hilbert transform of $A_I(f)$) and can be derived from Eqn.(2.43).

Thus, using the Convolution Theorem (Eqn.2.31)

$$\begin{aligned} A_R(f) &= \mathcal{F} \{ a_o(t) \cdot sgn(t) \} \\ &= \mathcal{F} \{ a_o(t) \} * \mathcal{F} \{ sgn(t) \} \\ &= A_I(f) * \frac{1}{\pi f} \end{aligned} \quad (2.44)$$

making use of the fact, that (Ref.2.4)

$$\mathcal{F} \{ sgn(t) \} = \frac{1}{j\pi f} \quad (2.45)$$

Equation (2.44) is one way of writing the Hilbert transform. Writing it out in full, it can be expressed as

$$\mathcal{H}\{G(f)\} = \tilde{G}(f) = \frac{1}{\pi} \int_{-\infty}^{\infty} G(\phi) \frac{1}{f-\phi} d\phi \quad (2.46)$$

Note that, contrary to the Fourier transform, the Hilbert transform does not change the independent variable, and the result is in the same domain as the original function.

The Hilbert transform of a time function is thus defined in exactly the same way:

$$\mathcal{H}\{a(t)\} = \tilde{a}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} a(\tau) \left(\frac{1}{t-\tau} \right) d\tau \quad (2.47)$$

$$= \frac{1}{\pi} a(t) * \left(\frac{1}{t} \right) \quad (2.48)$$

Applying the convolution theorem to Eqn.(2.48) it can be shown that the Hilbert transform corresponds to a 90° phase shift.

Thus, $\mathcal{F} \{ \tilde{a}(t) \} = A_{\perp}(f) = A(f) \cdot \{-j sgn(f)\}$ (2.49)

Hence, the spectrum of the Hilbert transform can be obtained by multiplying positive frequency components by $-j$ (a phase shift of -90°) and negative frequency components by $+j$ ($+90^\circ$).

This procedure, illustrated in Fig.2.32, is evidently much simpler than evaluating Eqn.(2.47) directly.

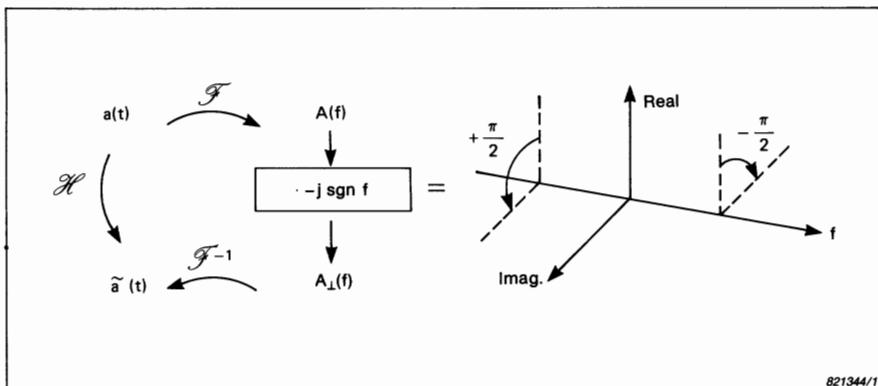


Fig. 2.32. The effect in the frequency domain of a Hilbert transform in the time domain

Fig.2.33 illustrates the results of two Hilbert transforms on a cosine function (as seen in the frequency domain). The first gives a sine function, and the second a minus cosine. This will be recognized as very similar to an integration, but only applies to spectra with a single frequency component ω_o as will be seen from Fig.2.34. Integration in the time domain corresponds to a division by $j\omega_o$ (multiplication by $-j\frac{1}{\omega_o}$) in the frequency domain which except for the scaling by $|\frac{1}{\omega_o}|$ is the same as a Hilbert transform.

2.6.1. Analytic Signals

An analytic signal is a complex time signal whose imaginary part is the Hilbert transform of the real part.

Thus, if $\tilde{a}(t) = \mathcal{H}\{a(t)\}$, then

$$\tilde{a}(t) = a(t) + j\tilde{a}(t) \quad (2.50)$$

is an analytic time signal.

To see what the application of analytic signals might be, it is interesting to investigate another means of representing frequency components than that used in Section 2.2. Instead of representing a general component as a sum of positive and negative frequency phasors (whose sum is always real) it can be represented as the projection on the real axis of a single phasor (of double

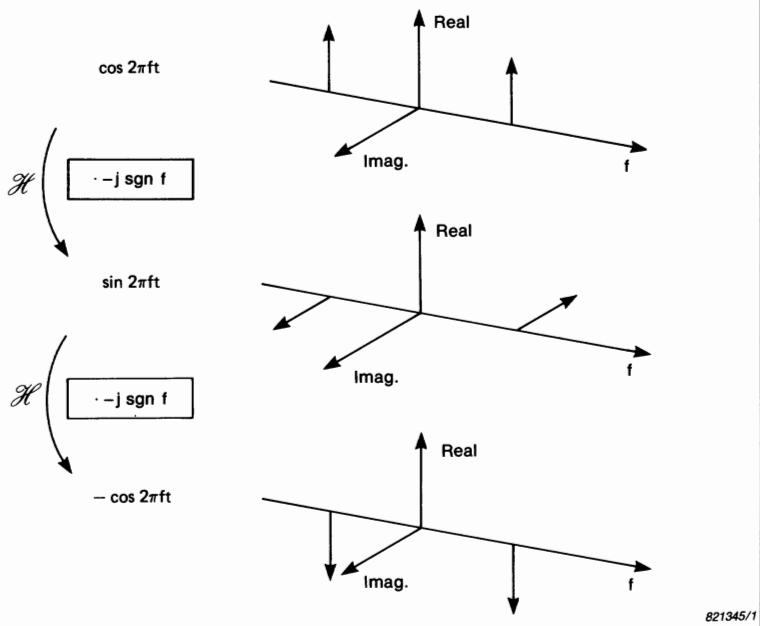


Fig. 2.33. Hilbert transforms of a sinusoid as seen in the frequency domain

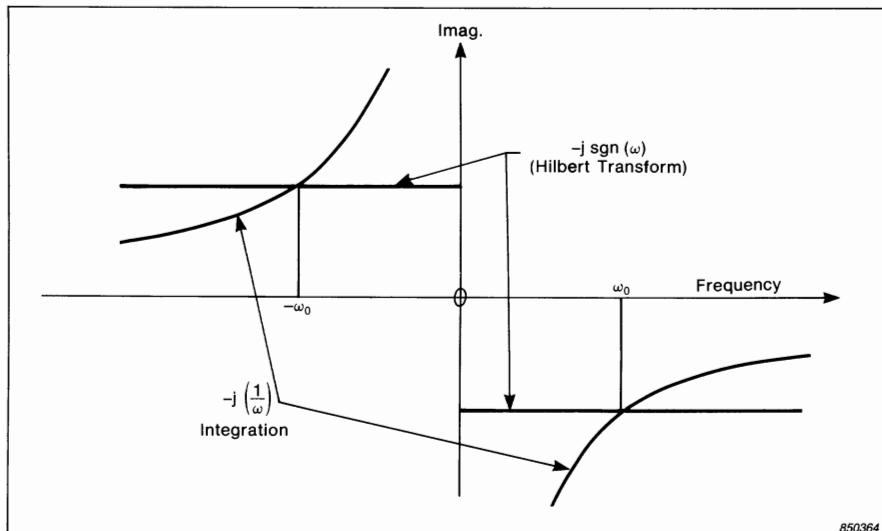


Fig. 2.34. Equivalence of frequency weighting functions for a single frequency component

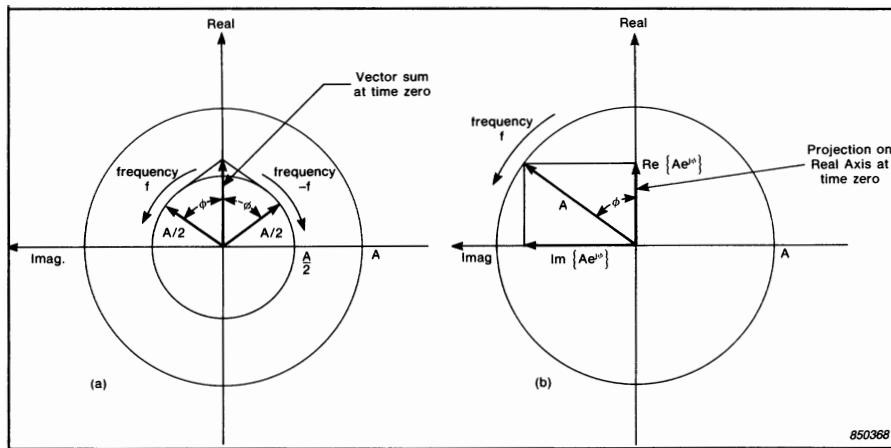


Fig. 2.35. Equivalence of vector sum and projection on Real Axis

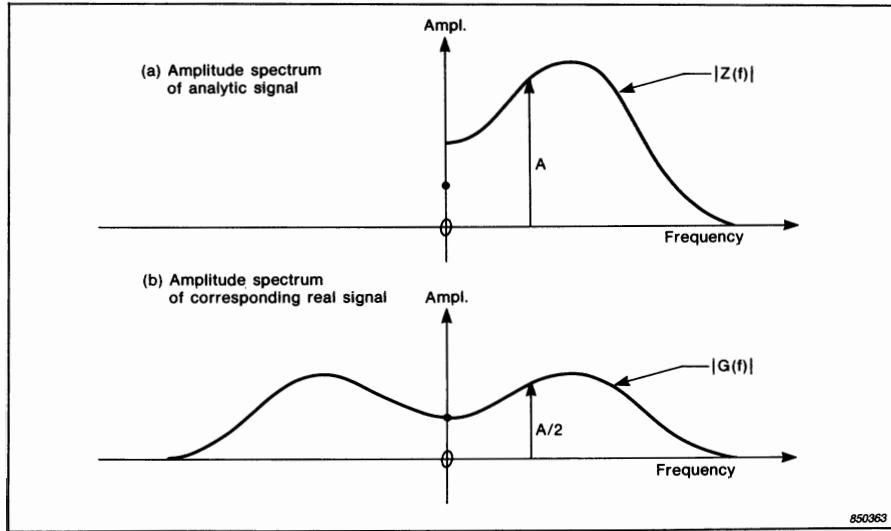


Fig. 2.36. One-sided vs. two-sided spectra

amplitude). Without loss of generality the frequency of this phasor can be taken as positive.

Fig. 2.35, which can be compared with Fig. 2.2, illustrates the equivalence of these two approaches. Thus, a time signal consisting of many frequency components could be expressed as the sum of complex conjugate pairs at plus and minus each frequency, or as the sum of the real parts of positive frequency

phasors alone, each with double amplitude. Fig.2.36 illustrates the two amplitude spectra corresponding to a typical situation, where the similarity to Fig.2.31 will be obvious. The spectrum for the model based on projections has twice the amplitude at all positive frequencies, the same amplitude at zero frequency, and zero amplitude for negative frequencies. Note that the phase spectra for the two models are identical (at least at positive frequencies) as will be obvious from Fig.2.35.

The time signal can be synthesized from either model by integrating over all frequencies. For the 2-sided spectrum of Fig.2.36(b), this is simply the inverse Fourier transform:

$$g(t) = \int_{-\infty}^{\infty} G(f) e^{j2\pi ft} df \quad (2.15)$$

For the one-sided spectrum of Fig.2.36(a) it can be expressed as

$$\begin{aligned} g(t) &= \int_{-\infty}^{\infty} \operatorname{Re} \{ Z(f) e^{j2\pi ft} \} df \\ &= \operatorname{Re} \left\{ \int_{-\infty}^{\infty} Z(f) e^{j2\pi ft} df \right\} \end{aligned} \quad (2.51)$$

i.e. the real part of the inverse Fourier transform of $Z(f)$

$$\text{where } Z(f) = 0 \quad , f < 0$$

$$Z(f) = G(f) \quad , f = 0$$

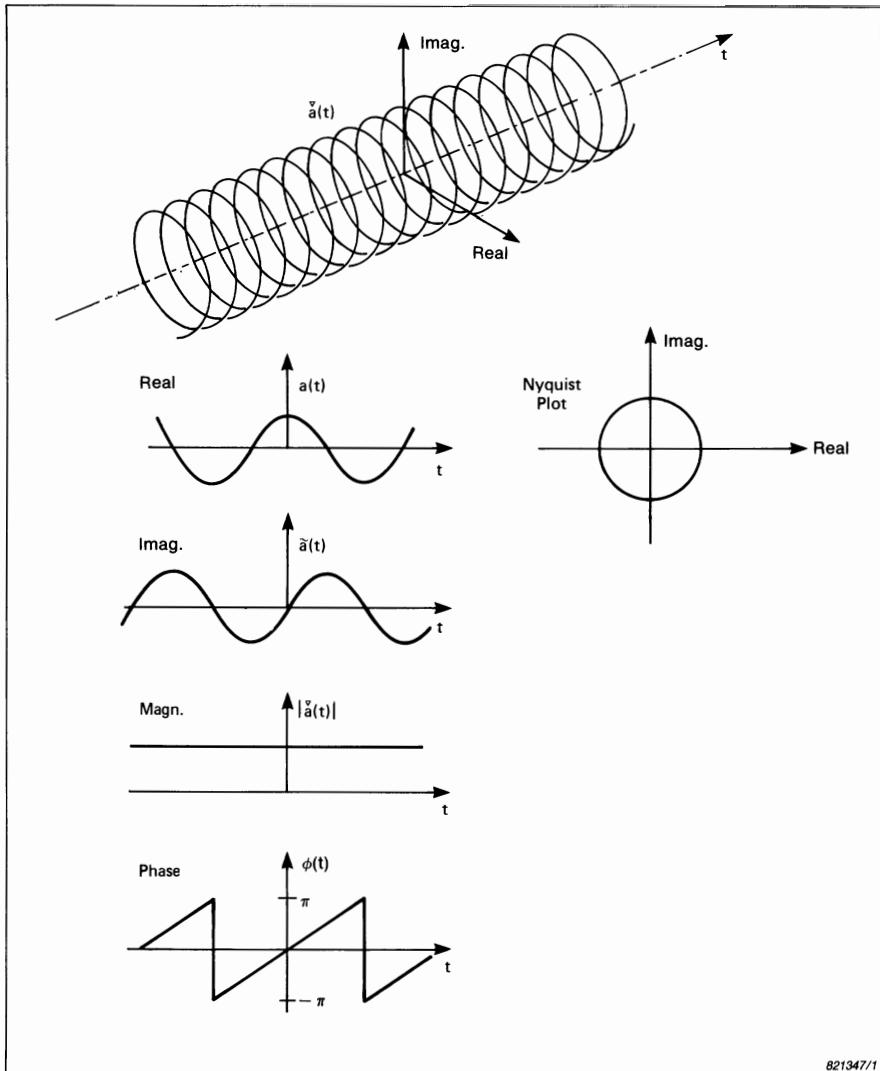
$$Z(f) = 2 G(f) \quad , f > 0$$

By analogy with Fig.2.31 it seems reasonable that the inverse Fourier transform of $Z(f)$, while complex, will have its real and imaginary parts related by a Hilbert transform. There are two differences, however. Firstly, instead of dividing $Z(f)$ into even and odd components, it is divided into conjugate even and conjugate odd components. The former is equal to $G(f)$, and transforms back to the real time function $g(t)$. The latter transforms to an imaginary time function, and still fulfils the requirement that it is equal to $G(f) \cdot \operatorname{sgn}(f)$. The derivation then follows that of Eqns.(2.40) through (2.45) with the additional difference that the use of the inverse rather than the forward Fourier transform results in the imaginary part being the Hilbert transform of the real part, rather than vice versa. The difference between forward and inverse Hilbert transforms can readily be shown in fact to be simply a change of sign.

$$\text{Thus if, } z(t) = \mathcal{F}^{-1} \{ Z(f) \}$$

$$\text{then } z(t) = g(t) + j\bar{g}(t) \quad (2.52)$$

and is thus an analytic signal.



821347/1

Fig. 2.37. The analytic signal corresponding to a cosine

It may be of interest to take a simple example for illustration purposes. Fig.2.37 shows the analytic signal derived from a cosine function. The imaginary part is thus a sine function, and the analytic function a spiral, or helix, the locus of a uniformly rotating phasor with constant (unit) amplitude. The phase angle $\phi(t)$ is a linearly increasing function of time (which is depicted, however, modulo 2π), and the slope of the phase curve represents the (constant) instantaneous frequency. Note that all the display modes normally associated with

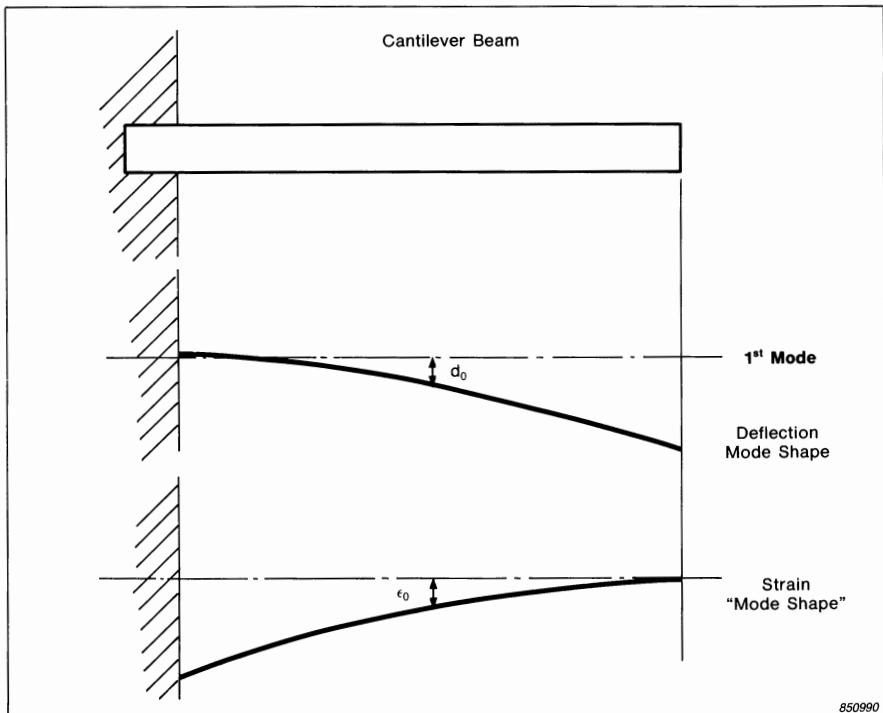


Fig. 2.38. Mode shapes for a cantilever beam in its first bending mode

frequency spectra viz. real, imaginary, magnitude, phase, and even Nyquist (polar) display are available for depicting analytic time signals.

Generation of a complex time signal from a measurable real component may not seem to have any physical relevance, but in fact the real and imaginary components of the analytic function can often be related to the two forms of energy always associated with a vibration or oscillation, viz., potential and kinetic energy. If, for example, the real part (or more strictly speaking its square) represents potential energy, then (the square of) the imaginary part would represent kinetic energy.

Fig.2.38, for example, shows a cantilever beam, and its first bending mode of vibration. Because this occurs at a single frequency the displacement mode shape depicted applies equally to its velocity (or acceleration). Also shown is its strain "mode shape" which is quite different (in fact the second derivative of the displacement mode shape). At the extremity of an oscillation, expressed in terms of the typical displacement coordinate d_0 , the velocity is zero, and all energy is in the form of internal strain energy. If a typical value of strain in this

situation is ϵ_0 , then the total strain energy can be represented as $C_1 \epsilon_0^2$ where C_1 is a constant with the appropriate physical dimensions. In the undeformed position, all the strain energy has been converted to kinetic energy which in terms of a typical velocity coordinate v_0 can be expressed as $C_2 v_0^2$, where C_2 is another constant. Since the total energy is the same the constants C_1 and C_2 must be scaled such that

$$C_1 \epsilon_0^2 = C_2 v_0^2 = C \quad (2.53)$$

The values ϵ_0 and v_0 represent the extremities of a sinusoidal oscillation, and at any other time in the cycle the total energy is also constant, i.e.

$$C_1 \epsilon^2(t) + C_2 v^2(t) = C \quad (2.54)$$

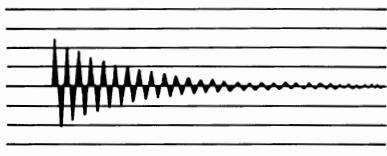
It is under these conditions that $\epsilon(t)$ and $v(t)$ must be Hilbert transforms of each other, except for a scaling constant defined by Eqn.(2.53). In this simple case of a constant amplitude oscillation, if $\epsilon(t) = \epsilon_0 \cos 2\pi f_0 t$ then $v(t) = v_0 \sin 2\pi f_0 t$. The total energy at any time is given by

$$\begin{aligned} & C_1(\epsilon_0 \cos 2\pi f_0 t)^2 + C_2(v_0 \sin 2\pi f_0 t)^2 \\ &= C[\cos^2(2\pi f_0 t) + \sin^2(2\pi f_0 t)] \\ &= C \quad \text{Q.E.D.} \end{aligned}$$

Note that the coordinates ϵ_0 and v_0 are typical coordinates representing the peak deflection in a particular mode shape and can be considered to be a scaled version of the peak value almost anywhere in the system. Nodal points in either mode shape are here excluded, however, because it is then impossible to satisfy Eqn.(2.53). In the particular case of a cantilever beam, as in Fig.2.38, the coordinates at either end of the beam cannot be used, since the free end is a node for the strain mode shape, and the built-in end is a node for the velocity mode shape.

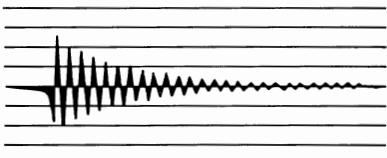
Consideration of Fig.2.36 and the frequency functions $G(f)$ and $Z(f)$ confirms these energy considerations in the frequency domain as well. Neglecting the DC component (which does not enter into the oscillation), each (positive) frequency component in $Z(f)$ has twice the amplitude and thus four times the energy of the equivalent component in $G(f)$. On the other hand, the negative frequency side of $G(f)$ represents the same amount again, so that overall $G(f)$ (and thus $g(t)$) has half the total energy contained in $Z(f)$, which represents the sum of the kinetic and potential energies.

Even where the total energy is not constant, such as in a damped free oscillation, the analytic signal can sometimes be used to represent the potential and kinetic energy components in the model.

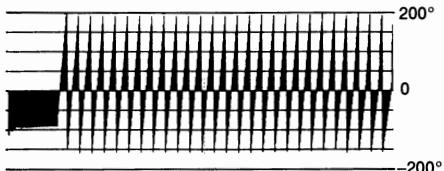


Impulse Response
of a Single Degree
of Freedom System

a) Real part $a(t)$

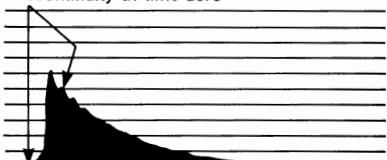


b) Imag. part $b(t) = \mathcal{H}\{a(t)\}$



d) Phase $\phi(t) = \tan^{-1} \left\{ \frac{b(t)}{a(t)} \right\}$

Deviations caused by
discontinuity at time zero



c) Amplitude $A(t) = \sqrt{a^2(t) + b^2(t)}$



e) Log amplitude, $\log A(t)$

850370

Fig. 2.39. Impulse response of a single degree of freedom system

Fig. 2.39 shows the impulse response for a single degree of freedom system, a damped one-sided sinewave, and its Hilbert transform. After time zero, the Hilbert transform pair represent damped sinusoids, 90° out-of-phase, but with the same “envelope” function, an exponential decay. The square of this amplitude function represents the total energy at any time. The phase curve once again has the constant slope corresponding to constant oscillation frequency, and is seen to be independent of the amplitude function. On the other hand, the abrupt discontinuity at time zero, where the impulse is applied, is seen to give problems in that the Hilbert transform exists for negative time and is thus non-causal. Clearly, the division into kinetic and potential energy does not apply in this region, as all energy is supplied instantaneously at time zero, and thus the

model breaks down. It appears that the model can only be applied for moderate rates of addition or subtraction of energy to or from the system.

On the other hand, the use of the Hilbert transform to generate the “envelope” and instantaneous phase of an oscillating function leads to another application of analytic signals, viz. the representation of modulated signals.

2.6.2. Modulation

So far, time signals have been considered as a sum of phasors with fixed amplitude and rotational velocity (i.e. frequency). It is in fact the purpose of Fourier analysis to derive these fixed phasors.

In the case of modulated signals, it is often easiest initially to model the signal in terms of a single phasor with varying amplitude and rotational speed (which can be termed “instantaneous frequency”). It would be possible to have a model with a complex conjugate pair of such phasors, but it is perhaps easier to deal with the projection on the real axis of a single phasor.

A generally modulated signal can thus be expressed as

$$g(t) = \operatorname{Re} \left\{ A(t) e^{j\phi(t)} \right\} \quad (2.55)$$

where the amplitude $A(t)$ and instantaneous phase angle $\phi(t)$ are functions of time. The instantaneous frequency $f(t) = d\phi(t)/dt$ in rads/s or $\frac{1}{2\pi} d\phi(t)/dt$ in Hz. It will be appreciated that if $A(t) e^{j\phi(t)}$ contains only positive frequencies then it will be an analytic signal.

As an example, take a purely amplitude modulated signal. In this case

$$g_a(t) = \operatorname{Re} \left\{ A(t) e^{j2\pi f_0 t} \right\} \quad (2.56)$$

where f_0 is the (constant) carrier frequency and $A(t)$ describes the (real) amplitude modulation function or “envelope” of the signal (normally including a DC offset).

Figure 2.40 uses the convolution theorem to derive the spectrum of a carrier frequency component, $g_c(t)$, amplitude modulated by a lower frequency sinusoid (plus DC component), $g_m(t)$. The resulting spectrum comprises a pair of sidebands spaced around the carrier frequency, f_c , by an amount equal to the modulation frequency, f_m . In the figure, both sinusoids were taken as cosines for illustration purposes, but in the general case it will be seen from Fig.2.41 that the phase relationships of the upper and lower sidebands must still be such that their vector sum is always aligned with the carrier frequency component. Fig.2.42 depicts the analytic signal corresponding to the same amplitude modulation as Fig.2.40. This can now be considered as a single rotating phasor with

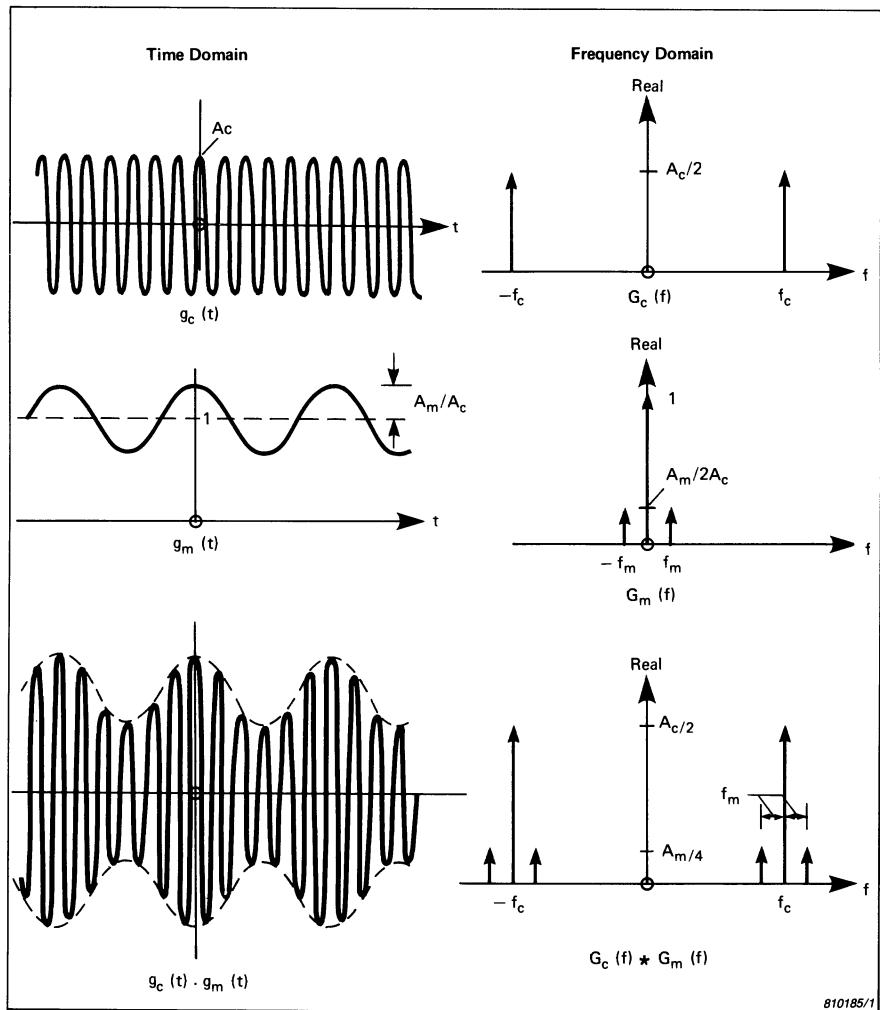


Fig. 2.40. Derivation of the spectrum of an amplitude modulated signal

amplitude fluctuating sinusoidally between two limits. The purpose of Fourier analysis, as typified by Fig.2.40, has been to decompose this single varying phasor into a sum of three constant amplitude phasors as illustrated in Fig.2.41.

For phase or frequency modulation, the decomposition is not quite so simple, but the representation as an analytic signal can still be made. Fig.2.43 depicts such an analytic signal, representing single frequency modulation of a constant amplitude carrier.

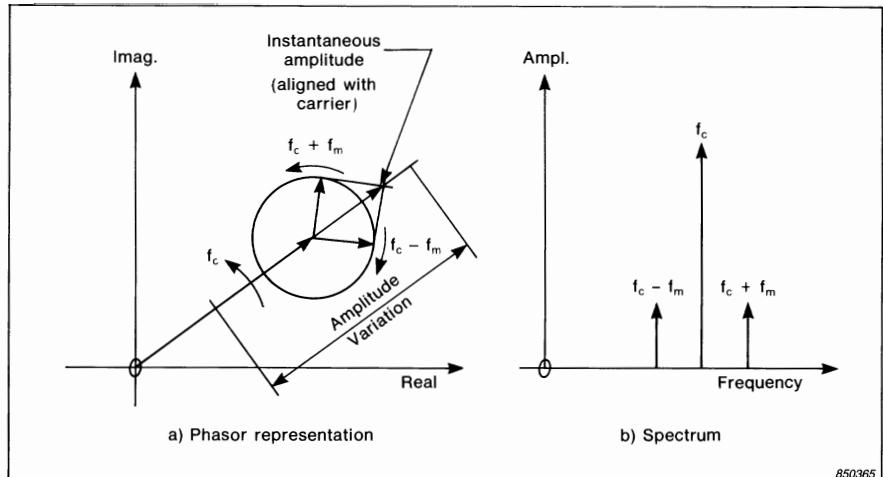


Fig. 2.41. Amplitude modulation as a sum of phasors

850365

The mathematical expression for the analytic signal is

$$g_p(t) = \operatorname{Re} \left\{ A_0 e^{j(2\pi f_0 t + \beta \sin 2\pi f_m t)} \right\} \quad (2.57)$$

where β is the maximum phase deviation, in radians, from the linearly increasing phase of the carrier component with frequency f_0 . This case of a phase modulation by a single frequency can also be interpreted as a frequency modulation by a single frequency. The instantaneous phase of the analytic signal is given by

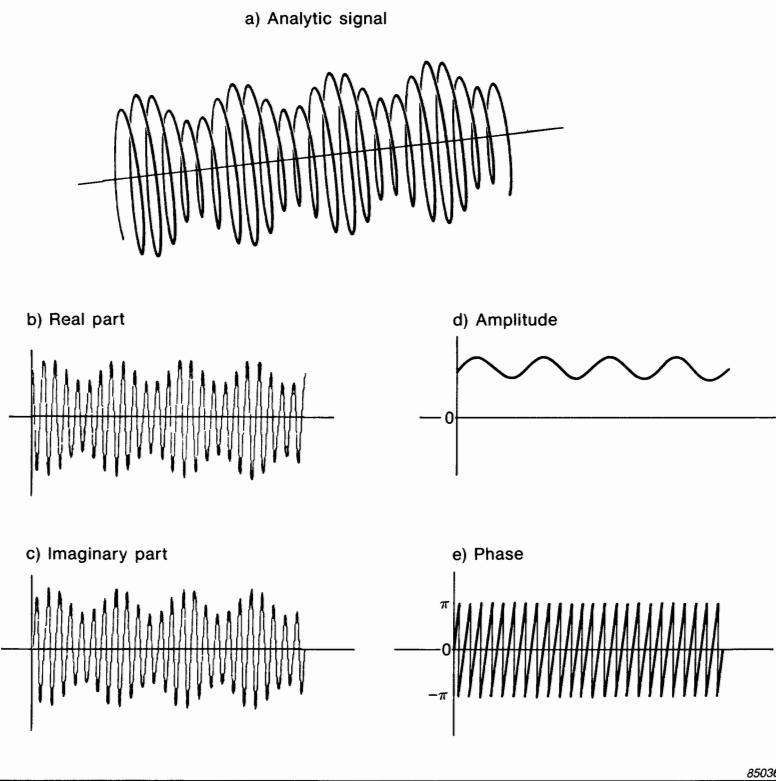
$$\phi(t) = 2\pi f_0 t + \beta \sin 2\pi f_m t \quad (2.58)$$

and differentiating this to obtain the instantaneous frequency, gives

$$\begin{aligned} f(t) &= \frac{1}{2\pi} \frac{d}{dt} (2\pi f_0 t + \beta \sin 2\pi f_m t) \\ &= f_0 + \beta f_m \cos 2\pi f_m t \end{aligned} \quad (2.59)$$

The maximum frequency deviation Δf is thus βf_m , and β ($= \Delta f/f_m$) is known as the "modulation index". Fig.2.43 depicts the instantaneous phase and frequency for a typical case, and illustrates β and Δf .

The decomposition of a phase modulated analytic signal into a sum of constant frequency phasors is somewhat complex, and involves Bessel functions. Fig.2.44 (from Ref. 2.5) illustrates the (one-sided) amplitude spectra for various values of modulation index β . It can be shown that the relative ampli-



850369

Fig. 2.42. Analytic signal for amplitude modulation

tude of the carrier frequency component is given by $J_0(\beta)$, and of the n^{th} order sidebands by $J_n(\beta)$, this being a Bessel function of the first kind, of order n and argument β .

It is seen from Fig.2.44 that for $\beta < 1$ most information is contained in the first two pairs of sidebands, and for $\beta \ll 1$ only one pair of sidebands is required. It is interesting to examine these two cases in more detail, as they give considerable insight into the interaction with amplitude modulation.

Fig.2.45(a) shows the amplitude modulation case of Fig.2.41, in a coordinate system rotating with the (positive) carrier frequency component f_0 (i.e. multiplication by $e^{-j2\pi f_0 t}$, or subtraction of frequency f_0 from all components in the spectrum). As before, the vector sum of the upper and lower sideband phasors is always in line with the carrier phasor.

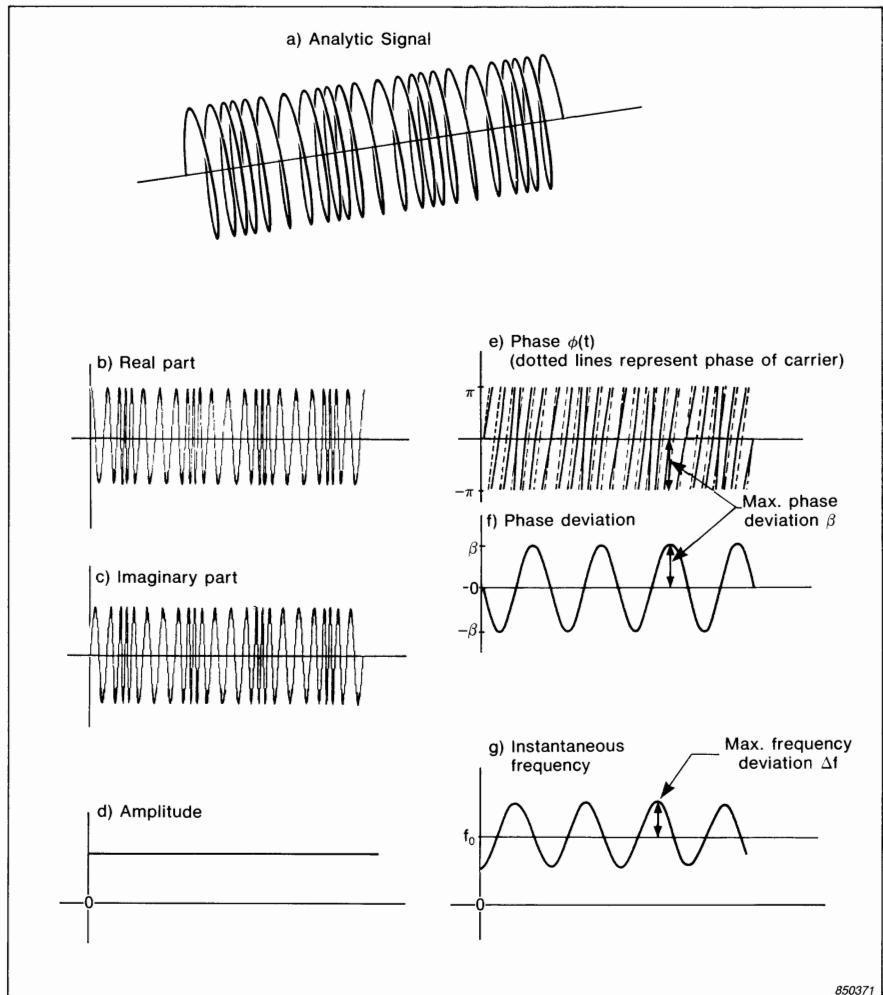


Fig. 2.43. Analytic signal for phase and frequency modulation

If, as illustrated in Fig.2.45(b), the phase of one of the sidebands is reversed, they will now add up to give a phasor of sinusoidally varying amplitude at right angles to the carrier component. The resultant (hypotenuse) formed by vector addition with the carrier will thus to a first approximation have roughly constant amplitude, but sinusoidally varying phase around the carrier. This model of phase modulation, with only one pair of sidebands, will obviously only be valid for phase deviation $\beta \ll 1$. For larger values of β , for example 1 radian, it is necessary to add higher order sidebands to compensate for the varying length of the hypotenuse. This is illustrated in Fig.2.46, where the actual values of

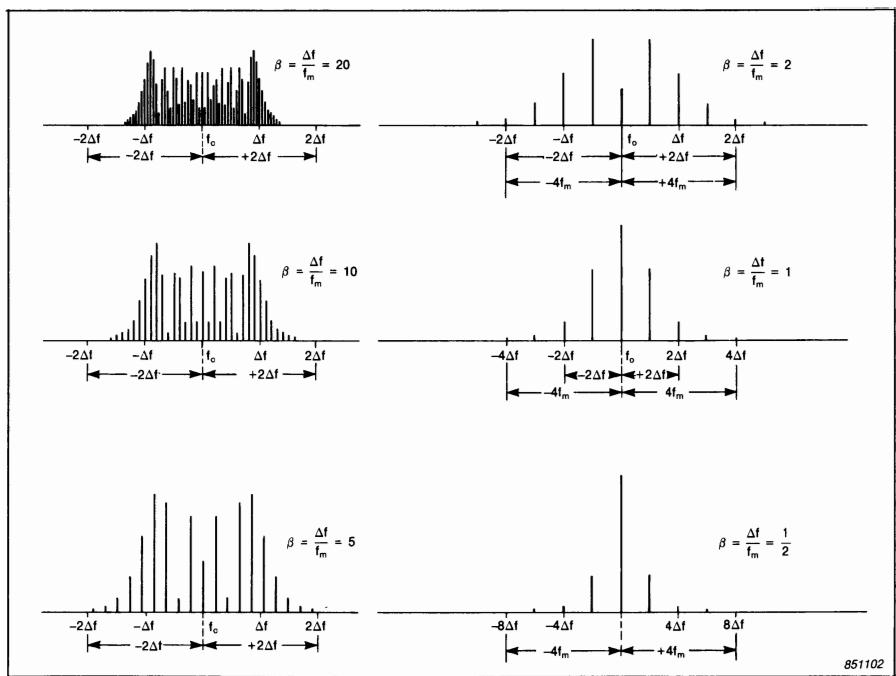


Fig. 2.44. Amplitude spectra of sidebands for frequency modulation with various values of modulation index β

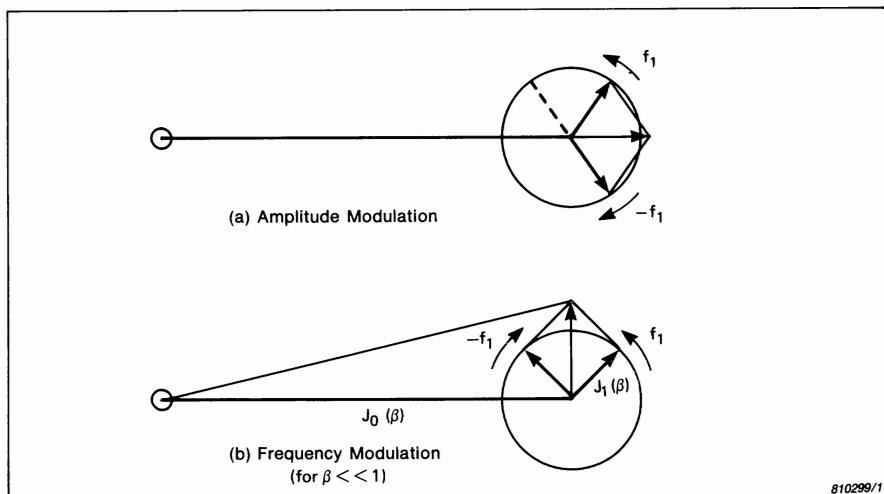


Fig. 2.45. Comparison of phase relationships of sidebands for amplitude and phase modulation

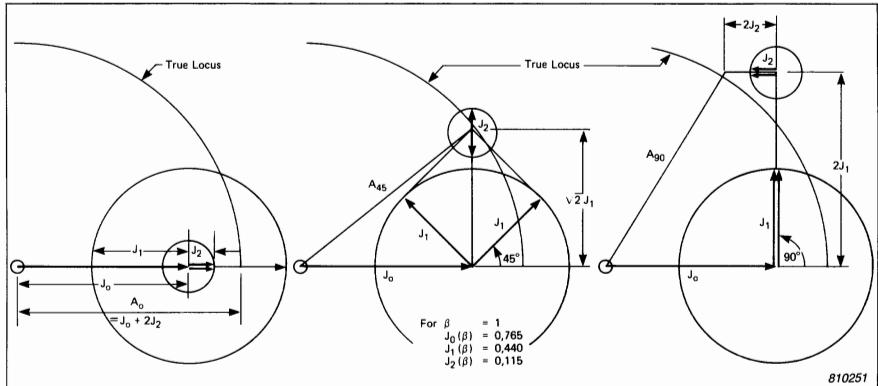


Fig. 2.46. Illustration of how the second order sidebands compensate for the varying length of the hypotenuse

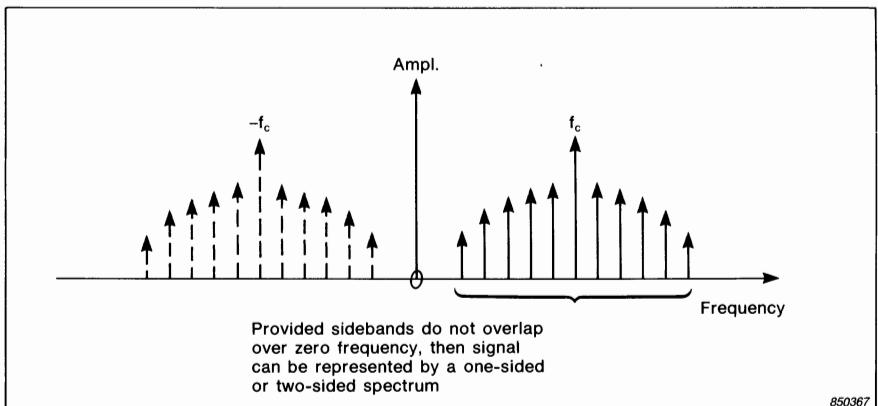


Fig. 2.47. Restriction of extent of sidebands for a single carrier component

$J_0(1)$, $J_1(1)$ and $J_2(1)$ have been used in a graphical construction of the resultant phasor, using up to second order sidebands. Over a time period corresponding to a quarter revolution of the first order sidebands (of length $J_1(1)$), the second order sidebands (of length $J_2(1)$) rotate a half revolution. A_0 , A_{45} and A_{90} represent the length of the resultant (using only two pairs of sidebands) for angular displacements of the first order sidebands of 0, 45, and 90 deg, respectively. In the figure it is seen that they compare very favourably with the constant amplitude vector (represented by the circle segment, the "true locus") which results from including all sidebands.

It is thought that this example illustrates how the sidebands from pure amplitude modulation, or pure phase modulation, would interact in the more general case. Even though either, in isolation, gives symmetrical sideband patterns, the phase relationships on either side of the carrier are different, often giving reinforcement on one side, and cancellation on the other.

It will also be appreciated that the use of analytic signals to model amplitude and phase modulated signals is only simply applicable when the modulation sidebands do not overlap over zero frequency, as illustrated in Fig.2.47. In practice this means that for amplitude modulation, the carrier frequency must be greater than the highest modulating frequency. For frequency modulation it must be greater than both $4f_m$ and $2\Delta f$ (see Fig.2.44), where f_m is now the highest modulating frequency, and Δf is the maximum frequency deviation corresponding to the maximum modulating signal amplitude. For combined amplitude and frequency (or phase) modulation, the lowest sideband from the phase modulated part ($e^{j\phi(t)}$) must be higher than the highest frequency in the amplitude modulation signal ($A(t)$).

3. FILTER ANALYSIS OF STATIONARY SIGNALS

The classical method of obtaining the frequency spectrum of an electrical signal is to pass it through a number of analog filters with different centre frequencies (or one filter whose centre frequency is moved over a frequency range) and measure the transmitted power at each frequency. This subject is thus treated first, even though FFT techniques (see Chapters 4, 5, 6, 7, 8) are currently more important, in particular for constant bandwidth analysis. Analysis using filters (including digital filters) still has advantages for constant percentage bandwidth analysis.

With the filter fixed at one centre frequency, its output will be the result of convolving the input signal with the filter impulse response. In the frequency domain this corresponds to a multiplication of the two (complex) frequency functions. Thus the transmitted signal will have an amplitude spectrum equal to the product of the two individual amplitude spectra and consequently a power spectrum (i.e. amplitude squared) equal to the product of the two power spectra (Fig.3.1). At the same time the phase relationships of the various signal compo-

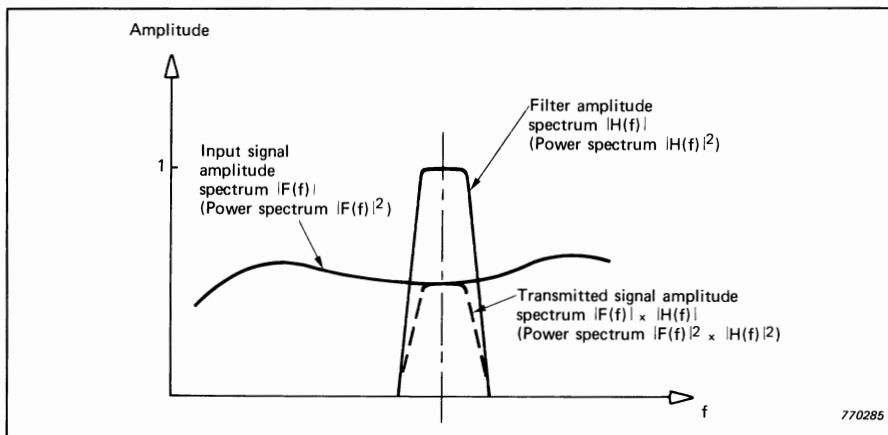


Fig. 3.1. Amplitude and power spectra for a filtered signal

nents will be modified by the filter, but this has no influence on the transmitted power.

Fig.3.2 shows a typical analyzer system in simple block diagram form. In order to obtain a complete spectrum the filter centre frequency must be stepped or swept over the frequency range of interest, and the rate at which this can be done is determined by the delays which arise in each of the major components, viz. filter, detector and recorder. Accordingly, each of these will be considered in detail, including a discussion of any practical details which can influence the results.

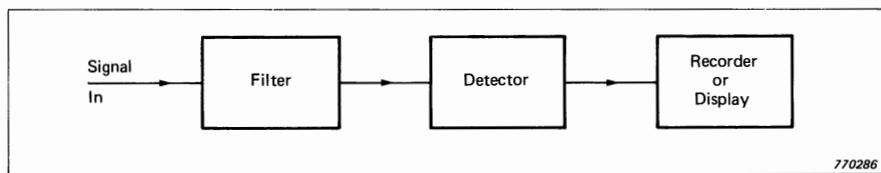


Fig. 3.2. Block diagram for a basic analyzer system using filters

Originally, all three elements were analog, but recent years' development in digital techniques has led to a situation where it is often preferable to use digital filters and detectors. In this case the signal is analog-to-digital (A/D) converted at the input, giving a continuous stream of digital samples. The digital filter receives the sequence of input data values, carries out some digital operation on them and outputs a corresponding sequence of digital values which are filtered in some way with respect to the input (Fig.3.3). In fact the discussion will be limited to so-called "recursive digital filtering". In contrast to FFT, which operates on whole blocks of data at a time, recursive digital filtering is a continuous process and for every input data value an output data value is obtained. In this way, digital filtering is similar to analog filtering, and in fact it is

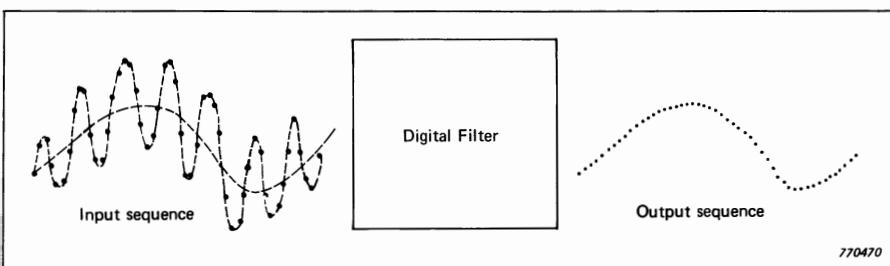


Fig. 3.3. Input to and output from a digital filter

possible to design digital filters with properties similar to those of virtually any practical filter. It is even possible to design digital filters which are not physically realisable in analog form, but these generally have undesirable properties and will not be considered here.

Hence, most of the discussion in this chapter will apply equally to analog and digital filtering. Section 3.4 discusses the special considerations applicable to analog analysis, and Section 3.5 the special considerations applicable to digital analysis. Chapter 3 is limited mainly to analysis of stationary signals, since analysis of transients and non-stationary signals are treated in Chapters 5 and 6, respectively. These chapters will, however, make reference to properties of filters, detectors and recorders as discussed in the following.

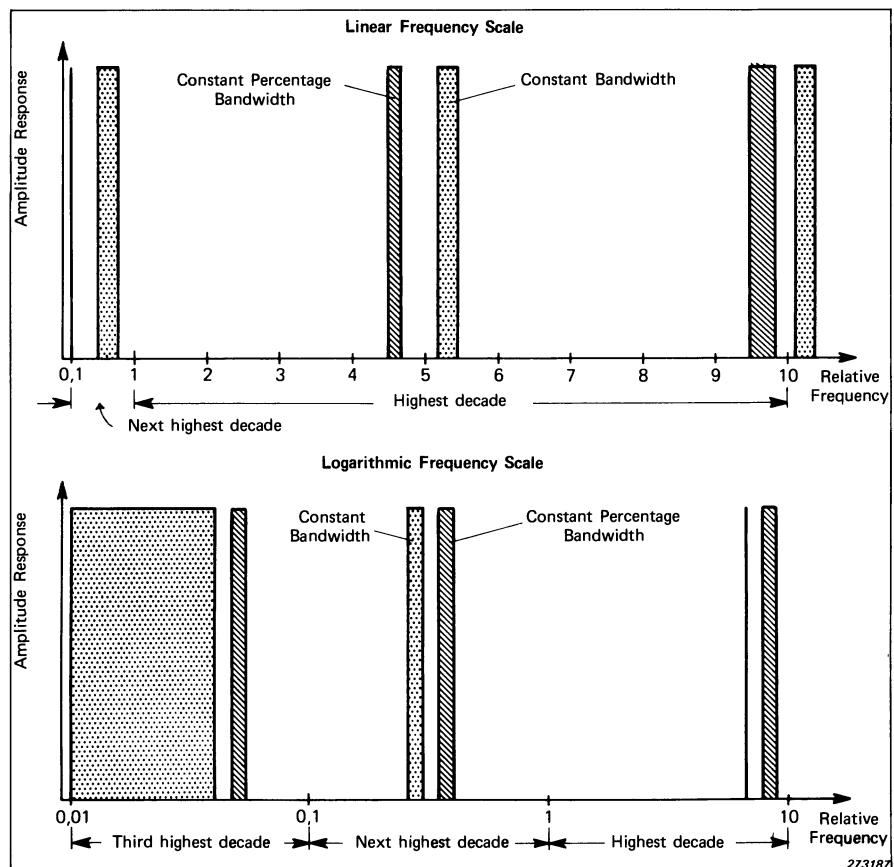


Fig. 3.4. Difference between a constant bandwidth analyzer and a constant percentage bandwidth analyzer

3.1. FILTERS

A basic choice to be made is between constant absolute bandwidth and constant relative (percentage) bandwidth where the absolute bandwidth is a fixed percentage of the tuned centre frequency. Fig.3.4 compares these two alternatives on both linear and logarithmic frequency scales and illustrates one of the most fundamental differences between them.

Constant bandwidth gives uniform resolution on a linear frequency scale and this, for example, gives equal resolution and separation of harmonically related components and this will facilitate detection of a harmonic pattern (Fig.3.5). However, the linear frequency scale automatically gives a restriction of the useful frequency range to (at the most) two decades as is evident from Fig.3.4.

Constant percentage bandwidth, on the other hand, gives uniform resolution on a logarithmic frequency scale and thus can be used over a wide frequency range of 3 or more decades (Fig.3.4.). Another feature of constant percentage bandwidth is that it corresponds to constant Q-factor (amplification ratio of resonance peaks) (Fig.3.6). It is thus both natural and efficient to analyze spectra dominated by structural resonances on a logarithmic frequency scale with a constant percentage bandwidth somewhat narrower than the narrowest resonant peak.

Other grounds for using a logarithmic frequency scale (though not necessarily constant percentage bandwidth) are:

- a) Small speed changes in, say, a machine only give a lateral displacement of the spectrum, thus simplifying direct comparison.

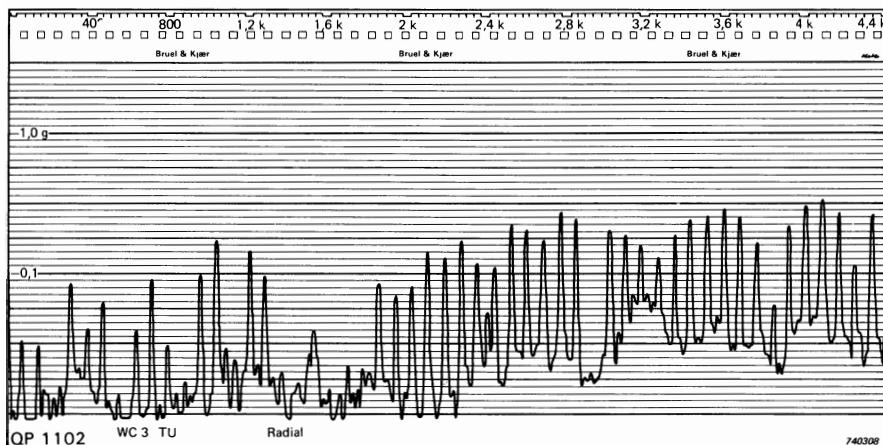


Fig. 3.5. Vibration spectrum having many harmonically related components

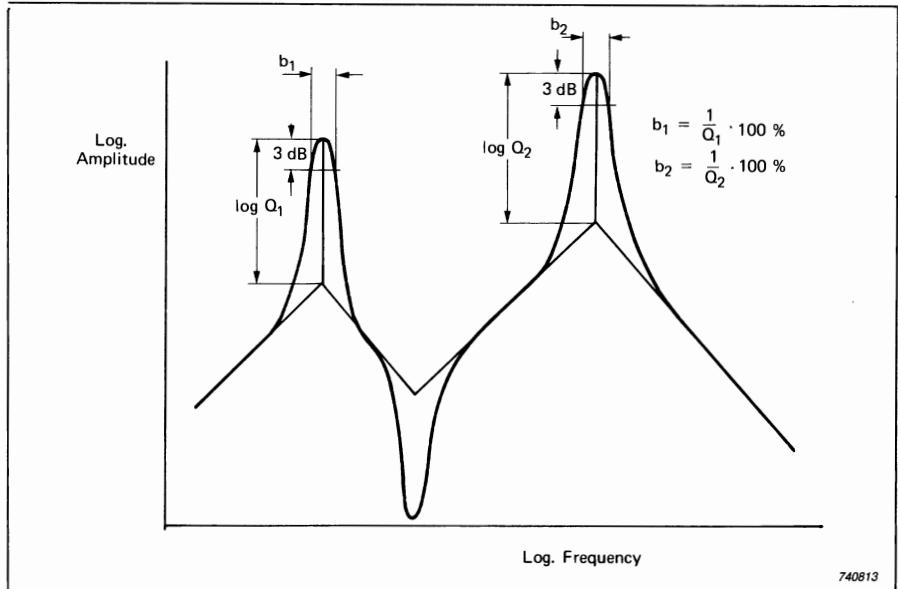


Fig. 3.6. Relationship between the amplification factor Q and the relative bandwidth b

- b) Certain relationships can most easily be seen on log-log scales such as for example, integration, which gives a change in slope of -20 dB/decade and thus means that constant velocities and displacements are represented by straight lines on an acceleration vs. frequency diagram.

It is worth paying particular attention to two special classes of constant percentage bandwidth filters, viz. octave and third octave filters since these are widely used, in particular for acoustical measurements. The former have a bandwidth such that the upper limiting frequency of the passband is always twice the lower limiting frequency, resulting in a bandwidth of 70,7%. This can be derived as follows:

If f_l = lower limiting frequency
 f_u = upper limiting frequency
 f_o = nominal centre frequency

Then $f_u = 2f_l$

and $f_o = \text{the geometric mean} = \sqrt{f_u \cdot f_l} = \sqrt{2f_l^2} = \sqrt{2} f_l$

The absolute bandwidth = $f_u - f_l = f_i$

and the relative bandwidth = $\frac{f_u - f_l}{f_o}$

$$= \frac{f_i}{f_o} = \frac{f_i}{\sqrt{2}f_l} = \frac{1}{\sqrt{2}} = 70,7\%$$

Internationally standardized centre frequencies for octave filters are laid down in IEC Recommendation 225 which specifies a set of contiguous filters based on a reference centre frequency of 1000 Hz.

Thus it can be seen that it is possible to cover 3 decades in frequency with 10 octave bands ranging from 22,5 Hz (lower limiting frequency for 31,5 Hz centre frequency) to 22,5 kHz (upper limiting frequency for 16kHz centre frequency).

Third octave filters are obtained by dividing each octave band into three geometrically equal sub-sections, i.e. $f_u = 2^{1/3} f_l$ and by coincidence this is equal to one-tenth of a decade since

$$\log_{10}(2^{1/3}) = 1/3 \log_{10}(2) = 1/3 \cdot 0,3 = 0,1 = 1/10 \log_{10}(10) = \log_{10}(10^{1/10})$$

By the same procedure as for octave filters, the percentage bandwidth of third octave filters can be derived as:

$$\frac{2^{1/3} - 1}{2^{1/6}} = 23,1\%$$

3.1.1. Practical Filter Characteristic

As mentioned in Section 2.3 practical filters deviate from ideal filters in several ways as illustrated in Fig.2.10. Provided the “ripple” within the passband is kept within acceptable limits, the main characteristics of a practical filter will be its bandwidth and selectivity.

The so-called “noise bandwidth” of a filter has already been defined in Section 2.3 as the width of an ideal filter which transmits the same power from a white noise source, for the same reference transmission level in the passband. This definition is most relevant when dealing with random signals, since results derived for ideal filters are then generally applicable to practical filters with the same noise bandwidth.

Another bandwidth which can be associated with a filter is its “3dB bandwidth” and this is simply the width of the power transmission characteristic at the “3dB points”, which as their name implies lie 3dB below the nominal (unity) amplification (Fig.2.10). It happens that this is often very close to the Noise Bandwidth and since it is much more easily measured, it is commonly used. The

3dB bandwidth is perhaps most relevant when dealing with deterministic signals, since it gives information about how well sinusoidal components can be separated. Only filters with a relatively poor selectivity will have a 3dB bandwidth substantially different from the noise bandwidth.

The *bandwidth* of a filter gives information as to its ability to separate components of approximately the same level. The *selectivity* indicates its ability to separate components of widely different levels. The most basic parameter indicating selectivity is known as the "Shape Factor". This is normally defined as the ratio of the width of the filter characteristic where the attenuation of the flanks is 60 dB, to its 3 dB bandwidth (Fig.3.7). In some cases, such as where the dynamic range is in any case less than 60 dB, terms such as "40 dB Shape Factor" may be used to represent the equivalent factor obtained using the breadth of the characteristic at 40 dB attenuation (Fig.3.7). Shape factor is normally used for constant bandwidth filters which have a characteristic which is symmetrical on a linear frequency scale, but it could be applied more generally.

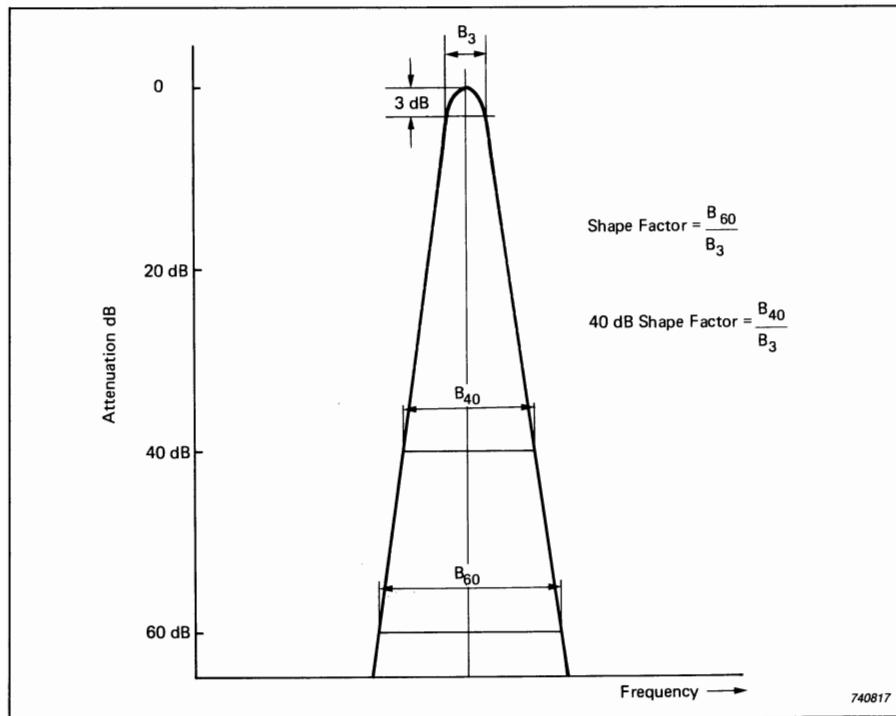


Fig. 3.7. Shape factor

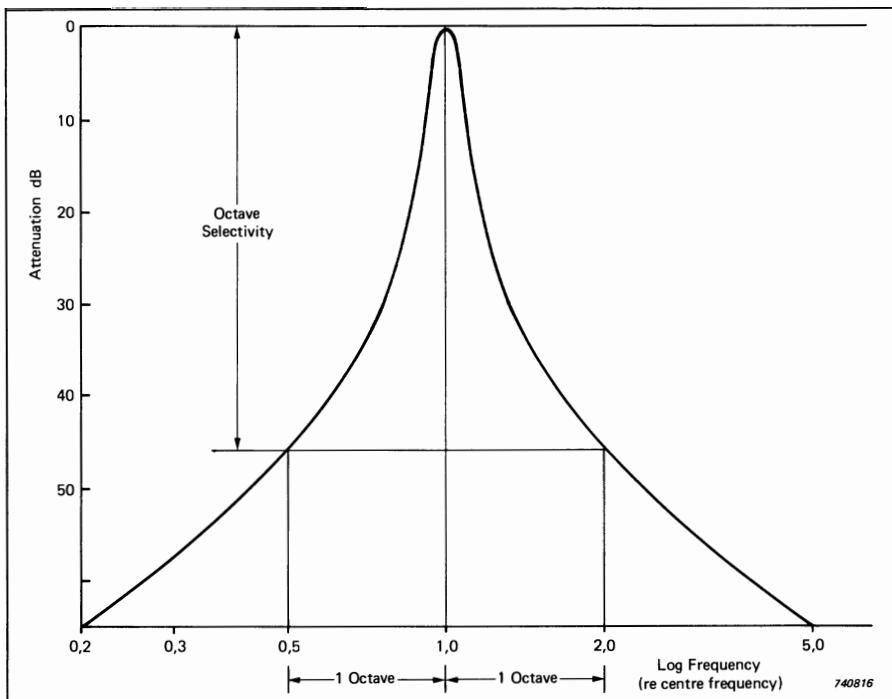


Fig. 3.8. Octave selectivity

For constant percentage bandwidth filters, however, which have a characteristic which is symmetrical on a logarithmic frequency scale, it is more common to use "Octave Selectivity", which gives the attenuation of the filter characteristic one octave on either side of the centre frequency (Fig.3.8).

3.1.2. Filter Response Time

When a signal is suddenly applied at the input of a filter, it takes some time before the latter responds. If the signal is sinusoidal and has a frequency within the passband of the filter, then the output will finally be a sinusoid of the same frequency and with the same amplitude as the original (assuming the amplitude characteristic has an amplification of 1 within the passband). The time required for the amplitude to approach its final value, however, is of the order of $1/B$ where B is the filter bandwidth. In principle, the total response time has two components, a "dead time" dependent on both the bandwidth and the number of poles in the filter circuit, and the "rise time", dependent on the bandwidth

only. Most practical filters have either four or six poles, and the response time can be taken to a first approximation as the reciprocal of the bandwidth.

This may be expressed in the form

$$BT_R \approx 1 \quad (3.1)$$

where T_R = filter response time.

Note that this may be modified as follows

$$\left(\frac{B}{f} \right) \cdot (fT_R) \approx 1$$

$$\text{i.e. } bn_R \approx 1 \quad (3.2)$$

where b = relative bandwidth,

n_R = number of periods of frequency f in time T_R
e.g. for $b = 1\%$, n_R will be approx. 100 periods, etc.

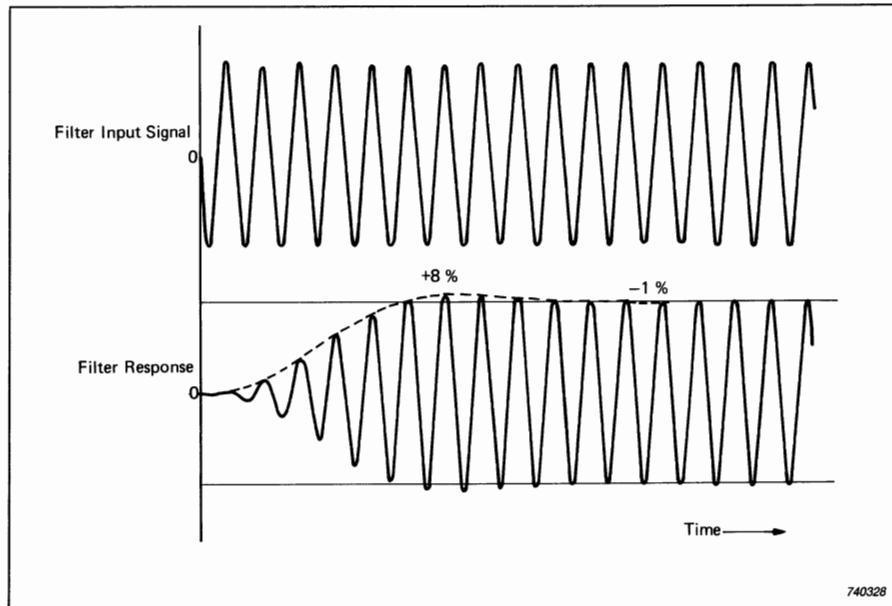


Fig. 3.9. Typical filter response from a 1/3 octave filter

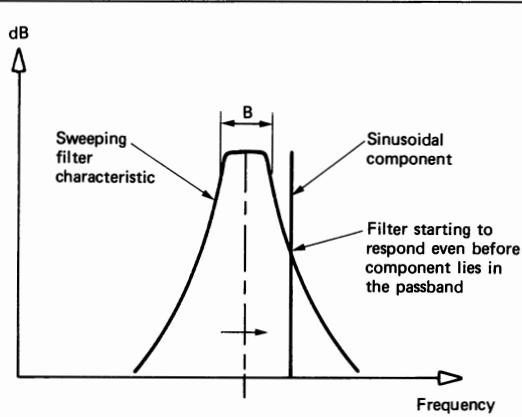
Thus, Equation (3.1) will be most applicable to constant bandwidth filters, while Equation (3.2) is most applicable to constant percentage bandwidth filters.

These relationships may be thought of as another manifestation of the physical requirement that a measurement with bandwidth B requires a measurement time of at least $1/B$.

Fig.3.9 shows the response of a typical $1/3$ -octave filter to a suddenly applied sinusoid. Since the bandwidth is 23.1% one would expect the response time to be 4.3 periods, whereas it is seen to be actually of the order of 5 or 6 depending on the desired accuracy. This gives an idea of the order of accuracy of Equations (3.1) and (3.2), although $1/3$ -octave filters, with their relatively steep filter characteristic, probably represent an extreme case. Another mitigating factor with swept frequency analysis is that the application of the sinusoid is not sudden, as would be the case with an ideal filter, but more gradual as the filter flank moves over the sinusoidal component (Fig.3.10). All-in-all, the equations (3.1) and (3.2) give a good estimate of the delays introduced for the case of swept filtration taking into account that they are used to select parameters such as paper speed, which normally are varied in steps of $3:1$.

3.2. DETECTORS

When a signal has been passed through a filter the output is still in the form of an AC signal which varies continuously with time. It is necessary to measure the power of this signal to obtain the desired frequency spectrum component. This



770287

Fig. 3.10. Gradual signal application with a sweeping filter

can be done mathematically by squaring the instantaneous value of the signal to obtain the instantaneous power and then integrating this over a defined time interval (the averaging time) to obtain an average value. The longer the averaging time, the smaller will be the variations in this average value, but the longer it takes to obtain the result. It is often desired to extract the square root of this mean square value to give the Root Mean Square (RMS) amplitude, since this has the same dimensions as the input signal. It also gives the practical advantage that the same output voltage range represents twice the dB range and thus gives a wider dynamic range in general.

Finally, it may be desired that the Mean Square or RMS signal be logarithmically converted to give a result in dB.

All these functions can be performed by electronic circuits known as detectors, giving a result with a high degree of accuracy with respect to the theoretical value.

3.2.1. Squaring

All high quality (so-called "true RMS") detectors square the input signal in one way or another. On the other hand, so-called "averaging detectors" perform only a rectification (as an approximation to squaring) and then average the rectified signal, but although the relationship between the "average" and "RMS" level is well-known for a sinusoidal signal, the average value for a more complex signal depends on the phase relationships between different frequency components, whereas the RMS value does not.

For many years B&K utilised the "Wahrman" detector, named after its inventor, which achieves squaring of the (rectified) signal by means of a characteristic whereby the parabola representing the squaring relationship is approximated by a number of straight lines. Fig.3.11 illustrates a number of such parabolae for a typical detector, along with the straight lines which approximate them. The reason for the number of different parabolae is connected with square root extraction (see Section 3.2.3) and will not be gone into here; instead consideration will be limited to a single parabola which corresponds to a given RMS level. For a given parabola, it will be seen that the larger the instantaneous level of the input signal, the further out along the parabola one is working, and that beyond the last breakpoint in the approximating curve the more will be the deviation between the true squared value and the value obtained from the circuit. The ratio of peak to RMS level of a signal is known as its "crest factor", and it will be seen that the more break points there are in the approximating curve, the higher will be the crest factor which can be accommodated. As an example, 4 breakpoints give an accuracy within 1/2 dB for crest factors up to 5.

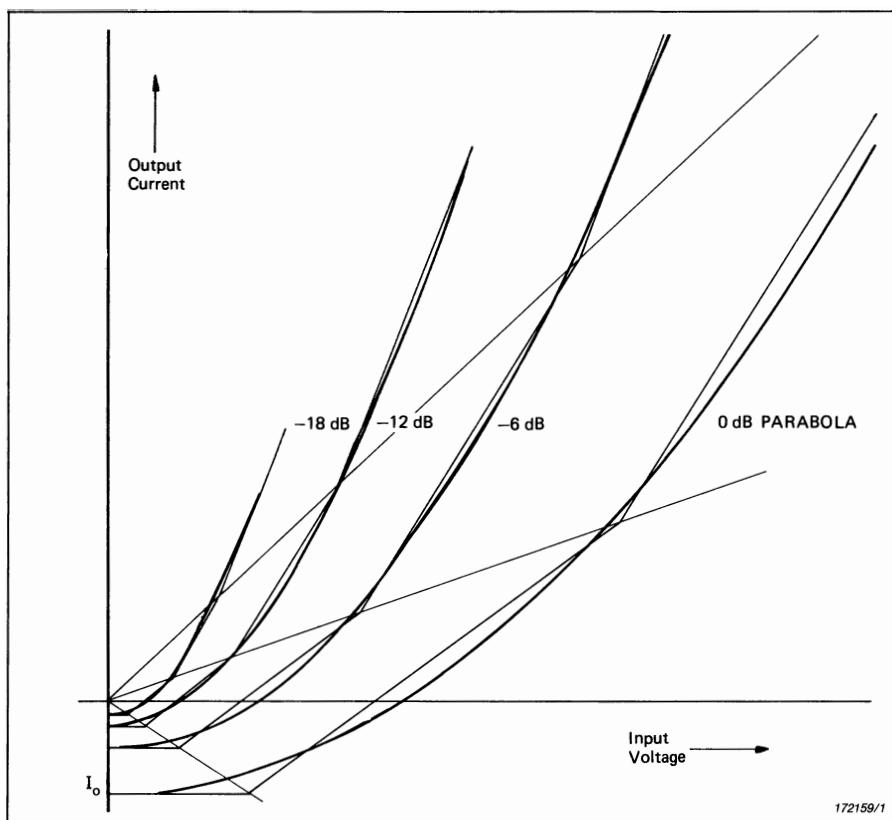


Fig. 3.11. Straight-line approximations to the square-law parabolae

In most newer instruments another type of detector known as an LMS (logarithmic mean square) detector is used. This makes use of the logarithmic characteristic of certain diodes to logarithmically convert the input signal. Squaring is achieved by amplifying the log converted signal by a factor of 2 and mean square averaging carried out in a part of the circuit where the signal has been exponentiated (antilogarithmically converted). The output voltage is proportional to the logarithm of the mean square value and can thus be scaled directly in dB. In comparison with the Wahrman detector already described, it can be said to have a true squaring characteristic, in general with a very high crest factor capability.

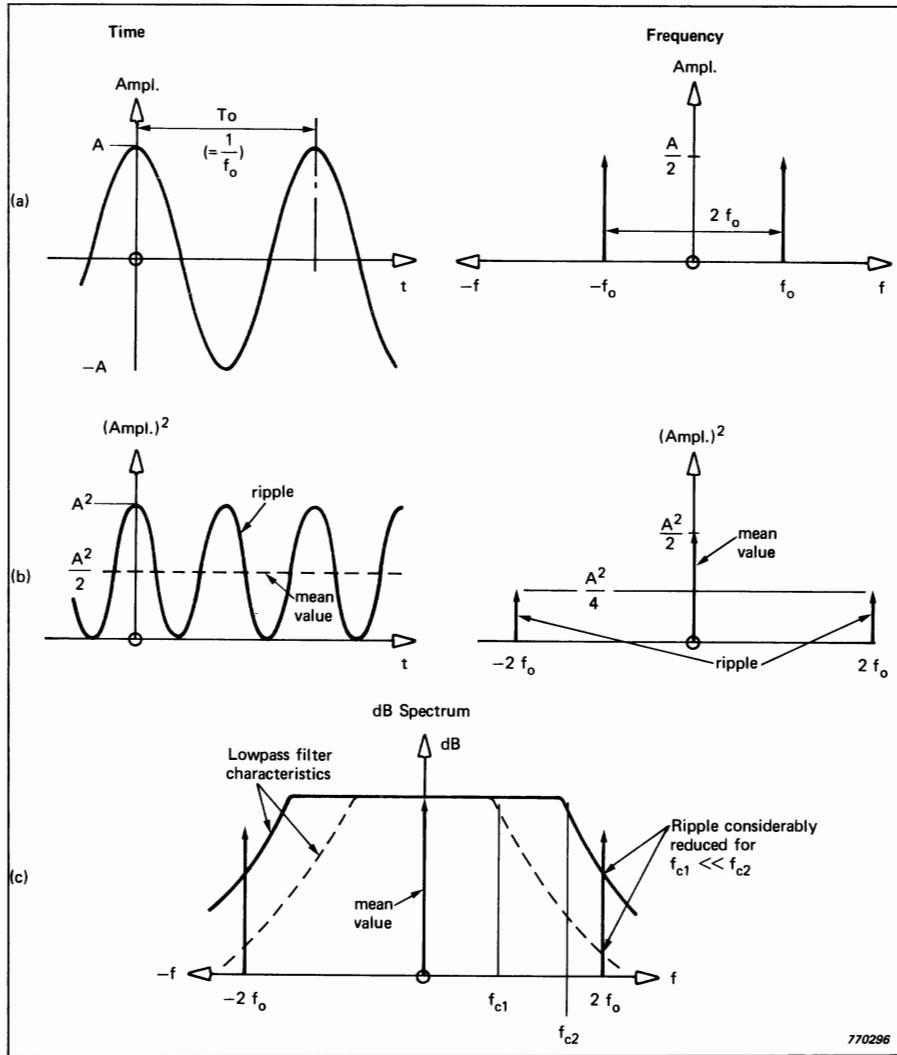


Fig. 3.12. Averaging as a low-pass filtration

3.2.2. Averaging

The purpose of averaging is to suppress the fluctuations in the squared rectified signal from the squaring circuit, thereby obtaining an expression for the mean power, which represents the mean square spectral estimate. It should

be kept in mind that Fourier analysis assumes integration over all time, which can be represented by the following equation:

$$\bar{y} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} y(t) dt \quad (3.3)$$

where the bar represents the average value.

In practice, of course, it is necessary to limit the averaging time to a finite value, and the effect of doing this will now be examined.

The effect on Equation (3.3) is to remove the limiting operation, and this results in a finite fluctuation which gets smaller as the averaging time gets longer. Suppression of the fluctuation can be considered as a lowpass filtration, and Fig.3.12 illustrates why this is so for a typical sinusoidal component of frequency f_o and amplitude A .

Fig.3.12(a) shows the original cosine signal in both time and frequency domains, while Fig.3.12(b) shows the result of squaring it. Note that the spectrum of Fig.3.12(b), while obviously corresponding to the adjacent time signal, can also be derived by convolving the spectrum of Fig.3.12(a) with itself (corresponding to the multiplication of the cosine wave by itself). Thus, in the convolved spectrum there are only components when two or more delta functions from the original spectrum line up, i.e. for displacements (f in Eqn. 2.30) of 0, $+2f_o$ and $-2f_o$. Moreover, for displacement $f = 0$, two sets of delta functions line up so that the total component is $A^2/4 + A^2/4 = A^2/2$ whereas with $f = \pm 2f_o$, the positive frequency delta function of one spectrum lines up with the negative frequency of the other and the resulting component is thus only $A^2/4$. The true mean value is evidently the DC component in Fig.3.12(b), viz. $A^2/2$, and this can be obtained by lowpass filtering the squared signal. Fig.3.12(c) illustrates the effect of filtering with two cut-off frequencies, the lower cut-off frequency giving less ripple and thus corresponding to a longer averaging time. The relationship between lowpass filter characteristic and effective averaging time will now be examined in some detail.

From Equation (3.3) the short-term average of a function $y(t)$ from $-T/2$ to $+T/2$ is given by $1/T \int_{-T/2}^{T/2} y(t) dt$, but more generally the running average obtained at time t over the previous T will be seen to be

$$\langle y \rangle_t = \frac{1}{T} \int_{t-T}^t y(\tau) d\tau \quad (3.4)$$

as illustrated in Fig.3.13(a).

It will be found, however, that this can equally well be written as the (scaled) convolution equation:

$$\langle y \rangle_t = \frac{1}{T} \int_{-\infty}^{\infty} y(\tau) g(t - \tau) d\tau \quad (3.5)$$

where $g(\tau)$ is defined as follows:

$$\begin{aligned} g(\tau) &= 1 & 0 < \tau < T \\ g(\tau) &= 0 & \text{otherwise} \end{aligned} \quad (3.6)$$

To make this more obvious, Fig.3.13(b), (c), (d) and (e) illustrate $g(\tau)$, $g(-\tau)$, $g(t-\tau)$ and $y(\tau) g(t-\tau)$ respectively.

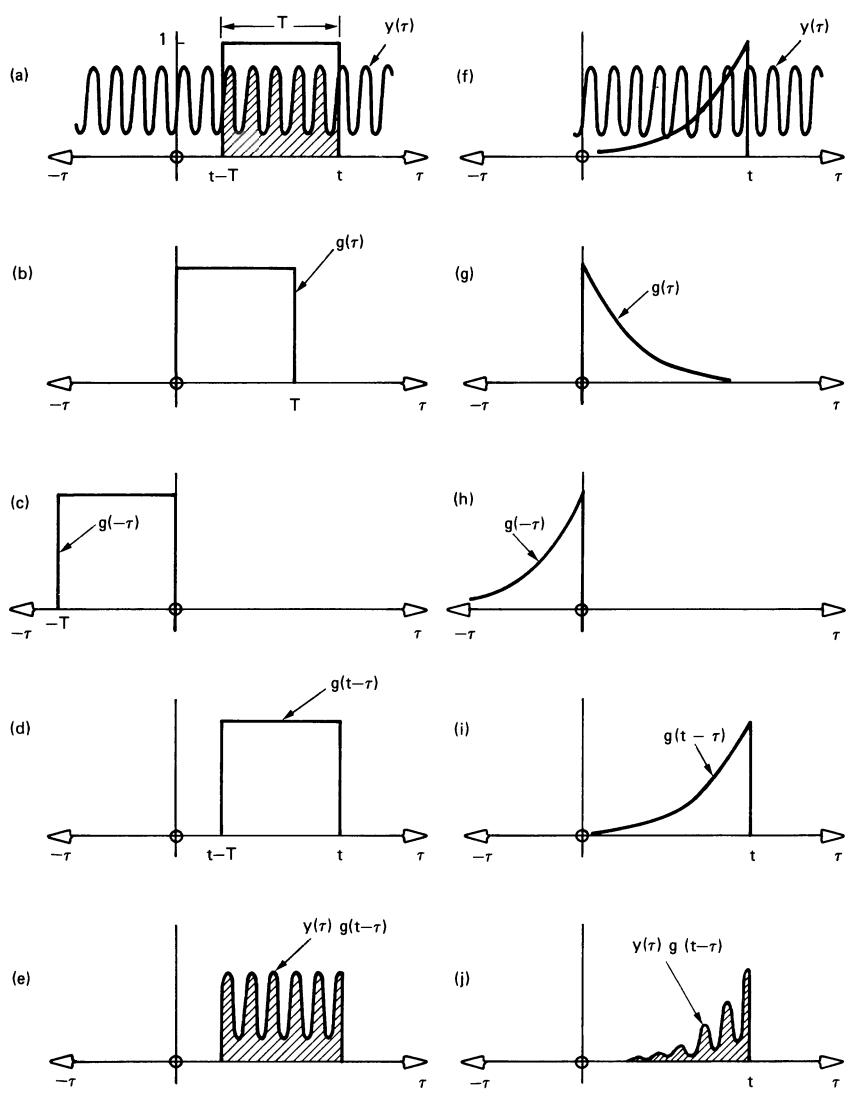
This convolution in the time domain corresponds to a multiplication in the frequency domain of the respective Fourier transforms. The amplitude spectrum of the rectangular function $g(t)$ defined in Eqn. (3.6) is once again the well-known $|\sin x/x|$ function with zeroes at multiples of $1/T$ (Appendix A) and this is found to be a low-pass filter characteristic with a slope of $-20 \text{ dB per decade}$ from a cut-off frequency of $1/\pi T$. The positive half of this characteristic is illustrated on log-log scales in Fig.3.14. Note that the zeroes at multiples of $1/T$ correspond to integral numbers of periods, and thus as pointed out in Section 2.2.1 integration over any integral number of periods completely removes the fluctuation. Note also that the phase characteristic of the Fourier transform of $g(\tau)$ has no influence on the power transmission, only on the phase shift of the ripple component.

Comparing Eqn. (3.5) with Eqn. (2.28) it will be seen that $g(\tau)$ as defined in Eqn. (3.6) can be interpreted as the required impulse response of a circuit to give running linear integration over averaging time T . This would be extremely difficult to realise in practice, and it is more common to use other electronic circuits with a low-pass filtering effect. The most commonly used circuit is a so-called RC circuit which has an impulse response defined by:

$$\begin{aligned} g(t) &= 0 & -\infty < t < 0 \\ g(t) &= e^{-t/RC} & 0 < t < \infty \end{aligned} \quad (3.7)$$

The employment of such a characteristic is referred to as "RC-averaging" or "exponential averaging" since the impulse response is a decaying exponential with most weight on the most recent part of the signal.

Fig.3.13(f) to (j) illustrate the resulting convolution for comparison with the rectangular function of Eqn. (3.6).



*Fig. 3.13. Running average as a convolution
 (a) – (e) Linear weighting
 (f) – (j) Exponential weighting*

Fig.3.15 also compares a typical RC characteristic with a rectangular one. It has intentionally been scaled in such a way that $T = 2RC$, since with a peak response equal to twice that of the rectangular response the area under the two curves is the same, and thus the scaling will be the same for averaging of stationary signals. More importantly, however, it is found that with $T = 2RC$, the two filter characteristics in the frequency domain have the same asymptotic lines and the same effective bandwidths. (See Fig.3.14.) The frequency response corresponding to the exponential characteristic of Equation (3.7) and drawn in Fig.3.14 is derived in Appendix A.

Thus, for an averaging time equal to several ripple periods (as will normally be required in practice to reduce ripple to an acceptable level) it will be seen that because the lowpass filter characteristics have the same asymptotic lines the maximum ripple which can be obtained with linear averaging over time T_A is the same as that obtained by exponential averaging with an RC-time constant such that

$$T_A = 2RC \quad (3.8)$$

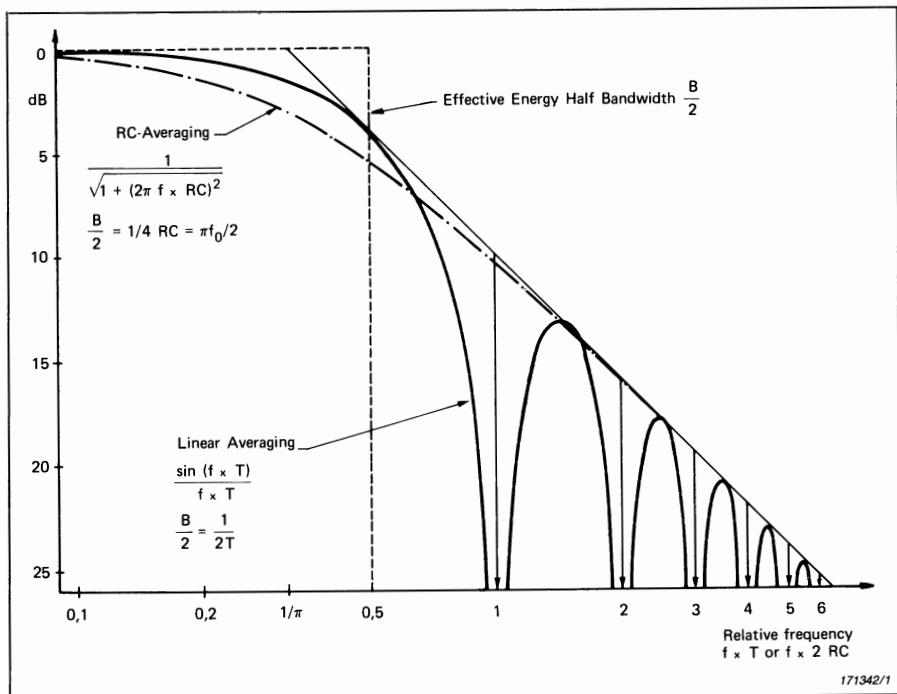


Fig. 3.14. Comparison of linear and RC-averaging in frequency domain

and thus Eqn. (3.8) expresses the equivalent averaging time for an RC circuit applied to a sinusoidal signal.

The ripple obtained for averaging over various numbers of periods (Ref.3.1) is illustrated in Fig.(3.16).

When the signal coming from the filter is narrow band random noise (i.e. the result of filtering wideband noise with a narrow band filter) it is found (Appendix B) that the relative fluctuation (variance) of the result is proportional to the noise bandwidth of the lowpass filter characteristic. It has already been shown that for the $|\sin x/x|$ characteristic resulting from a rectangle of length T the noise bandwidth is $1/T$ (see Appendix A) and in Appendix A it is demonstrated that the noise bandwidth of the lowpass filter characteristic corresponding to exponential averaging is also the same for the situation illustrated in Fig.3.14 (i.e. with the same asymptotic lines) and thus Equation (3.8) also expresses the equivalent averaging time for stationary random signals. In Ref.3.1 it is shown that the same equivalence holds for trains of impulses (see Fig.3.17) so that it can be taken that Equation (3.8) holds for all stationary signals.

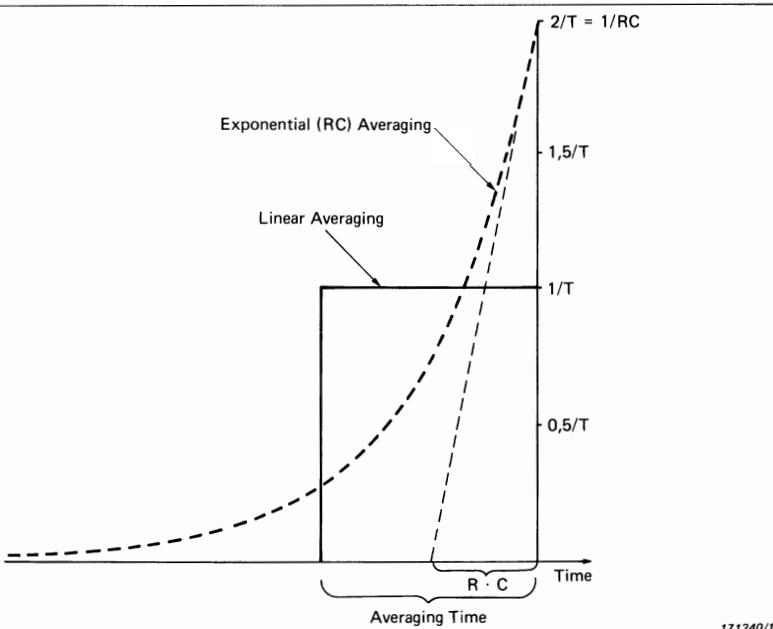


Fig. 3.15. Weighting curves for linear and RC-integration

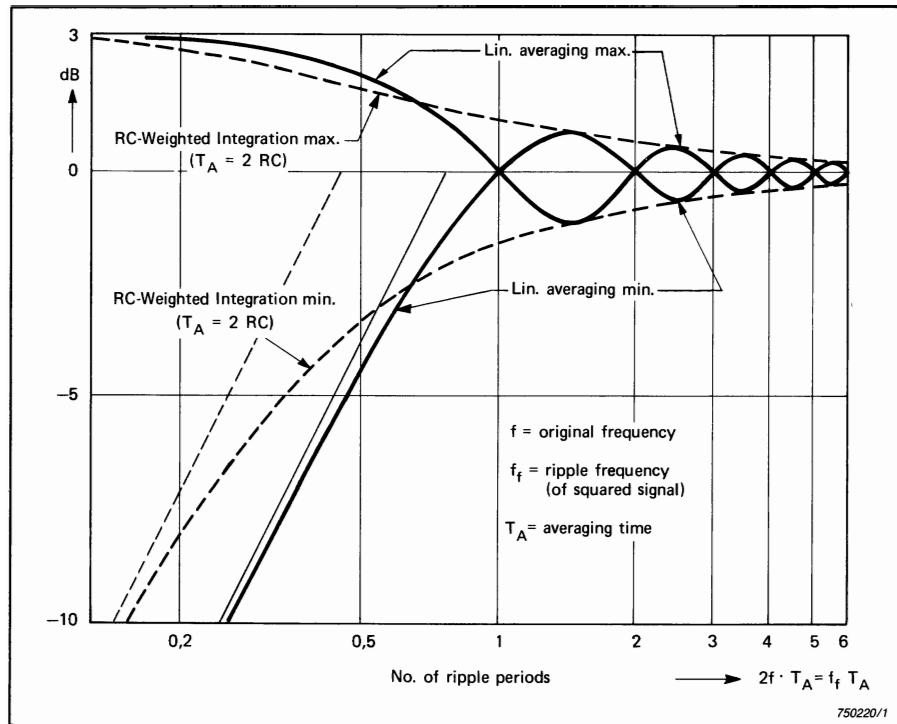


Fig. 3.16. Comparison of ripple components for linear and RC averaging of squared sinewaves

Figs. 3.16 and 3.17 can be used for choosing averaging times for measurements on deterministic signals, according to the desired level of ripple. For random signals, the approach is somewhat different. In Appendix B it is shown that when a narrow band noise of bandwidth B is applied to an RMS detector with averaging time T_A the relative standard deviation of the measured RMS level is expressed by the equation:

$$\epsilon = \frac{1}{2\sqrt{BT_A}} \quad (\text{for } BT_A \gg 1) \quad (3.9)$$

or, in decibel form $\epsilon = \frac{4,34}{\sqrt{BT_A}} \text{ dB}$ (3.9.a)

What this means is illustrated in Fig. 3.18 for three different averaging times. The result is itself a random variable but with a probability distribution which lies closer and closer to the true value (i.e. that for infinite averaging time) the longer the averaging time. Thus it is possible, though unlikely, for a shorter

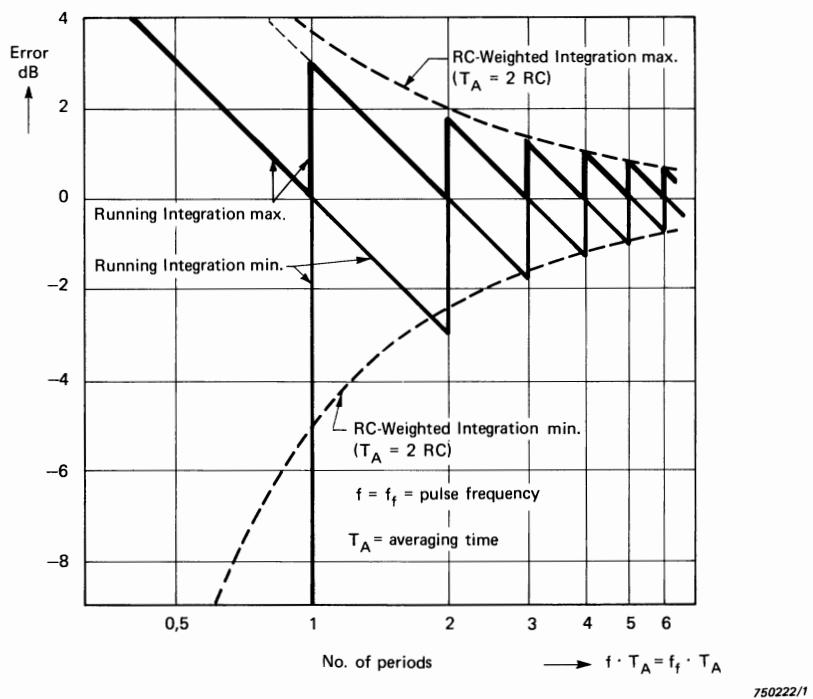


Fig. 3.17. Comparison of ripple components for linear and RC averaging of pulse trains

averaging time to give a more accurate result, but on the other hand, by integrating over the different regions of the probability density curve it can be shown that for a single estimate there is a 68,3% chance of it being within $\pm \epsilon$ of the true value, 95,5% chance of it being within $\pm 2\epsilon$ and 99,7% chance of it being within $\pm 3\epsilon$.

Note that the equivalence expressed in Equation (3.8) does not apply directly to transient signals, but this point is discussed in some detail in the sections dealing with the analysis of transients.

3.2.3. Square Root Extraction

In Section 3.2.1 it was explained how a squaring circuit could be used to square an input signal $x(t)$ such that the output $y(t)$ could be represented by the equation

$$y(t) = kx^2(t) \quad (3.10)$$

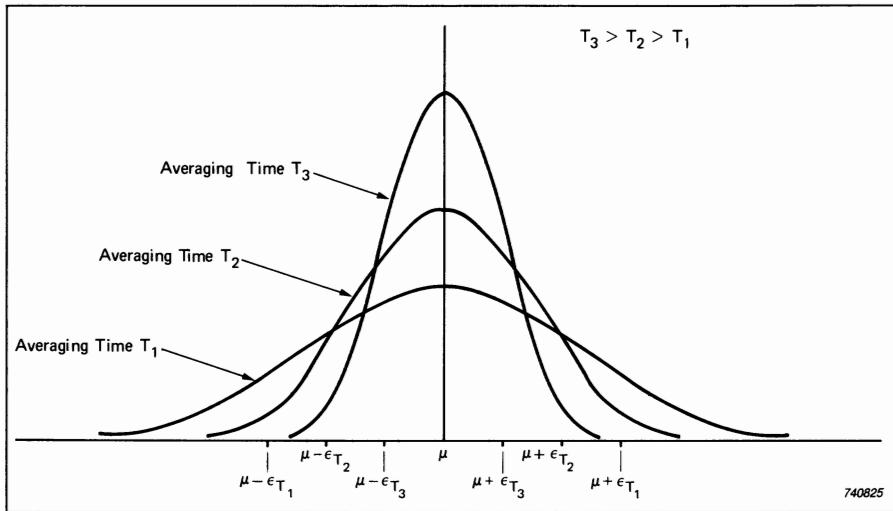


Fig. 3.18. Influence of the averaging time on the error distribution

where k is a constant determining the size of the parabola. (Note that k is large for a narrow parabola and vice versa.)

In Section 3.2.2 it was shown that it was possible using an RC circuit to obtain the average of the squared signal. Thus:

$$\overline{y(t)} = k \overline{x^2(t)} \quad (3.11)$$

In the foregoing it was assumed that the parameter k was a constant, but it is possible to feed back the circuit output signal in such a way that k is variable and proportional to the reciprocal of the output level. The output will thereby be modified, and will be called $y'(t)$. Thus, k can be expressed in the form:

$$k = \frac{k_1}{y'(t)} \quad (3.12)$$

where k_1 is a constant
and $y'(t)$ is a slowly varying function of time because of the inherent averaging time.

Substituting this in an equation analogous to (3.11) gives

$$y'(t) = \frac{k_1}{y'(t)} \overline{x^2(t)}$$

or $\{y'(t)\}^2 = k_1 \overline{x^2(t)}$

and $y'(t) = \sqrt{k_1 \overline{x^2(t)}}$ (3.13)

a scaled approximation to the RMS value of $x(t)$. It will be seen that a large RMS level gives a small k factor and thus a wide parabola as illustrated in Fig.3.11.

The square root extraction properties of the variable-parabola Wahrman detector are discussed in considerably more depth in Ref.3.1 where a more rigorous derivation of its mode of operation is given.

For the LMS detector, as already mentioned, the output voltage is inherently proportional to the logarithm of the mean square value. If a signal proportional to the RMS value is desired, then the reverse of the process at the input is carried out; the signal is first scaled down by a factor of two and then antilogarithmically converted.

3.2.4. Logarithmic Conversion

The result will often be required on a logarithmic, or decibel, amplitude scale and for other than LMS detectors this may be achieved in a logarithmic conversion circuit at the analyzer output, or by using a logarithmic potentiometer in the Level Recorder used to record the spectrum.

3.2.5. Detector Response

The effect of the finite detector response time on a swept frequency analysis is to give both a delay and an error on peaks (and valleys) in a spectrum compared with the true result which would be achieved with an infinitely slow rate of sweep (see Fig.3.19). Moreover, in particular with RC averaging, there is a limit to the steepness of slope which can be reproduced. These factors have been studied in some detail in Ref.3.1 and the results are expressed in Figs.3.20, 3.21 and 3.22. Fig.3.20 expresses the errors which will be experienced at peaks and valleys in a spectrum with running linear integration and RC-integration. Fig.3.21 gives information as to the corresponding delays. Finally, Fig.3.22 gives the delays which will be experienced at sloping areas in the spectrum. Here it is seen that for RC-integration there is an asymptotic limit of $8.7 \text{ dB}/2\text{RC}$ which is the maximum rate at which the detector level can fall, and this will often govern the rate of sweep.

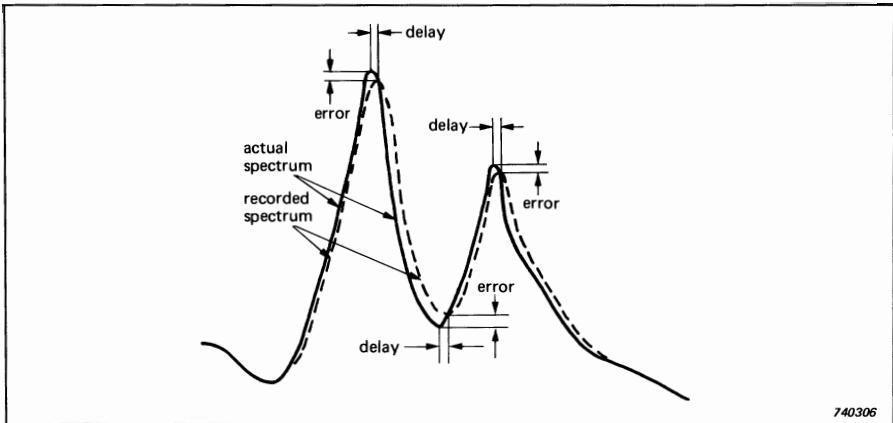


Fig. 3.19. Error and delay (bias errors) in writing out peaks and valleys in a spectrum

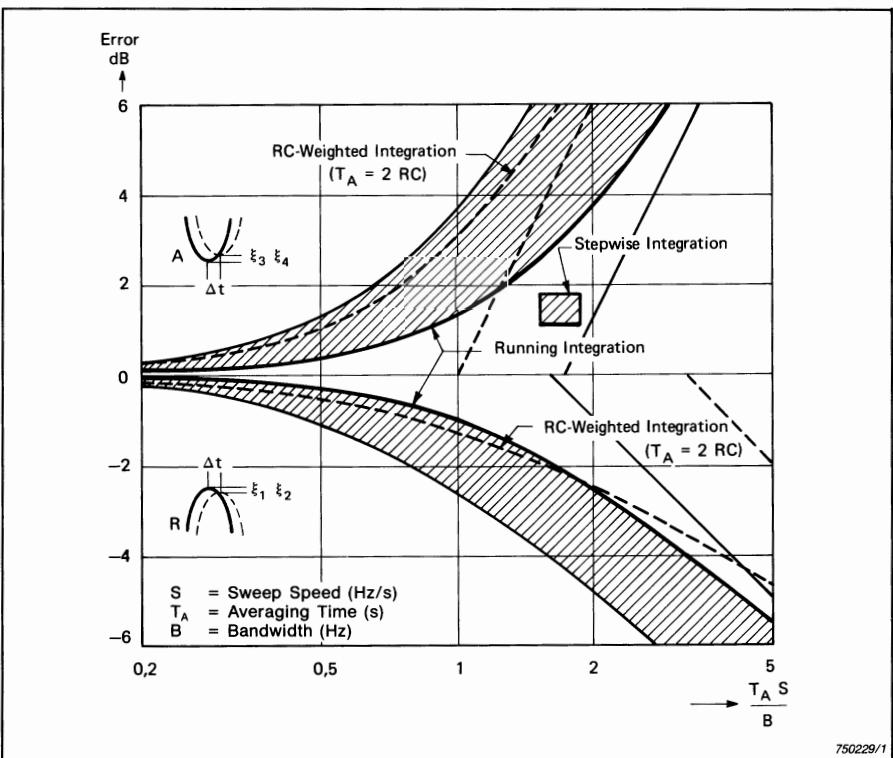


Fig. 3.20. Errors at peaks and valleys in a recorded spectrum

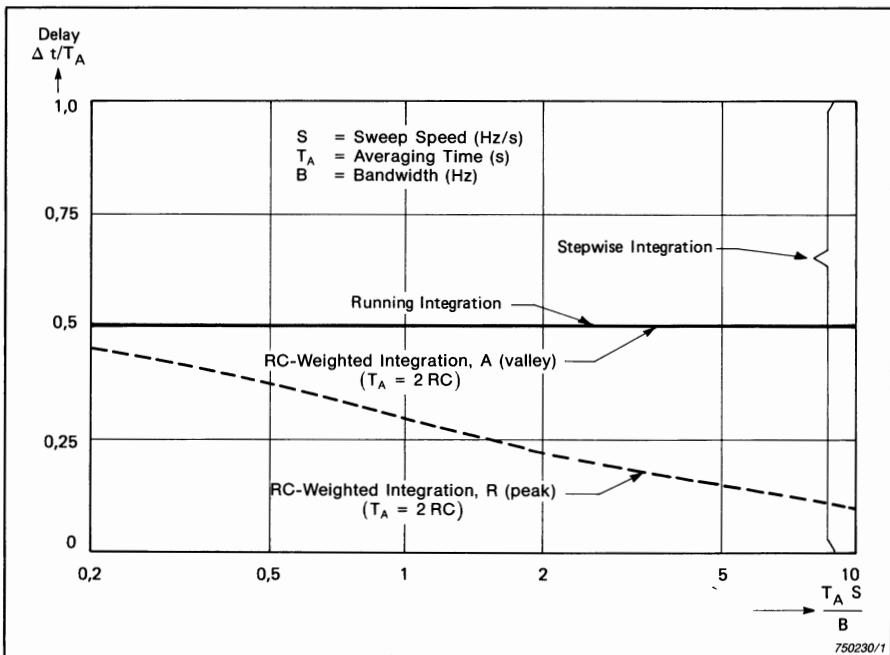


Fig. 3.21. Delays at peaks and valleys in a recorded spectrum

3.3. RECORDERS

The level recorder used to record the spectrum can be the factor which most limits the sweep speed but its influence is greatly affected by whether it is operating in "AC-recording" or "DC-recording" mode.

3.3.1. AC Recording

AC-recording is the case where the filtered AC signal is applied directly to the recorder input. With the recorder set to "RMS", the signal is rectified and squared as for a normal detector but the averaging results entirely from the response of the pen drive system which acts as a lowpass filter.

Fig.3.23 shows typical frequency response characteristics (Ref.3.2) for the 2307 Level Recorder (for 50 mm paper), which illustrate the fact that higher writing speeds give a higher cut-off frequency and thus a shorter averaging time as would be expected from the discussion of Section 3.2.2. However, the level recorder circuits have a non-linear characteristic because of feedback of the error signal and also because of velocity limiting circuits which come more and more into play the larger the fluctuations. Accordingly, the lowpass filter

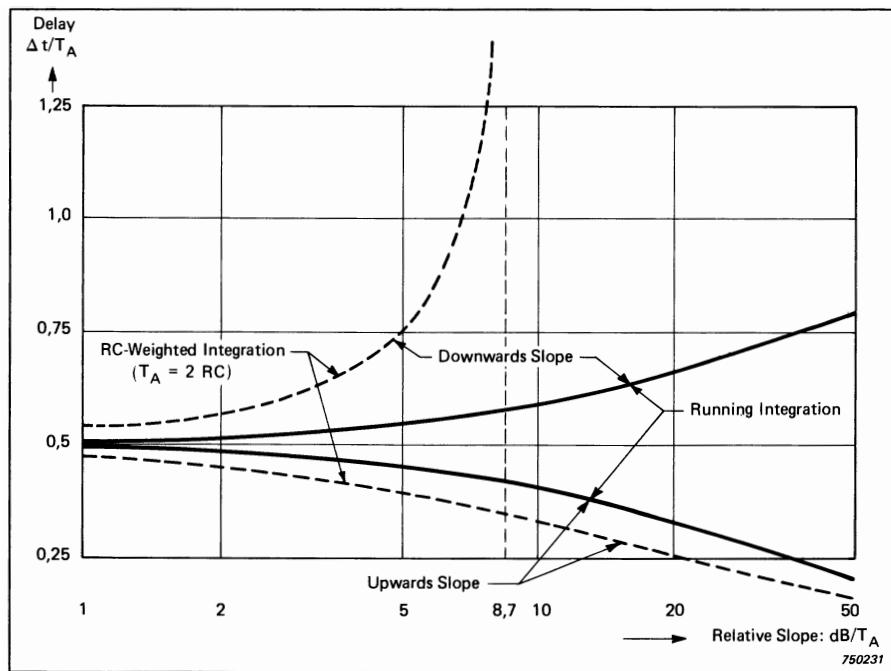


Fig. 3.22. Delays at sloping areas in a recorded spectrum

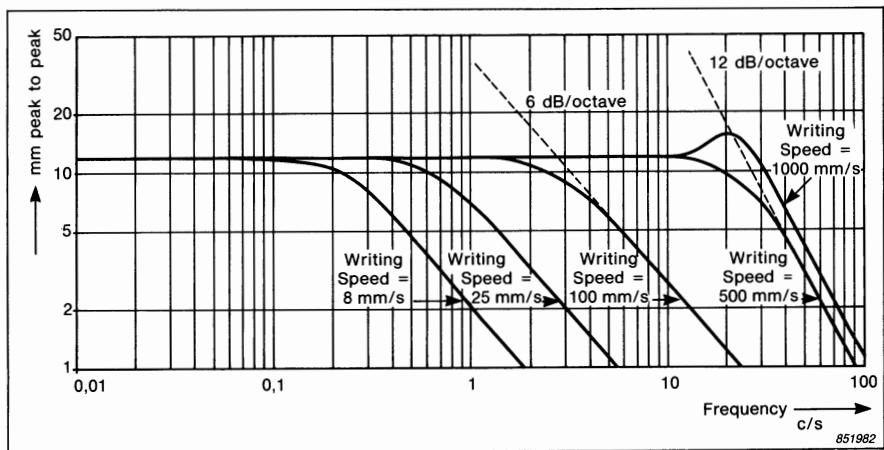


Fig. 3.23. Typical frequency response curves for "normal" setting of the Recorder control knob marked "Potentiometer Range dB". The curves were measured for a fixed maximum amplitude and different writing speeds

characteristics illustrated in Fig.3.23 are not general but apply to a particular set of conditions, and in general there is a range of averaging times which apply to a given set writing speed. This range is illustrated in Fig.3.24 which shows the relationship between averaging time and writing speed for the B & K Level Recorders Types 2305, 2307, 2306 and 2309 (i.e. the results are not identical for all of them, but all tend to fall in the range indicated by the shaded area). In general the lower values of averaging time apply for small fluctuations while the higher values apply to large fluctuations.

Thus, in order to obtain a given averaging time with AC-recording, Fig.3.24 may be used to select an appropriate writing speed, and then the rate of sweep may be limited by the error introduced on peaks and valleys or, more likely, by being able to reproduce the steepest part of the spectrum with a limited writing speed.

3.3.2. DC Recording

In DC recording, the RC-averaged RMS level is applied to the input of the recorder which is then operated in "DC" mode. This requires of course that the analyzer being used has a DC output with a sufficiently large dynamic range (i.e. > 50 dB) but this being the case it is generally better to use DC recording for the following two reasons:

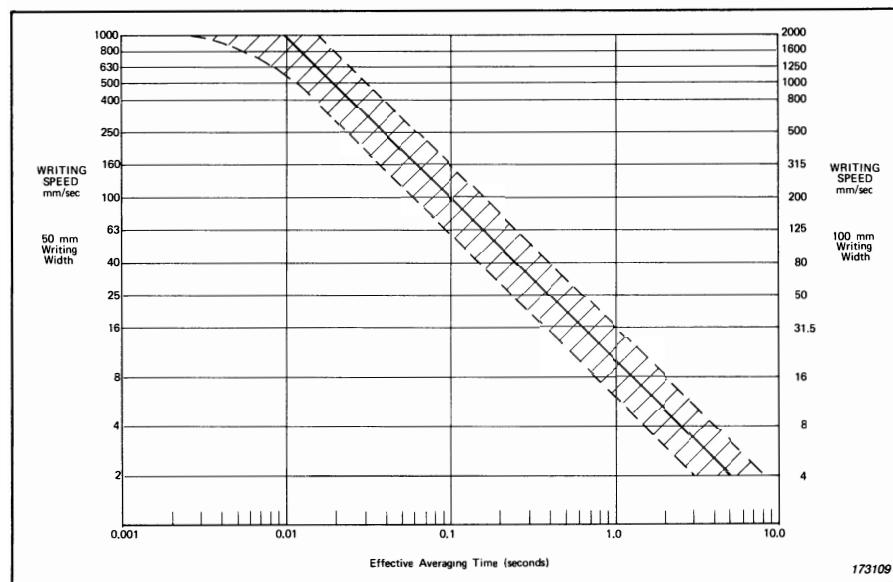


Fig. 3.24. Averaging time vs writing speed

- 1) The averaging time is much better specified
- 2) It is generally possible to analyze faster with DC recording

Note that even in DC recording the pen system will still add an extra delay corresponding to its inherent averaging time. However, with both 2305 and 2307 Level Recorders it is always possible to select a Writing Speed of 1000 mm/s (100 mm paper). From Fig. 3.24 it is seen that this corresponds to an averaging time of about 0,02 s and is thus negligibly small in comparison with the smallest averaging time available for DC recording (0,1 s). It is later shown (Section 3.6.2.7) that for this choice of writing speed the limitation on reproduction of slopes is also determined by the averaging time and not by the recorder.

For the small portable Level Recorders Types 2306 and 2309 the maximum Writing Speed is 250 mm/s (50 mm paper), but on the other hand, they are usually used in combination with detectors with a minimum averaging time of 0,25 s, and will still not limit sweep speeds in DC-recording.

For the newer portable recorder Type 2317, true RMS averaging detectors are built-in, and so the recording can always be treated in the same way as a DC recording with the appropriate averaging time.

3.4. ANALOG ANALYZER TYPES

Analog analyzers fall into the following broad categories:

- (1) Discrete stepped filter analyzers
- (2) Sweeping filter analyzers
- (3) Parallel analyzers (real-time)
- (4) Time compression analyzers (real-time)

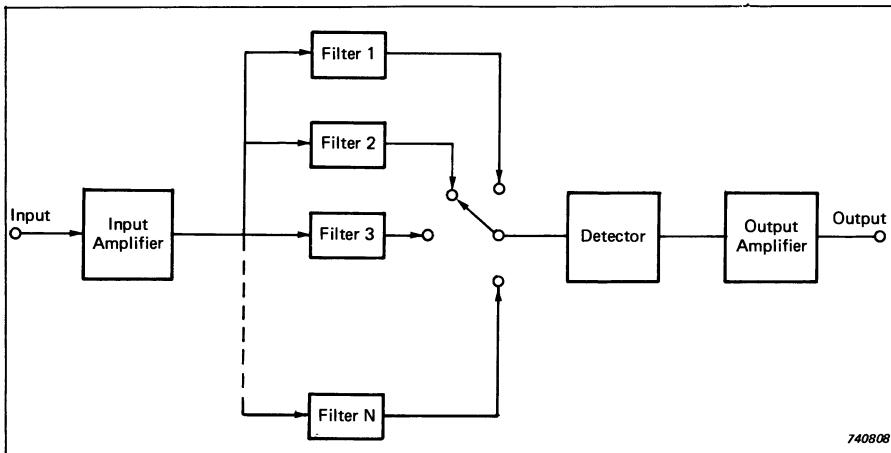
Time compression analyzers have been included, even though they make use of digital techniques, because the actual frequency analysis is carried out by analog (heterodyne) techniques.

The mode of operation and other main features of each of these types will now be discussed in more detail.

3.4.1. Discrete Stepped Filter Analyzers

Fig. 3.25 shows a simple block diagram of a typical fixed filter analyzer. The signal, after conditioning by an input amplifier, is applied in parallel to a bank of filters, contiguous in the frequency domain, which together cover the frequency

range of interest (Fig.3.26). A detector is connected sequentially to the various filter outputs, and thus successively measures the output power in each frequency band. Note that it is not necessary to wait for the filter response time, only that of the detector.



740808

Fig. 3.25. Stepped filter analyzer

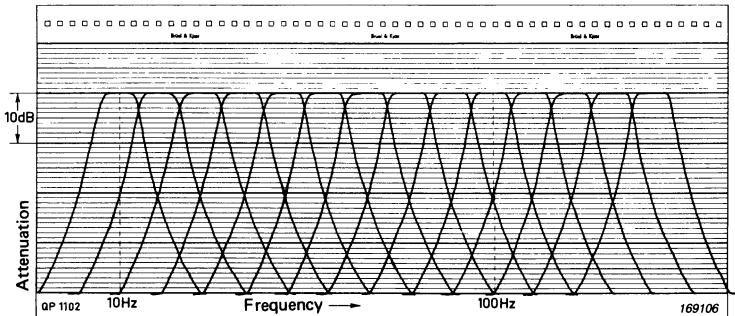


Fig. 3.26. Frequency response of several adjacent 1/3-octave filters

The rate of stepping through the various filters is often controlled by and synchronized with the speed of a connected level recorder, which can thus be used to record the resulting spectrum. A typical 1/3-octave spectrum obtained in this way is shown in Fig.3.27.

A more up-to-date way of achieving the same effect as a filter bank is to have only two filters whose centre frequencies can be changed by switching between different electronic components. These two filters can be designed to step up alternately in frequency in such a way that the signal is always being applied to the next filter in the series while measurements are being made with the current filter. This gives effectively the same result as a complete bank of parallel filters in the sense that every filter has had adequate time to respond before the detector is coupled in to its output.

3.4.2. Sweeping Filter Analyzers

For narrow-band analysis it is more common to use a single filter with tunable centre frequency, as illustrated in block diagram form in Fig.3.28. The filter can be either of constant bandwidth or constant percentage bandwidth type. Fig.3.29 shows an analysis obtained in this way on the same signal as in Fig.3.27, in the first place with the same bandwidth (23,1%) and then with 3%

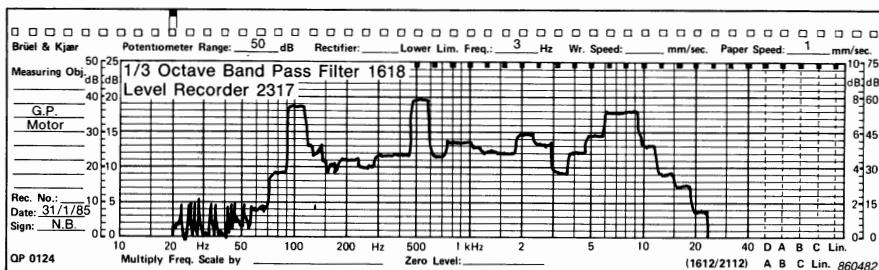


Fig. 3.27. Typical spectrum from a 1/3-octave filter set

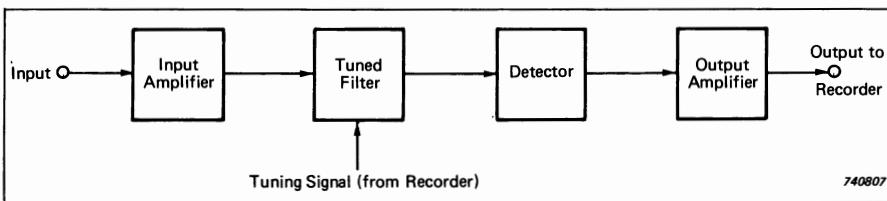


Fig. 3.28. Sweeping filter analyzer

bandwidth. Instead of a number of spectrum estimates at discrete frequencies, the resulting spectrum is now continuous in frequency. Each point on the curve, however, represents an integration of the true spectrum over a frequency range corresponding to the filter bandwidth. Considering only the positive frequency components in both signal spectrum and filter characteristic, it will be seen that the result is effectively a convolution of the two functions. (It may in fact be of assistance in the interpretation of convolution to think of it as a sweeping filtration of one function with the other reversed).

3.4.3. Real-time Parallel Analyzers

The two types of analyzers already discussed are known as sequential or serial analyzers, since the analysis is carried out sequentially at each frequency. Thus, the assumption is implicitly made that the signal being analyzed is stationary, because otherwise the analysis at one frequency would have no connection with the analysis at another frequency (which is made on a different section of the time signal). Sometimes a signal is forced to be stationary by recording a section of it on a tape loop (or equivalent) which is played back repetitively.

So-called real-time analyzers obtain the whole spectrum in parallel from the same section of signal, and are thus not only able to follow changing signals, but can also obtain the spectrum very much faster than sequential analyzers. Perhaps the most direct way of performing such a real-time analysis is simply to apply the signal to a parallel bank of filter/amplifier/detector channels as

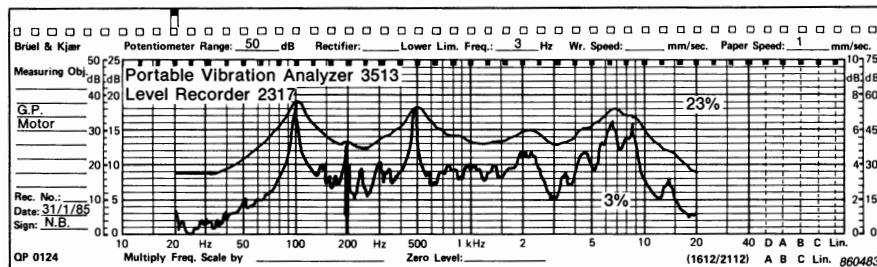


Fig. 3.29. Sweeping filter analysis for comparison with Fig. 3.27

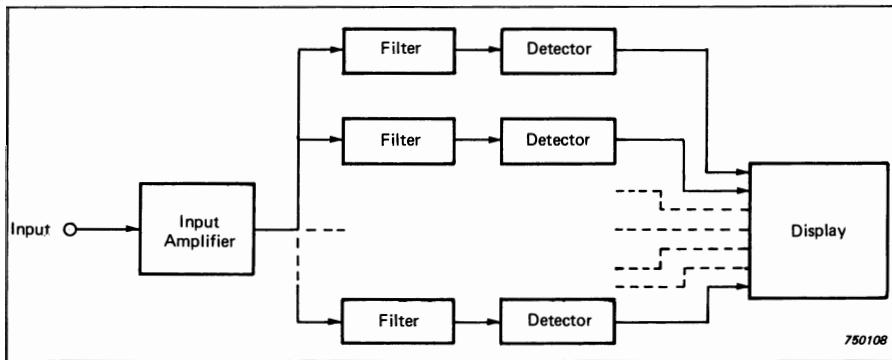


Fig. 3.30. Real-time parallel analyzer

illustrated in block diagram form in Fig.3.30. Such an analyzer is called a parallel analyzer and the first real-time 1/3-octave analyzers (e.g. B & K Type 3347) were made in this way. The speed with which the results are generated makes it desirable to be able to view them on a continuously updated screen, and also to be able to transfer them rapidly in digital form to a computer or other digital device. Fig.3.31 shows the B & K Analyzer Type 3347.

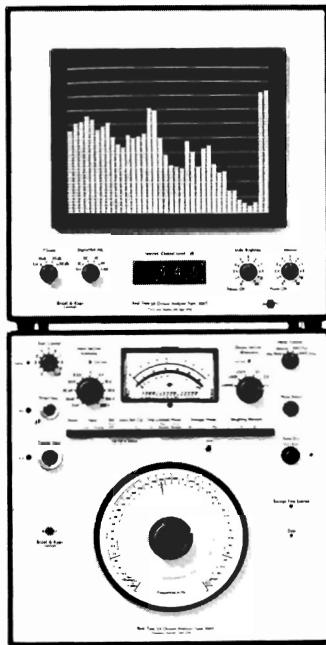


Fig. 3.31. Real-time 1/3-octave Analyzer Type 3347

The analog parallel analyzer has now been superseded by analyzers based on digital filtering, but its operation will be discussed where relevant since a number are still in service.

3.4.4. Time Compression Analyzers

Time compression analyzers were the first successful real-time narrow band analyzers, but have now been superseded by FFT analyzers (Chapter 4). The time compression principle can be explained as follows:

Recording a signal at one speed (for example on a tape recorder) and playing it back M times faster, allows the use of filters which are M times larger (for the same relative resolution) and a corresponding reduction in analysis time. With a tape recorder the signal must first be recorded in its entirety and then played back, but in a time compression analyzer, the signal is continuously digitised, and the samples entered into a digital memory, in between being played back at high speed and analyzed using a high frequency heterodyne filter.

Samples come from the analog-to-digital (A/D) converter at a rate determined by the sampling frequency (typically 3 times the frequency range), and are entered into the memory to replace the oldest values. With a 400-line analyzer (e.g. the now obsolete B & K Type 3348) the memory contains 1200 samples, of which 3 are replaced for every cycle (when operating at the highest real-time frequency). The digitised record is played back through a D/A converter at least 400 times faster than it was recorded and analyzed in a frequency range 400 times higher than the actual. One line out of the 400 is analyzed for each playback, and thus after all 400 lines have been analyzed the complete memory has been updated. For frequency ranges less than the maximum the analysis still goes on at the same rate, but the memory is replenished correspondingly slower, and there is a certain amount of redundancy in successive spectra.

In contrast to FFT analyzers (Chapter 4) where all frequency lines are obtained from the same data record, each line in a time compression analysis may be obtained from a slightly different record.

3.5. DIGITAL FILTERING

As an introduction to digital filtering, the example will be taken of a simple lowpass filter of the RC type such as is widely used for “exponential averaging”.

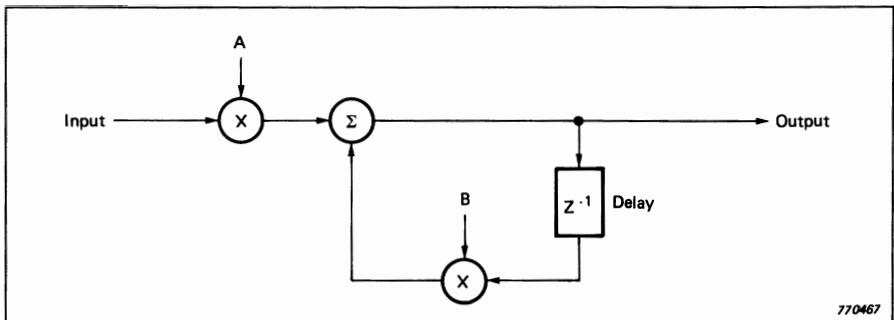
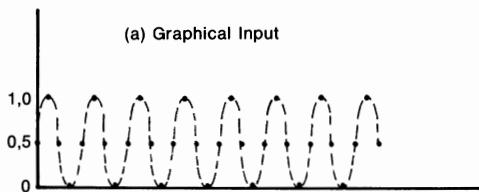


Fig. 3.32. Schematic diagram for a single-pole filter

Fig.3.32 is a schematic diagram of such an averager, illustrating how it can be made up of standard digital components, e.g. adders, multipliers, a delay unit and a read-only memory (ROM) for storing the constants A and B .

For each sample period, the new data sample entering the filter is multiplied by the constant A and added to B times the previous output value which has been stored in a delay unit. Consider the example where constant A is 0,1 and constant B is 0,9. In order not to change the amplification, the sum of A and B must be unity, so that in fact there is only one independent parameter with which the properties of the filter can be varied. This parameter can be related to the equivalent averaging time. Note that the averaging time is defined only in relation to the sampling period, which is the same as saying that the lowpass filter cut-off frequency is only defined in relation to the sampling frequency (the reciprocal of the sampling period).

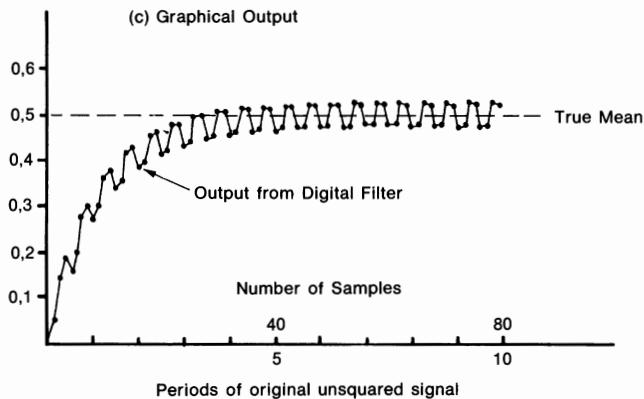
Fig.3.33 illustrates the effect of carrying out the operations of Fig.3.32 on a sampled squared sinusoidal signal. The input signal, varying between 0 and 1, could be considered to be the output from a squaring circuit of a sinusoidal signal of amplitude 1. It has somewhat fortuitously been sampled (4 times per period) at the values 0, 0,5 , 1,0 and 0,5 but since this obeys Shannon's Sampling Theorem (see p.30) no error will be introduced. It can for example be checked that the RMS level of the samples over one period is the same as that of the continuous sinusoid. Thus, the first sample (0,5) is multiplied by 0,1 and added to 90% of zero, giving 0,05. The next sample (1,0) is also multiplied by 0,1 and added to 90% of 0,05 giving 0,145 and so on. It is seen that there is first a period in which the filter output rises, but that finally the output value levels out though with a small fluctuation (see Fig.3.33). The correct average is of course 0,5 , but the output fluctuates about this with a ripple sampled at the four values 0,472, 0,475 , 0,525 and 0,528. These samples no longer correspond to the peaks as with the original signal (because of a phase shift) but it can be



(b) Tabular Input/Output

Input Data	0,5	1,0	0,5	0	0,5 repeated	1,0	0,5	0
Output Data	0,05 0,292 0,396 0,441	0,145 0,363 0,457 0,497	0,181 0,377 0,461 0,497	0,162 0,339 0,415 0,448	0,196 0,355 0,423 0,453	0,277 0,420 0,481 0,508	0,299 0,428 0,483 0,507	0,269 0,385 0,435 0,456
				.	.			
	0,475 0,475	0,528	0,525	0,472	0,475	0,528	0,525	0,472

T0107KG00



860611

Fig. 3.33. Effect of filter from Fig.3.32 for $A = 0,1$ and $B = 0,9$

calculated that the amplitude of the fluctuation is 0,0375 corresponding to $\pm 0,3$ dB. This agrees well with the result obtained from Fig.3.16 for the equivalent analog detector. The RC-time constant in this case is approx. 10 sample periods, giving an equivalent averaging time of 20 sample periods. The squared

sinusoid was sampled 4 times per period corresponding to 8 times per period for the original unsquared signal. Thus the averaging time corresponds to 2,5 periods of the original signal, and it is this which gives 0,3 dB ripple when inserted in Fig.3.16.

3.5.1. The Z-transform

Although is it not intended to go into details, it may be useful to explain some points of terminology which are often heard in connection with digital filters. One of these is the so-called z-transform.

The z-transform is defined (Ref.3.3) as:

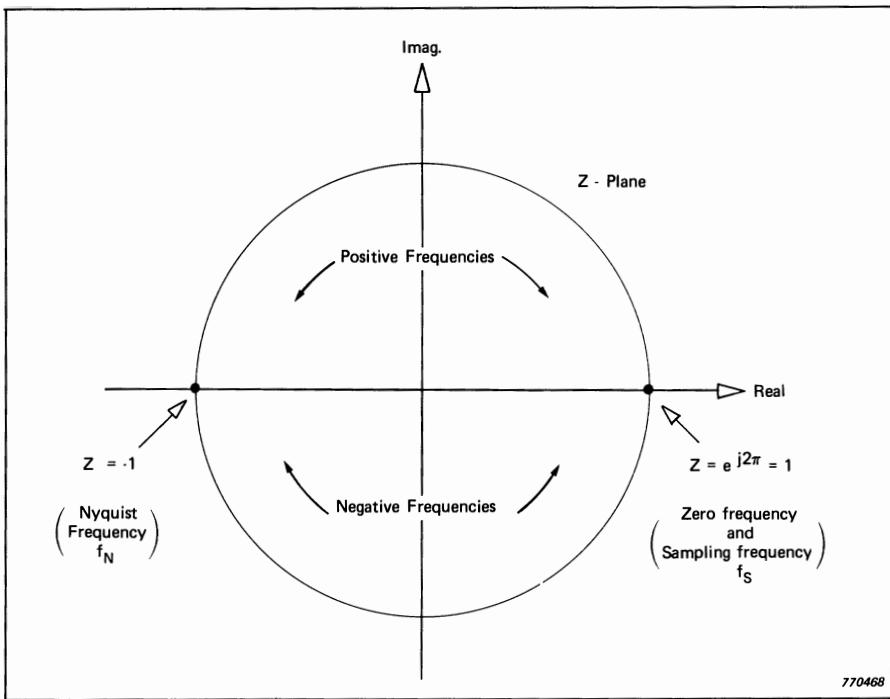
$$G(z) = \sum_{n=-\infty}^{\infty} g(n \Delta t) z^{-n} \quad (3.14)$$

where z is a complex variable.

By making the substitution $z = e^{j2\pi f \Delta t}$ (a circle of radius 1) it will be seen that Eqn.(3.14) reduces to that form of the Fourier transform which applies to sampled time functions, viz. Eqn.(2.16), and the latter can thus be considered as the z-transform evaluated on the unit circle (i.e. $|z| = 1$). Note that on the unit circle, the sampling frequency $f_s (= 1/\Delta t)$ gives $z = e^{j2\pi} = 1$, as do all multiples of it, which is another expression of the periodicity of the frequency spectrum alluded to in Fig.2.6 part 3. Fig.3.34 illustrates this, and shows for example how the Nyquist frequency f_N is located at $z = -1$; angles from 0 to π represent the frequencies 0 to f_N , while the angles from $-\pi$ ($=\pi$) around to 0 represent the frequencies from f_N to f_s , or equally, the negative frequencies from $-f_N$ to zero, because of the periodicity.

The z-transform bears the same kind of relationship to the discrete time Fourier transform (Eqn.(2.16)) as the Laplace transform does to the Fourier Integral Transform (Eqn.(2.14)), but applies to discrete time sequences rather than continuous functions. In particular, it has very similar properties with respect to difference equations as the Laplace transform does with respect to differential equations, in general reducing the solution of an N^{th} -order equation to finding the roots of an N^{th} -order polynomial in the transform domain (Ref.3.3).

Those familiar with Laplace transform methods will remember that a filter whose function is described by an N^{th} -order differential equation in general has N poles in the Laplace domain, each one corresponding to the root of an N^{th} -order polynomial which forms the denominator of the Laplace transform of the impulse response (i.e. the transfer function) (Appendix C). The same applies to the z-transform for the case of an N^{th} -order difference equation (which may be an approximation to the differential equation or a system characteristic in its own right).



770468

*Fig. 3.34. Frequencies in the z-plane
(conventional representation with real axis horizontal)*

It is of interest to see how a delay corresponds to an integration, and the simplest example of this is provided by the RC-lowpass filter already discussed. The differential equation applicable to its characteristic is:

$$\frac{dh(t)}{dt} = -\frac{h(t)}{RC} \quad (3.15)$$

which says that the rate of *fall* of the capacitor voltage is proportional to the voltage and inversely proportional to the RC time constant. As already given in Eqn.(3.7) the solution to this equation is:

$$h(t) = e^{-t/RC} \quad (3.16)$$

for unit input.

The simple (backward) difference equation corresponding to Eqn.(3.15) may be derived as:

$$\frac{\Delta h}{\Delta t} = \frac{-h}{RC}$$

or $\frac{h(n) - h(n-1)}{\Delta t} = -\frac{h(n-1)}{RC}$

giving $h(n) = h(n-1) \left[1 - \frac{\Delta t}{RC} \right]$ (3.17)

This shows how the value of $h(n)$ (i.e. the integration of dh/dt) can be obtained from the previous value $h(n-1)$ making use of the constant $\Delta t/RC$. Thus a single integration involves one delay.

3.5.2. The General Multiple-Pole Filter

The difference equation equivalent to a 2nd order differential equation involves two delays (i.e both $h(n-1)$ and $h(n-2)$ are required in order to calculate $h(n)$), and in fact the general flow diagram for a 2nd order system (2-pole filter) is as illustrated in Fig.3.35.

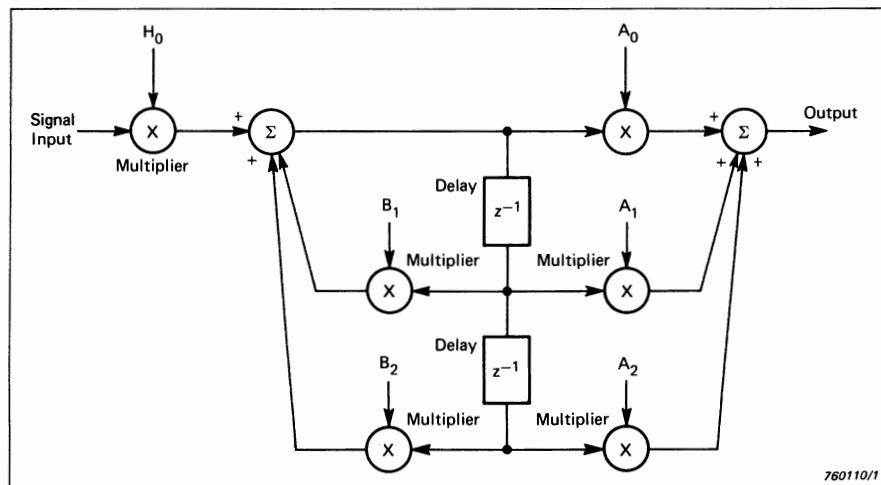


Fig. 3.35. Generalised block diagram of a 2-pole recursive digital filter

It is possible to generate digital equivalents of all the well-known filter types e.g. Butterworth, Chebyshev, etc. (Ref.3.4). Because of the periodicity of the frequency function, they can never be exactly the same, but by careful design the differences can be made negligible. Multiple-pole filters (generally required

to achieve steep filter flanks) can be formed by cascading 2-pole sections, with appropriate choice of coefficients.

One of the main advantages of a digital filter is that the same hardware can be used to generate virtually any filter shape (with the same number of poles) just by changing the filter coefficients used in the calculations. For example, it is possible to change the same calculating unit from a bandpass to a lowpass filter, or to change the centre frequency or bandwidth just by changing the filter coefficients. Once decided, the filter coefficients completely determine the filter properties, and for example these do not change with time; the filter never needs trimming.

Digital filters are best adapted to logarithmic frequency scales and constant percentage bandwidth, in contrast to FFT which intrinsically gives constant bandwidth on a linear frequency scale.

To illustrate these points, the design of the B & K Real-Time 1/3-octave Analyzer Type 2131 will be considered.

3.5.3. The Brüel & Kjær Digital Frequency Analyzer Type 2131

The Digital Frequency Analyzer Type 2131 is depicted in Fig.3.36. Although considerably more compact than the analog analyzer it replaced (the B & K Type 3347, see Fig.3.31), it has basically the same function, viz. producing averaged 1/3-octave spectra in real-time in the frequency range up to 20 kHz. Its major advantages, however, are that it has a larger dynamic range (display range 60 dB instead of 50 dB), lower limiting frequency 1,6 Hz as standard (instead of 20 Hz), a choice between octave and 1/3-octave bandwidth, a choice between linear and exponential averaging (instead of exponential only) and a choice between a range of constant averaging times and constant BT_A products giving constant statistical confidence (standard deviation of the error $< 0,5$, 1 or 2 dB). There are other advantages, but the above are those which derive from its construction as a digital filter, and it is that which is of interest here. The way in which the advantages arise will become obvious from a consideration of its mode of operation.

Before looking at the overall construction of the analyzer, it is of interest to consider the design of the individual filter units. In place of the general 2-pole diagram of Fig.3.35, they have a special schematic diagram as illustrated in Fig.3.37. This has the obvious advantage that fewer adders and multipliers are required, while by making use of the so-called "matched z-transform" it is still possible to design equivalents of the well-known bandpass and lowpass filter types. It will be appreciated that the three coefficients remaining provide the required flexibility to determine both the resonant frequency and Q-factor for a damped resonator, and to vary the overall amplification.

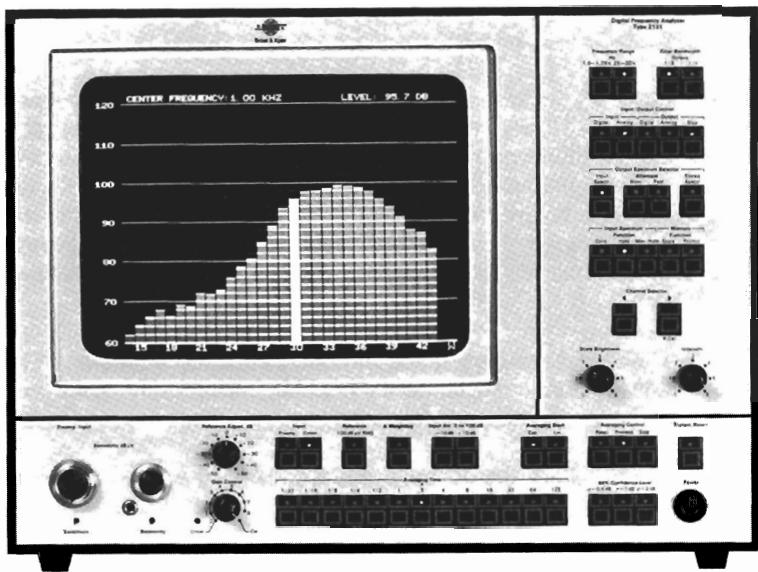


Fig. 3.36. The B & K Digital Frequency Analyzer Type 2131

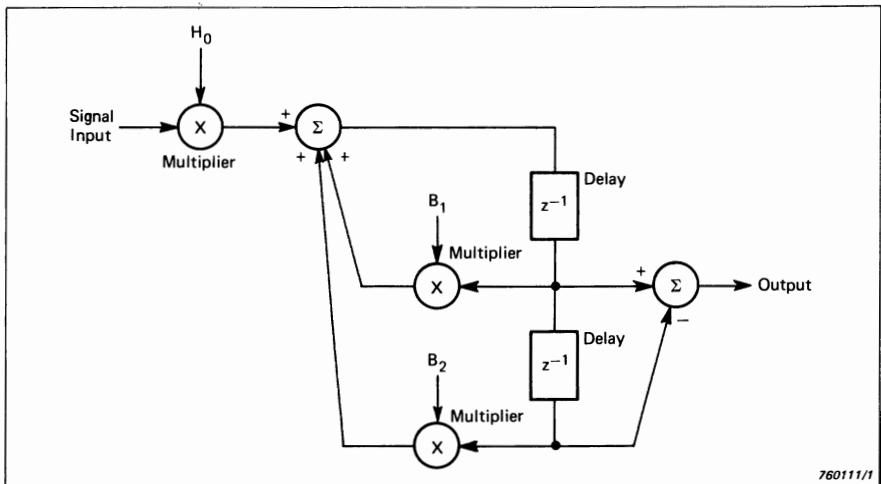


Fig. 3.37. Block diagram of 2-pole digital filter used in the 2131

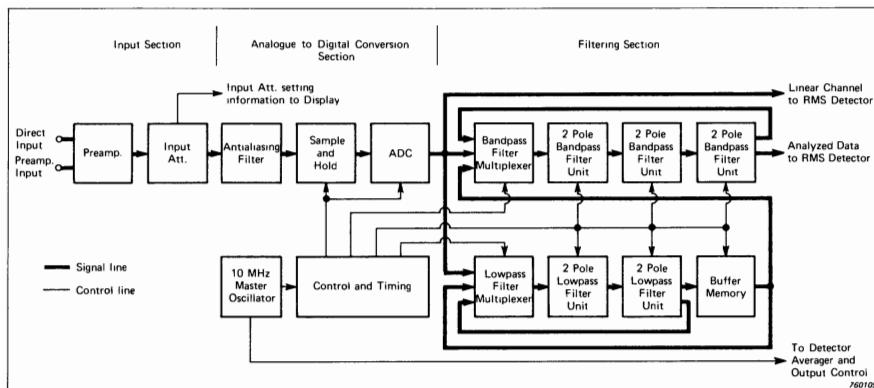


Fig. 3.38. Block diagram of input and filter section of 2131

Fig.3.38 is a block diagram of the input and filter section of the 2131. It will be seen that the signal, shortly after entry, is converted into digital form and from then on all operations are digital. Before analog-to-digital (A/D) conversion, the signal is first lowpass filtered with a 12-pole analog lowpass filter having its cut-off at 27 kHz which is above the highest frequency of interest, viz. 22,4 kHz (the upper limiting frequency of the 1/3-octave centred on 20 kHz). This is done to avoid aliasing (section 2.2.4). The A/D converter used gives 12-bit resolution, and this in combination with 15-bit calculations and 9-bit filter coefficients, gives a resulting spectrum with more than 69 dB dynamic range. The sample rate of the A/D converter is 66,667 kHz.

Each sample coming from the A/D converter is passed simultaneously through a 1/3-octave bandpass filtering section and a lowpass filtering section. In fact, each sample is passed through each section three times for the following reasons:

- 1. 1/3-octave filtering.** The 1/3-octave filter section consists of three 2-pole filter units in series and for each pass, coefficients are used which give a 6-pole Chebyshev filter of 1/3-octave bandwidth. For each pass, the filter coefficients are changed so as to obtain successively the three 1/3-octave centre frequencies in each octave (e.g. 20 kHz, 16 kHz and 12,5 kHz in the highest octave). These three filter characteristics are illustrated in Fig.3.39.
- 2. Low-pass filtering.** The low-pass filter section consists of two 2-pole filter units in series. Thus, during the three passes used to obtain the three 1/3-octave filtered values, it is possible to circulate the data value three times through the lowpass filter section, achieving 12-pole lowpass filtration (in this case, incidentally, a Butterworth filter was used). The cut-off frequency

of the lowpass filter is one octave lower than the previous maximum frequency content.

The reason for the lowpass filtration is that it makes it possible to discard every second sample without losing any further information, i.e. once the highest octave in frequency is filtered away it is quite valid to use half the previous sample rate while still complying with Shannon's sampling theorem (see footnote p.30). These lowpass filtered samples with half the sampling frequency can now be fed to the bandpass filter section, and since the filter characteristics are defined only in relation to the sampling frequency, the same filter coefficients will now give the three 1/3-octave filters one octave lower in frequency (see Fig.3.39).

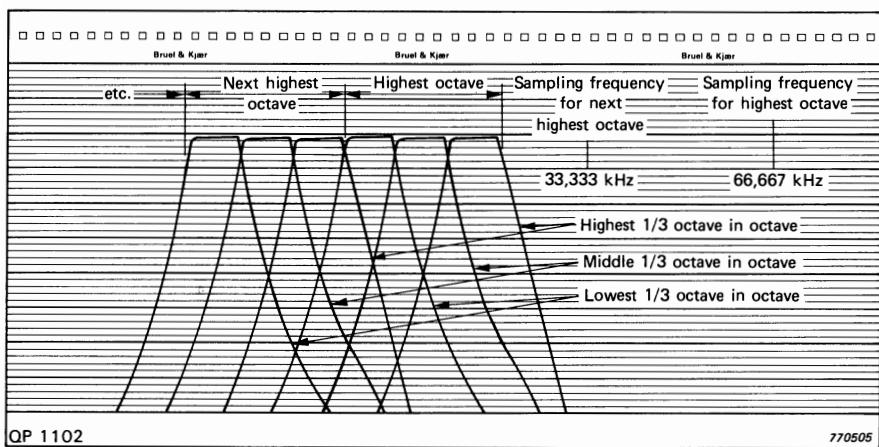


Fig. 3.39. Filter characteristics vs. sampling frequency

In a similar manner, the same filtered samples can be fed back to the lowpass filter section and again filtered to one octave lower, once again allowing each second sample to be discarded, and so on.

It is possible to continue in this way for all lower octaves, thus obtaining the complete 1/3-octave spectrum. This explains the presence of the multiplexers at entry to both filter sections. These must keep track of where the next sample to be filtered is located and where the result is to be placed. It will be found that provided it is possible to process a sample from the A/D converter (i.e. in the highest frequency octave) plus one other sample in each sample period, it is possible to produce a parallel real-time spectrum for all octaves up to and including the highest. The limitation at the low frequency end is not in calculating capacity, only in being able to store the results, and so in the 2131 the

frequency range is limited to a little over 4 decades. The reason why this is possible can be understood by reference to Fig.3.40. This is a table showing the order of processing, at least of the first few samples in each octave, both in the 1/3-octave section and in the lowpass filter section. The lowpass filtered values (with every second one discarded) are fed back to the next lower octave in both filter sections. Considering the number of samples to be processed in each octave, and calling the number in the 16 kHz octave M , the number in the 8 kHz octave is $M/2$, in the 4 kHz octave $M/4$, and so on. Thus the total number of samples to be processed in all octaves below the highest is $M(1/2 + 1/4 + 1/8 + \dots) = M$, i.e. the same as the number of samples in the highest octave alone. Consequently, as shown in Fig.3.40, it is only necessary to process one data value from the highest octave plus one other in each sample period.

This ability to timeshare efficiently when the frequency scale is based on octaves is the major reason why digital filters are so well adapted to logarithmic frequency scales and constant percentage bandwidth. Another results from the fact that the filter characteristics are relative to the sampling frequency and thus tend to be symmetrical and uniform on a logarithmic frequency scale, a desirable property in constant percentage bandwidth filters. This question of filter characteristic is discussed in some detail in section 3.5.4 since it represents one of the major differences from spectra generated originally by FFT techniques and then converted to constant percentage bandwidth.

The operation of the analyzer can be converted to octave band filtering, once again basically by changing the filter coefficients. Since only one filter is to be calculated in each octave it is possible to recirculate the data values more than

BANDPASS FILTER																
B.P Filter Octave	SAMPLING PERIOD NUMBER															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
16 kHz	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
8 kHz	11				51		71		91		111		131		151	
4 kHz		12				52				92				132		
2 kHz				13							93					
1 kHz								14								
500 kHz																15

LOWPASS FILTER																
L.P.Filter Cut-Off Freq.																
12 kHz	11	21	31	41	51	61	71	81	91	101	111	121	131	141	151	161
6 kHz	12		32		52		72		92		112		132		152	
3 kHz		13			14					93		94				
1.5 kHz								15								
750 kHz																16
375 kHz																

Fig. 3.40. Operation of 2131 digital filter unit

once through the bandpass filtering section. In the 2131, two passes are used (3 would have been possible) giving 12-pole filters of the Chebyshev type.

So far, only the actual filtering process has been considered, but it is important to consider the averaging as well, since many of the advantages of digital filters over competitive analyzers stem from the correct use of digital averaging (correct in the sense that the sampling theorem is obeyed, and thus no information is lost).

Linear averaging is achieved simply by adding together the squares of the filter output samples[§] divided by the total number to be averaged. This number is of course equal to the averaging time multiplied by the sample rate for the octave in question. The averaged result is held and is logarithmically converted into dB before display.

Exponential averaging is achieved as discussed in the early part of Section 3.5. The equivalent averaging time is based on the Eqn.(3.8) (i.e. $T_A = 2RC$) and thus applies directly to stationary signals. As for linear averaging it is the squared values of the filter output samples which are averaged, and the results are converted to dB before display.

Averaging with **constant standard error**, which is available with exponential averaging only, is achieved by making the averaging time in each octave inversely proportional to the frequency so that a constant BT_A product is obtained for all frequencies (at least for the lowest 1/3-octave filter in each octave; the two higher filters have the same averaging time and thus somewhat higher BT_A product). Since for constant percentage bandwidth a constant BT_A product corresponds to a given number of periods (cf. Eqn.(3.2)) in this case it will correspond to a given number of samples (for constant sampling ratio i.e. ratio of sampling frequency to filter centre frequency), and thus the *number of samples* averaged will be the same in all octaves. Three values of BT_A product are selectable giving standard error $\epsilon < 0.5, 1$ or 2 dB as desired (cf. Eqn.(3.9)).

3.5.3.1 1/12-octave Analysis

As an example of the flexibility of a digital filter such as the B & K Type 2131, the case of its use for 1/12-octave filtering will be discussed.

As mentioned previously, the filter characteristic obtained from a digital filter is determined entirely by the filter coefficients used in the calculations. These can be chosen to give 1/12-octave bandwidth, and in fact in the 2131 these extra coefficients are programmed as standard in the ROMs used for storing

[§]squared values are represented by 24 bits and sufficient bits are used in the calculations to avoid overflow

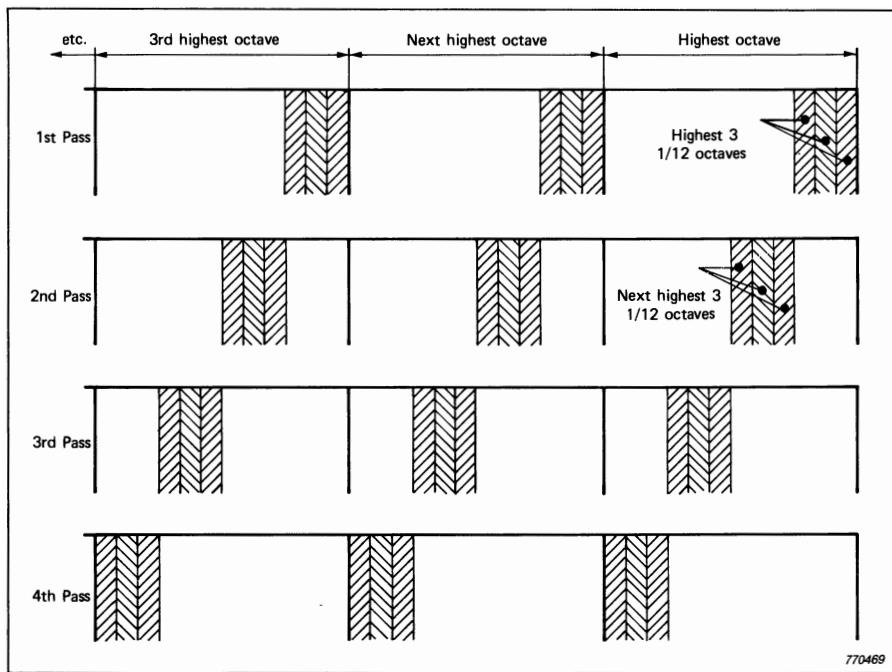


Fig. 3.41. Procedure for obtaining 1/12-octave spectrum in 4 passes

the coefficients (the storage space would otherwise have been unused). However, since there is only calculation and storage capacity to calculate three filters per octave at any one time, it is necessary to perform four separate analyses, each time accessing three sets of filter coefficients corresponding to three of the 1/12-octave passbands in each octave (see Fig.3.41). As will be evident from Fig.3.41, the results from the 4 passes must be interleaved in the correct order, and this is most easily done by a small computer. Fig.3.42 shows such a 1/12-octave spectrum produced by the 2131 in conjunction with a Tektronix Calculator (connected via the IEEE 488 Interface Bus). Not only can the calculator control the changing of the coefficients for each pass, it also stores all intermediate results and finally outputs the total 1/12-octave spectrum (sorted into the correct order) on its own display screen with correctly calibrated axes. Fig.3.42 was obtained as a direct copy of this displayed spectrum.

It should be emphasized that because of the necessity to make 4 passes, the operation of calculating 1/12-octave spectra is no longer real-time, and thus either the input signal must be stationary or exactly the same data must be recirculated 4 times.

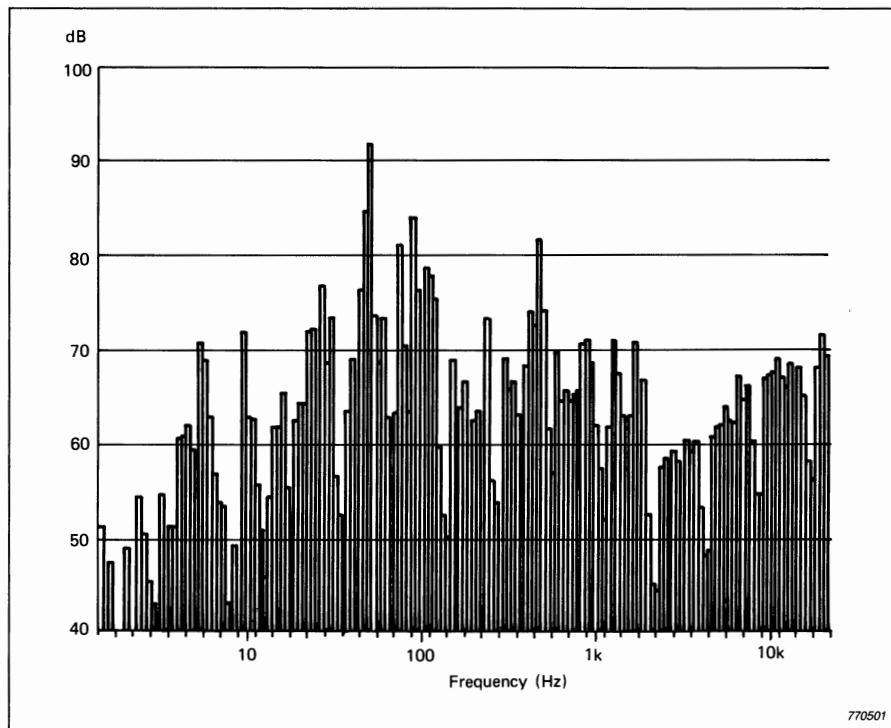


Fig. 3.42. 1/12-octave spectrum produced by 2131 in conjunction with a Tektronix calculator

3.5.4. Filter Characteristics

In general, digital filters are designed to resemble one of the wellknown classes of analog filters, e.g. Butterworth, Chebyshev, etc., but as mentioned previously there will always be some slight deviation for the following reasons:

1. Because of the periodicity of the frequency characteristics, there is some overlapping around the Nyquist frequency (aliasing) as illustrated in Fig.3.43. There is also a slight distortion of the characteristic (when the matched z-transform is used) but by careful design this can be made negligible. The overlapping around the Nyquist frequency can obviously be reduced by increasing the "sampling ratio" and for example in the 2131, it is only the highest 1/3-octave filter in each octave which is noticeably affected by folding. On the other hand, increasing the sampling ratio is undesirable since it gives a proportionate reduction in the real-time frequency capability and moreover if carried too far gives rise to stability problems. With the

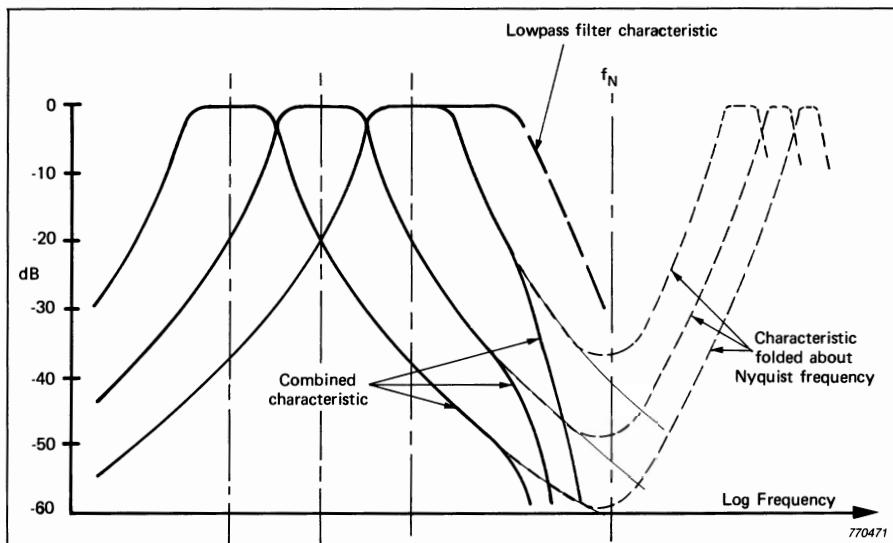


Fig. 3.43. Overall filter characteristics including lowpass filtration

2131, the sampling ratio varies from 3,3 for the highest 1/3-octave in each octave to 5,3 for the lowest.

2. The folding is counteracted by the lowpass filter characteristic of the anti-aliasing filters, which thus have an influence on the overall characteristics. The lowpass filter is the same, however, for all three 1/3-octave filters in each octave and thus has a different effect on their filter characteristics (see Fig.3.43).

It must be kept in mind that the deviations referred to above are very minor in nature, and in fact are only detectable from about -30 dB with respect to the reference level in the passband. The filters are well within the requirements of the most stringent standard specifications (e.g. Fig.3.46).

In this connection, it is worth making comparisons with another common method used to obtain constant percentage bandwidth filtration by digital means. This involves converting from a linear frequency scale constant bandwidth spectrum (obtained by FFT methods) by integrating over the appropriate number (and fractions) of lines in the spectrum. The integrated bandwidth must always be greater than that of the original spectrum for the result to be valid. With respect to filter characteristic, it will be seen from Fig.3.44 that because the original filter characteristic is symmetrical on a linear scale, the integrated characteristics will be as well, and thus will be unsymmetrical on a logarithmic

frequency scale. Moreover, over one decade the relative steepness of the flanks of filters obtained in this way (i.e. effectively the shape factor) will vary by a factor of 10:1. It is quite common to convert one decade at a time, and thus where two successive decades are fitted together there will be a sudden change of 10:1 in filter flank steepness (see Fig.3.45) even though both might be within the tolerances specified for a particular filter class (see for example Fig.3.46) which in reality are often quite wide.

It will be appreciated that this sort of variation in filter characteristic is at

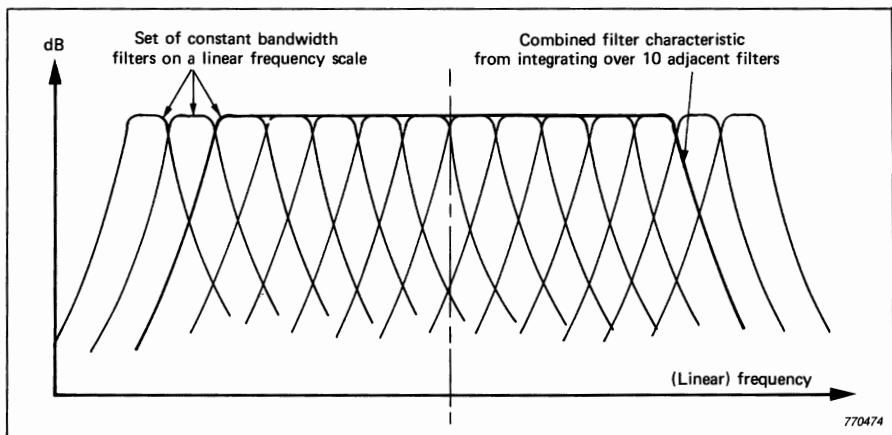


Fig. 3.44. Effect on filter characteristic of combining filters

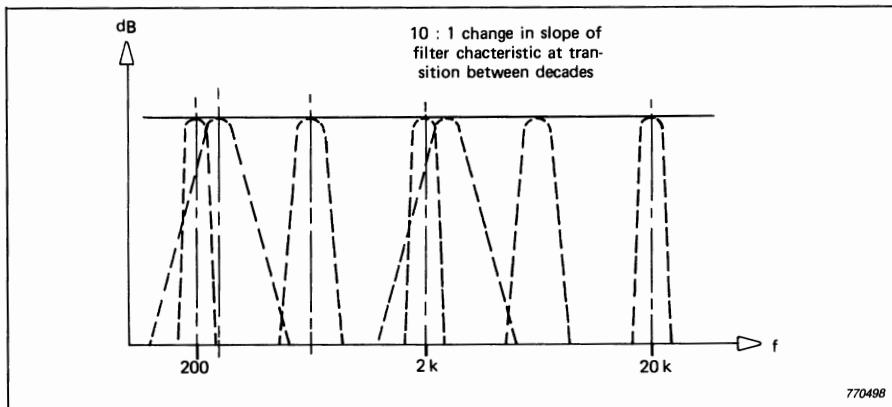


Fig. 3.45. Variations in filter characteristic with synthesized constant percentage bandwidth filters

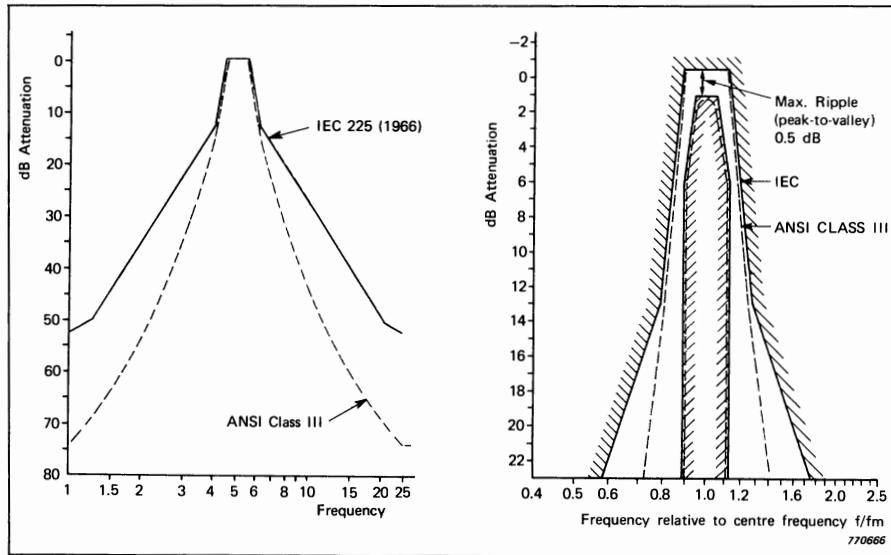


Fig. 3.46. 1/3-Octave filter specifications over 20 dB and 80 dB ranges

least an order of magnitude greater than that discussed in connection with digital filters, and is undesirable in cases where the filter flanks can have an effect on the results, e.g. for steeply sloping spectra.

The differences can be made considerably smaller by synthesizing FFT spectra obtained with intervals of one octave (rather than one decade) in frequency, and this procedure is a viable alternative (to digital filtering).

3.6. PRACTICAL ANALYSIS OF STATIONARY SIGNALS

The analysis of stationary signals, both deterministic and random, will now be discussed with respect to choice of analysis parameters for each of the types of analyzer discussed in Sections 3.4 and 3.5.

3.6.1. Stepped Filter Analysis (1/3 octave and octave)

In this case, the analysis bandwidth is determined by the analyzer, and the only choice to be made is between 1/3 octave and octave. The discussion here will refer to 1/3 octave unless otherwise specified.

The parameters to be chosen are the following:

1. Averaging time, T_A
2. DC or AC recording
3. Recorder writing speed
4. Recorder paper speed

Unless otherwise specified, it will be assumed that B & K standard calibrated paper of width 100 mm is used, on which 1/3 octave corresponds to 5 mm in the longitudinal direction, and that the recorder is a B & K Level Recorder Type 2307 or 2305.

3.6.1.1. Averaging Time T_A

For **deterministic** signals, the major requirement is that ripple be reduced to an acceptable level. Where only a single sinusoidal component is included in the bandwidth, Fig.3.16 indicates that the ripple will be reduced to less than $\pm 1/4$ dB for

$$T_A \geq \frac{3}{f} \quad (3.18)$$

where f is the frequency of the single component. Where there are several sinusoidal components within the bandwidth (and this is quite likely with 1/3-octave analysis) then Eqn.3.18 can still be used, with f interpreted as the minimum separation of any two components, since this will represent the lowest beat frequency[§]. When there are a large number of sinusoidal components in the bandwidth such as is likely at high frequencies, then strictly speaking the same considerations still apply but the situation becomes similar to the case of random signals and it might be simplest to treat the signal as random.

For **random** signals, Eqn.(3.9) can be used to select a suitable value for T_A , depending on the desired accuracy. Table 3.1 summarizes the value of standard deviation (in dB) obtained over a range of averaging times and frequencies. It will normally be the lowest frequency of interest which governs the selection of averaging time, but it will be noted that for every half decade increase in frequency it is possible to reduce the averaging time by a factor of $\sqrt{10}$, the normal steps in B & K measuring amplifiers. The newer filter sets can control the changes in averaging time automatically.

For **octave-band** filters, the bandwidth is of course 3 times larger, so that averaging times from Table 3.1 should be reduced by a factor of 3 (one step).

[§]The beat frequency will not be doubled by squaring (as for a single component) and therefore it may be found necessary on inspection to increase T_A even further

Std error for T_A	Centre Frequency (Hz)								
	2	6,3	20	63	200	630	2k	6,3k	20k
0,1 s						1,2	0,7	0,4	0,2
0,3					1,2	0,7	0,4	0,2	
1				1,2	0,7	0,4	0,2		
3			1,2	0,7	0,4	0,2			
10		1,2	0,7	0,4	0,2				
30	1,2	0,7	0,4	0,2					
100	0,7	0,4	0,2						
300	0,4	0,2							

T01038GB0

Table 3.1. Standard error (dB) for 1/3-octave filters in combination with various averaging times

3.6.1.2. DC or AC Recording

As mentioned in Section 3.3.2 it is generally better to use DC recording where this is possible, but where no DC output is available from the analyzer (or where it has a dynamic range less than 50 dB) it is necessary to use AC recording.

Where one has a choice, it may be worth checking both possibilities with respect to paper speed as in some cases AC recording is just as fast, and may be preferable for another reason (e.g. to obtain averaging times less than 0,1 s).

3.6.1.3. Writing Speed

For **DC recording** it is possible to select a writing speed of 1000 mm/s (100 mm paper) and this is always advisable with the Level Recorder Type 2307. With the Level Recorder Type 2305 it may be preferred to use a lower writing speed (to reduce hum which results from the low chopper frequency). The effect of using 315 mm/s will be very small, even with the lowest averaging time of 0,1 s.

For **AC recording**, the writing speed will be determined by the required averaging time, and can be selected using Fig.3.24.

3.6.1.4. Paper Speed

The selection of a suitable paper speed will be described for the case where it is the level recorder which controls the rate of stepping between filters. Newer filter sets control the rate of stepping in proportion to the averaging time which is changed automatically, but because the operation in that case is virtually

automatic, it does not require operator intervention.

In the first-mentioned case, the paper speed should be chosen such that the pen can attain the correct level for each 1/3-octave band and remain at that level for a while before shifting to the next filter (see Fig.3.47).

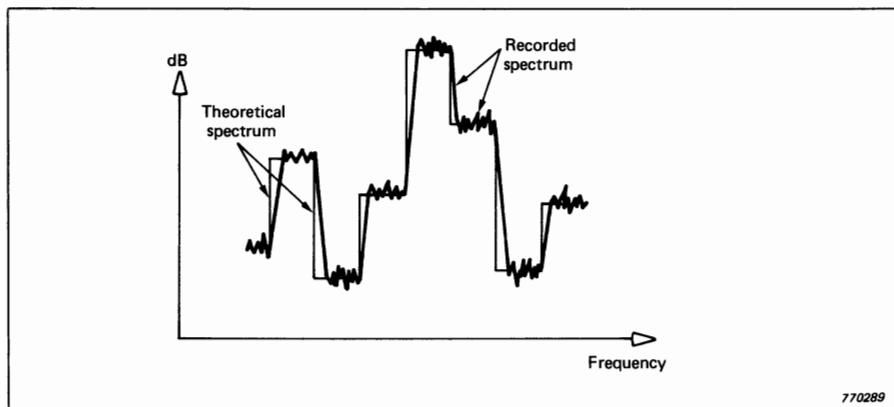


Fig 3.47. Typical recorded spectrum

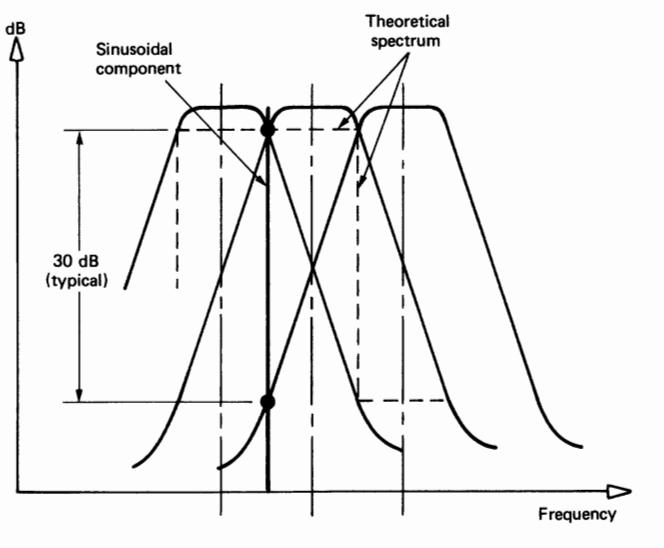
For **DC recording**, it will normally be the rate of fall of the detector which limits the sweep rate, and from Fig.3.22 of Section 3.2.5 it will be remembered that this has a maximum of 8,7 dB/averaging time.

Table 3.2 gives suitable paper speeds for the normal range of averaging times, based on the assumption that the time required to sweep over a bandwidth is equal to 5 averaging times. With B & K's steepest filter characteristics this means that the step between filters will normally be attained in less than 2 mm (i.e. 40% of the bandwidth), and in the very worst case of a single sinusoidal component located between two filters, within 75% of the bandwidth. This worst case situation is illustrated in Fig.3.48.

Averaging Time T_A (s)	0,1	0,3	1	3	10	30	100	300
Paper speed for DC recording (mm/s)	10	3	1	0,3	0,1	0,03	0,01	0,003

T01043GB0

Table 3.2. Paper Speeds for DC recording



770290

Fig 3.48. "Worst-case" situation involving the largest possible step between adjacent filters

In cases where the spectrum is known to be relatively flat, with a maximum fall of approx. 5 dB between adjacent filters, it will be possible to increase paper speed by one step (i.e. a factor of approx. 3). The time taken to sweep over one bandwidth will then be approx. 1,6 times the averaging time, and this is about the minimum time required for the detector to respond, even on peaks.

For **AC recording**, the same basic considerations apply, except that now it is the writing speed rather than the averaging time which will determine how rapidly the recording pen can adjust to a new level. Table 3.3 gives suitable paper speeds for the range of writing speeds available. Once again these may be increased by one step when the spectrum is known to be relatively flat.

Writing Speed (mm/s) (100 mm paper)	4	8 — 16	31,5 — 50	80 — 125	200 — 2000
Paper Speed for AC recording (mm/s)	0,1	0,3	1	3	10

T01044GB0

Table 3.3. Paper Speeds for AC recording

For **octave-band** analysis, with either DC or AC recording, it is generally best to use the same values as for 1/3 octave, because the filter flank steepness is often the same.

There is a chance, however, that the values recommended here can be increased by one step (in particular for octave analysis) and it is one of the features of stepped frequency analysis that this point can be checked by visual observation of the spectrum. This is not the case in general with swept frequency analysis.

3.6.2. Swept Filter Analysis

In this case there are several more factors to take account of, and in the most general case it will be necessary to consider the following:

- (1) Constant vs. constant percentage bandwidth
- (2) Linear vs. logarithmic frequency scale
- (3) The actual bandwidth, in Hz or percentage
- (4) Averaging time
- (5) Dwell time per bandwidth (which determines the sweep rate)
- (6) Recorder writing speed
- (7) Recorder paper speed

These will each be discussed in turn.

3.6.2.1. Constant vs. Constant Percentage Bandwidth

From the discussion in Section 3.1 it will be appreciated that constant bandwidth is most appropriate in the following situations:

1. Where the frequency scale is linear.
2. Where the spectrum is dominated by harmonics or other discrete components (i.e. deterministic signals).
3. Where comparisons are to be made with inherently constant bandwidth analysis (e.g. FFT or time compression).

and that constant percentage bandwidth is most appropriate in the following situations:

1. Where the frequency scale is logarithmic.
2. Where the signal is stationary random and the spectrum dominated by resonance peaks. Note that bandwidth compensation will be required if the result is to be calibrated as a spectral density. See Section 3.6.2.9. This compensation may be included in the analyzer as a power attenuation factor proportional to the bandwidth.
3. In order to account simply for fluctuations in machine speed, either within one record or between several records.
4. Where constant percentage bandwidth is specified, as is often the case with, for example, acoustic measurements or whole body vibration.
5. Where comparisons are to be made with inherently constant percentage bandwidth analyses (e.g. 1/3-octave).

In some cases it is possible to approximate constant percentage bandwidth by a series of constant bandwidth filters which step up automatically, for example, by a factor of $\sqrt{10}$ for every half decade in frequency. Note that in a broadband spectrum this would result in a sudden increase in spectrum level of 5 dB for every filter change, but that the step can be eliminated by using bandwidth compensation. At the same time this would automatically give a result in terms of spectral density. It is necessary, however, that the analyzer bandwidth is always less than that of any peaks in the signal spectrum, in order for the result to be valid.

3.6.2.2. Linear vs. Logarithmic Frequency Scale

From the discussion in Section 3.1 it will be appreciated that a linear frequency scale is most appropriate in the following situations:

1. Where the spectrum is dominated by harmonically related or other equi-spaced discrete components such as sidebands (and where constant bandwidth is used).
2. Where the frequency range is restricted.
3. Where comparisons are to be made with analyses which inherently have a linear frequency scale (e.g. FFT, time compression).

while a logarithmic frequency scale is most appropriate in the following situations:

1. Where constant percentage bandwidth filters are used.
2. Where a wide frequency range is to be covered.
3. In order to eliminate the effects of minor machine speed fluctuations by a lateral shift of the spectrum.
4. In order to emphasize relationships which are linear in log-log representations, e.g. integration (-20 dB/decade) and bandwidth compensation (-10 dB/decade).

Note that some analyzers (e.g. B & K Type 2120) employ a hyperbolic frequency scale which repeats for each half decade.

3.6.2.3. Choice of Bandwidth

It is found that the time required for a swept frequency analysis is proportional to the square of the reciprocal of the bandwidth, so it is important that the latter be chosen as large as possible while still giving adequate resolution.

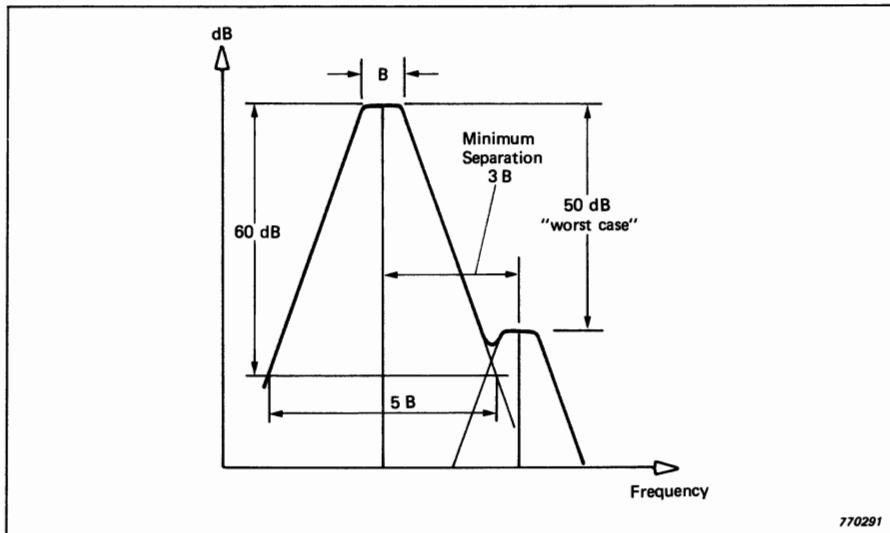


Fig 3.49. Minimum separation of sinusoidal components for filter shape factor 5 (typical)

For **deterministic** signals, as discussed in Section 2.4.1, it is desirable that the most closely spaced spectral components can be separated, and this will generally be the case if the bandwidth is made of the order of 1/3 of the minimum spacing (assuming a shape factor of about 5). See Fig.3.49.

For **random** signals, as discussed in Section 2.4.2, it is desirable for the percentage bandwidth to be chosen less than 1/3 of the bandwidth of the narrowest peak in the spectrum. In some cases it will only be possible to satisfy this requirement by using stepping constant bandwidth filters, since 1% bandwidth is typically the minimum constant percentage bandwidth available. As previously mentioned, bandwidth compensation can be used to obtain a smooth result.

For **mixtures** of deterministic and random components it is generally best to base the selection of bandwidth on the deterministic components. In this case, if stepping bandwidth is used to cover a wide frequency range, it is best not to use bandwidth compensation, but to recognize and accept the sudden changes in level of random components where the filter bandwidth changes.

3.6.2.4. Averaging Time T_A

Here, the discussion of Section 3.6.1.1 still applies, with the following additional remarks:

For **deterministic** signals, the minimum requirement is set by Equation (3.18) ($T_A \geq 3/f$) and solutions of this can be obtained from Fig.3.52, but often this will be small in relationship to the filter response time which determines the analysis speed. Where DC recording is used, it is wise to increase the averaging time up to the maximum which does not affect the sweep speed. This limit is given in terms of T_D , the "dwell time per bandwidth" (see Section 3.6.2.5), by the equation:

$$T_A \leq \frac{T_D}{4} \quad (3.19)$$

where allowance has been made for a filter characteristic with shape factor > 4.5 .

For **random** signals, the selection of T_A should be based on Equation (3.9) ($\epsilon = 1/(2 \sqrt{BT_A})$) and Table 3.4 lists values of standard error ϵ against BT_A product. Moreover, Fig.3.52 permits the selection of T_A for various bandwidths and centre frequencies for a BT_A product of 10. This value of T_A can be modified proportionally for other BT_A products.

BT_A	10	20	30	40	50	80	120
ϵ (dB)	1,5	1,0	0,8	0,7	0,6	0,5	0,4

T01039GB0

Table 3.4. Standard error ϵ vs. BT_A product

3.6.2.5. Filter Dwell Time T_D

The “filter dwell time” T_D is defined by the equation

$$S = \frac{B}{T_D} \quad (3.20)$$

where S is the sweep speed and B is the filter bandwidth

Note that S will be in Hz/s for B expressed in Hz, but will be in mm/s (paper speed) for B expressed in mm (equivalent length on the recording paper).

Thus T_D is the time taken to sweep over one bandwidth.

As mentioned previously, the sweep rate (and thus T_D) is limited by one or more of the factors:

1. Filter response time T_R
2. Averaging time T_A
3. Recorder writing speed W

and each of these will be considered in turn.

With respect to **filter response time** T_R , it will be appreciated that if T_D is made equal to T_R , then the recording of the whole spectrum will be delayed by an amount equal to the bandwidth (see Fig.3.50(a)). This is normally not acceptable, and it is customary to make

$$T_D = K_R T_R \quad (3.21)$$

where for example, $K_R = 4$ will give a delay equal to 1/4 bandwidth (see Fig.3.50(b)). This a good compromise as it corresponds to the average frequency error in the fixed filter case. Solutions are given for this case in the graph of Fig.3.52. Combining Equation (3.21) with Equation (3.1) of Section 3.1.2, one obtains the result:

$$BT_D = K_R \quad (3.22)$$

where K_A is to be chosen keeping in mind that the delay is equal to B/K_A . Equation (3.21) will normally govern for deterministic signals.

With respect to **averaging time** T_A , there are two basic considerations to be taken into account when selecting T_D . Letting:

$$T_D = K_A T_A \quad (3.23)$$

then K_A will either be limited by the maximum rate of fall of the detector ($8.7 \text{ dB}/T_A$, see Fig.3.22) which limits the ability to reproduce the filter characteristic, or by the maximum allowable errors and delays in recording peaks and valleys in the spectrum (Figs.3.20, 3.21). Table 3.5 allows selection of K_A based on these factors. With respect to the ability to reproduce the filter characteristic, this will primarily apply where the spectrum contains sinusoidal compo-

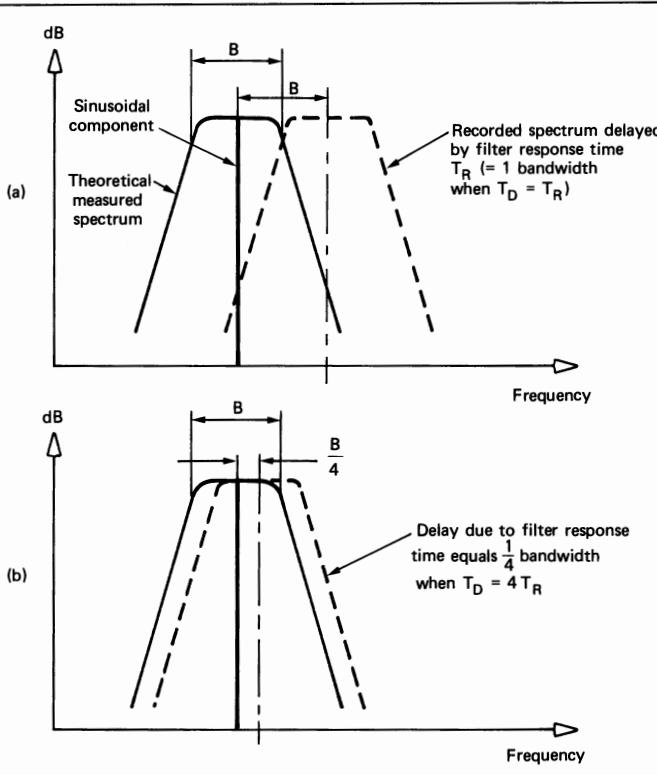


Fig 3.50. Effect of filter response time on recorded spectrum

nents. The filter shape factor to be used is not necessarily the actual value, but that corresponding to the steepest part of the characteristic. The error and delay on peaks and valleys (usually the former) will generally govern in the case of random signals where the analyzer bandwidth is normally less than that of the function being analyzed, and account has been taken of this in setting the values in the table.[§] Equation (3.23) will always govern for random signals and signals on a loop.

K_A	1	2	3	4	5
Filter Shape Factor	15	8	5,6	4,5	3,8
Error on peaks (dB)	1,2	0,5	0,4	0,2	0,15
Delay on peaks ($\times B$)	0,3	0,2	0,13	0,1	0,1
Error on valleys (dB)	3	1	0,5	0,25	0,2
Delay on valleys ($\times B$)	0,5	0,25	0,17	0,13	0,1

T01040GB0

Table 3.5. Choice of K_A according to filter characteristic or required accuracy

With respect to the **Recorder Writing Speed W**, it is not really necessary to calculate T_D , as the paper speed is determined directly from the writing speed (see Eqn.(3.25)). If it is of interest, then T_D can be calculated back from the paper speed (using Equation (3.24)).

3.6.2.6. Writing Speed W

Here, the discussion of sections 3.6.1.2 and 3.6.1.3 applies fully, but it is perhaps worth remarking that in the case of narrow band analysis, the advantages of DC vs AC recording are generally more marked, partly because the analysis time becomes more critical, and partly because the validity of the results is not so obvious by inspection, and thus the selection of parameters such as averaging time is made somewhat more conservatively.

[§]Errors and delays given in the table will be correct when the filter bandwidth is equal to that of peaks in the spectrum being analyzed. They may be underestimated when the signal bandwidth is less (e.g. discrete components), but then the filter shape factor would normally govern. When the filter bandwidth is less than that of peaks in the spectrum, the indicated errors will be conservative, and delays approximately correct

3.6.2.7. Paper Speed P

As for the discussion of section 3.6.1.4, the selection of paper speed depends primarily on whether DC or AC recording is used.

For **DC recording**, the maximum paper speed may be obtained from Equation (3.20), expressed in the form

$$P = \frac{B_{eq}}{T_D} \quad (3.24)$$

where B_{eq} is the equivalent of the bandwidth in mm along the recording paper.

There are two advantages in using this approach:

1. It is not necessary to go through the intermediate step of calculating the sweep rate in Hz/s.
2. Equation (3.24) is equally applicable to the case of constant bandwidth, where the bandwidth in Hz corresponds to a certain length in mm (on a linear scale), and to constant percentage bandwidth, where a given percentage corresponds to a certain length in mm (e.g. on a logarithmic scale).

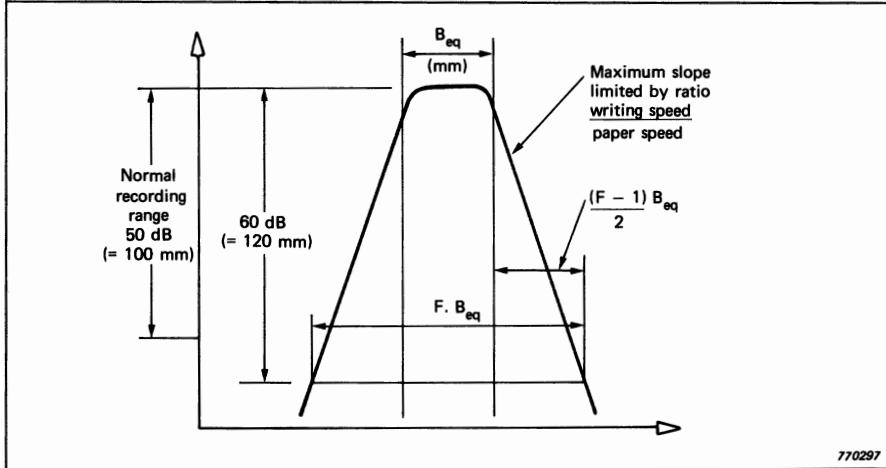


Fig 3.51. Limitation of writing speed in recording a filter characteristic of shape factor F

For **AC-recording**, the sweep rate will often be limited by the ability to write out the steepest filter flank with a given writing speed. Referring to Fig.3.51 it will be seen that the ratio of writing speed to paper speed must be greater than the slope of the filter characteristic which can be expressed as

$$\left(\frac{F-1}{2} \right) \frac{B_{eq}}{120}$$

where, as illustrated, F is the shape factor corresponding to the steepest part of the filter characteristic, and 120 mm is the paper width corresponding to 60 dB (i.e. assuming that 100 mm corresponds to 50 dB).

Thus, $P \leq \frac{B_{eq} W}{K_w}$ (3.25)

where $K_w = \frac{240}{(F-1)}$ (3.26)

and W is the writing speed corresponding to 100 mm paper width. Values of K_w can be obtained from Table 3.6.

K_w	100	75	50	40	30
Equiv. Shape Factor F	3,4	4,2	5,8	7,0	9,0

T01041GB0

Table 3.6. Choice of K_w according to filter shape factor

Where it is the averaging time which governs T_D (i.e. random signals and looped signals where T_A is greater than the loop length) then Equation (3.25) will always govern. Where it is the filter response time which governs T_D (i.e. continuous deterministic signals) then it must be checked whether Equation (3.24) or (3.25) governs.

It has already been shown (section 3.3.2) that for DC recording, if a writing speed of 1000 mm/s is selected, then there is no effective increase in averaging time. It can be shown that there will be no limitation with respect to reproduction of filter slope either. The minimum averaging time available for DC recording with B & K analyzers is 0,1 s. Inserting this in Eqn. (3.23) with $K_A = 2$ (corresponding to Shape Factor 8) gives $T_D = 0,2$ s. Inserting this in turn in Equation (3.24) gives

$$P \leq \frac{B_{eq}}{0,2} = 5 B_{eq}$$

Inserting the same shape factor in Equation (3.26) gives $K_W = 34,3$ and thus from Equation (3.25)

$$P \leq \frac{B_{eq} \cdot 1000}{34,3} = 29,2 B_{eq}$$

Accordingly, the limiting speed is almost 6 times faster, and thus Equation (3.24) will always govern.

3.6.2.8. Summary

It is probably necessary at this stage to summarize the most important information in the foregoing. Fig.3.52 indicates a logical procedure to be followed in the general case, and allows the most important parameters to be selected graphically. The following values have been chosen as most generally appropriate:

- (1) $K_R = 4$ (See Section 3.6.2.5) i.e. $BT_D \geq 4$
- (2) $K_A = 2$ (See Section 3.6.2.5) i.e. $T_D \geq 2 T_A$
- (3) $K_W = 50$ (See Section 3.6.2.7) i.e. $P \leq B_{eq} \cdot W/50$

Note that the reason for the somewhat more conservative approach in selecting K_W compared with K_A (with respect to filter shape factor) is that the writing speed gives the same limitation for rising and falling curves, while the averaging time primarily limits falling curves only (and thus peaks are more likely to be recorded correctly).

3.6.2.9. Calibration

In the analysis of **deterministic** signals it is usual to express the results directly as an amplitude spectrum, where the amplitudes have the same dimensions as the input signal, e.g. volts. If the amplitudes are represented on a logarithmic scale then a ratio of 10 in amplitude corresponds to 20 dB and Fig.3.53 can be used to convert to linear amplitudes within each 20 dB range. It is normal to set the full-scale value on the recording paper to a round figure (i.e. in 10 dB steps) and this can be done as follows:

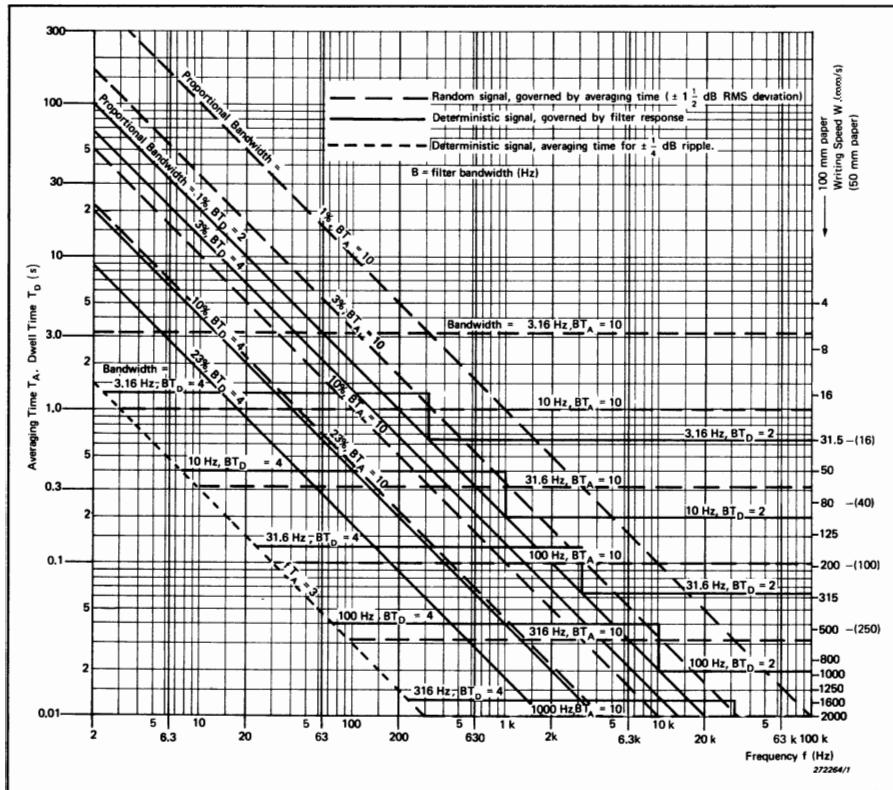


Fig. 3.52. SWEEP SPEED FOR SWEPT FREQUENCY ANALYSIS

For each bandwidth and frequency range, determine Filter Dwell Time, Averaging Time and Sweep Speed as follows:

1. Random signal: Dwell time T_D is always determined by averaging time T_A .

From the graph read off T_A from the appropriate line for $BT_A = 10$ (horizontal lines for constant bandwidth, sloping lines for constant proportional bandwidth). This value will correspond to ± 1.5 dB RMS error. For a higher BT_A product and consequent reduced error, increase the value of T_A proportionally. Calculate $T_D = 2 T_A$. For DC recording, calculate sweep speed $S = B/T_D$.

For AC recording, T_A is determined by recorder writing speed W and this can be read directly from the right-hand scale. Sweep speed can then be calculated from $S = BW/50$ (applicable to 50 dB potentiometer and 100 mm paper).

2. Deterministic Signal (periodic or quasi-periodic): Read T_D based on

filter response time from appropriate line for $BT_D = 4$ (or 2 where bandwidth $\leq 1\%$). Read also minimum averaging time T_A based on $\pm 1/4$ dB ripple from line $fT_A = 3$ (independent of bandwidth).

For DC recording, calculate sweep speed based on filter response as $S = B/T_D$.

For AC recording, read off writing speed W corresponding to T_A and calculate sweep speed based on recorder response as $S = BW/50$. The governing sweep speed is the lesser of this and the value as calculated for DC recording.

In all cases where sweep speed is governed by filter response time, increase T_A to the highest value which does not affect sweep speed. (i.e. $T_D/4$ for DC recording.)

Note (1) Averaging times will normally be constant over at least a half decade, whereas filter response times change automatically with frequency.

Note (2) S will be in Hz/s for B in Hz, but will be directly in mm/s for B expressed as equivalent recorder paper length in mm

1. Using the internal reference signal of the analyzer

This is typically 50 mV for B & K analyzers. With the analyzer in linear frequency mode (i.e. filter disconnected) and meter range 100 mV, the Level Recorder input attenuation is adjusted until the pen is 6 dB below full-scale. The full-scale on the paper then corresponds to the full-scale voltage of the meter for all the succeeding measurements. This full-scale voltage can easily be changed in 10 dB steps as required. Conversion of the scaling to mechanical units is straightforward provided the conversion factor of the input signal is known, (e.g. 10 mm/s per volt).

2. Using an external reference signal

This might for example be a pistonphone signal or the signal obtained from an accelerometer mounted on an accelerometer calibrator. The signal may

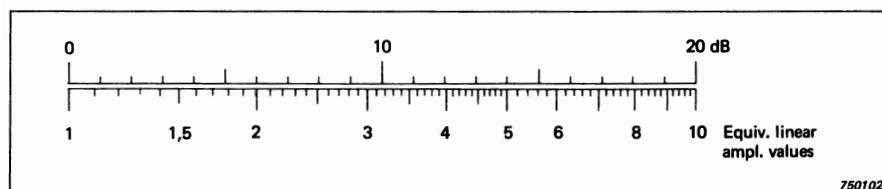


Fig. 3.53. Linear amplitude vs. dB levels

either be played directly into the analyzer input or recorded on a tape recorder. To illustrate the general method, the case will be taken of an accelerometer calibrator which gives a vibration level of 1 g peak (i.e. 0,7 g RMS). The signal from the measurement system is played into the input of the analyzer (which is in linear mode) and the recorder pen adjusted to 3 dB below full-scale. The full-scale will now correspond to 1 g RMS and all later measurements can be related to this, taking due account of any 10 dB steps in amplification of either the measurement, analysis, or recording systems.

In the case of **random** signals, the results may be expressed as an amplitude spectrum, but it may be preferred to express them as a power spectral density (PSD).

For **constant bandwidth** analysis this only involves a modification of the scaling as follows:

The PSD is obtained by squaring the linear amplitude and dividing by the filter bandwidth, and the result is of course independent of frequency.

e.g. for a full-scale level of 10 g and 100 Hz bandwidth
the full-scale power = 100 g^2
and full-scale PSD = $100 \text{ g}^2/100 \text{ Hz} = 1 \text{ g}^2/\text{Hz}$

Normally, this scaling must be adjusted each time the bandwidth is changed, but for an analyzer with "bandwidth compensation" it is adjusted automatically, by means of an attenuation proportional to the filter bandwidth, and the calibration need only be performed once, even where the bandwidth changes automatically during normal measurements. A sinusoidal reference signal may be used, but it is desirable to make the calibration adjustments with the filter selected, and thus the filter must be centred on the reference signal so as not to attenuate it.

For **constant percentage bandwidth** analysis of **random** signals the bandwidth varies proportionally with the frequency, and thus the conversion factor to calibrate as a PSD varies with frequency. On the normal log-log paper used for 1/3-octave analysis the conversion factor is linear and corresponds to a linear slope of -3 dB/octave (-10 dB/decade). Thus, if the conversion to PSD is performed at one frequency (e.g. 100 Hz) then it is relatively simple to draw lines of constant PSD through the values calculated at 100 Hz and sloping upwards at 10 dB/decade with increasing frequency. Fig.3.54 illustrates a typical example. Alternatively, an instrument may be used which gives a power attenuation proportional to frequency (i.e. -10 dB/decade) and if this is inserted before the analyzer input, the system can be calibrated as described for "bandwidth compensation". This is desirable in the case of analyzers which have a hyperbolic or other non-logarithmic frequency scale, since it is not then so simple to draw lines of constant PSD.

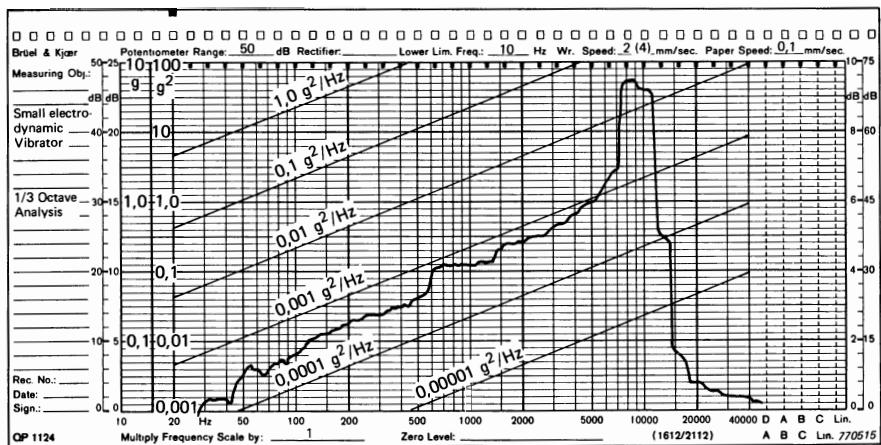


Fig. 3.54. Typical 1/3-octave power spectrum with lines of constant PSD

3.6.3. Digital filter analysis

The operation of the Digital Frequency Analyzer Type 2131 is almost self-explanatory, but it may help to give a guide to the selection of those parameters which are not decided automatically.

3.6.3.1. Linear vs. Exponential Averaging

Normally, linear averaging would be chosen in the following cases:

- (1) Where the signal sample is limited in length, e.g. a short tape recording.
- (2) Where it is desired to minimize the statistical error in a measurement made on a stationary random signal with a certain fixed record length.
- (3) Where linear averaging is specified or recommended.

Exponential averaging would be used in the following cases:

- (1) Where it is desired to monitor a continuous signal which may be slowly varying (and therefore not stationary over a time considerably longer than the averaging time).
- (2) Where it is desired to obtain a result with a uniform statistical error over all frequencies (i.e. uniform BT_A product).

3.6.3.2. Averaging Time T_A

The same basic considerations apply as discussed in section 3.6.1.1 for stepped 1/3-octave filtration, but it may be worth making the following remarks which apply specifically to the 2131 Digital Frequency Analyzer.

For continuous **deterministic** signals the easiest way to ensure that the averaging time complies with Eqn.(3.18) ($T_A \geq 3/f$) is to check visually that the fluctuations are acceptable, in particular in cases where f is to be interpreted as a beat or modulating frequency smaller than the analyzer bandwidth. It is quite likely that such modulating frequencies will be common to all frequency ranges and thus a constant averaging time will be optimal.

For **stationary random** signals it may be preferred to choose the averaging mode giving constant BT_A product and thus uniform error at all frequencies. One advantage of this is that the result is then known to be valid at all frequencies.

Where it is desired to get maximum information from a **limited record**

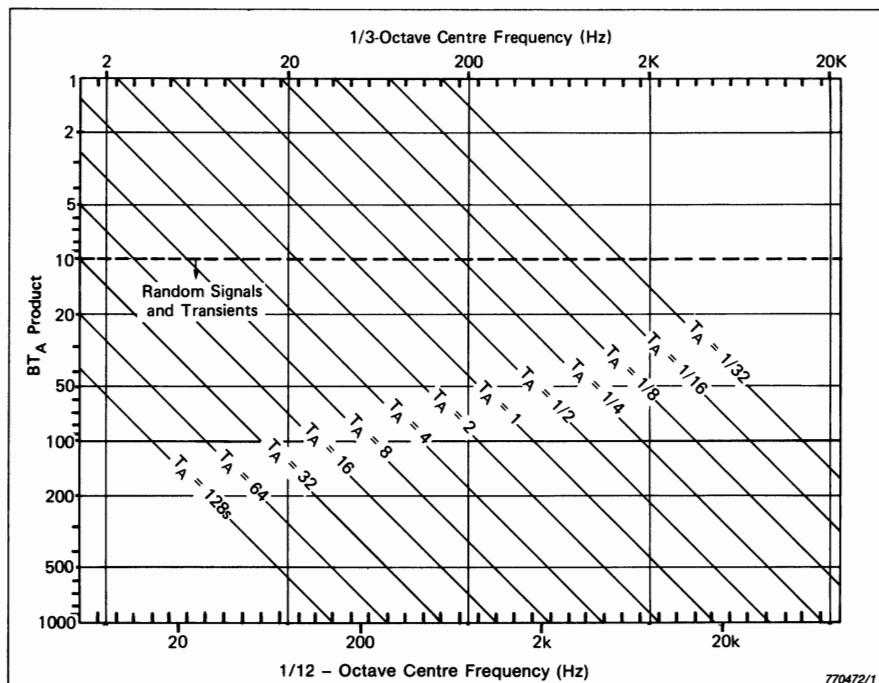


Fig. 3.55. BT_A product vs. averaging time T_A for the 2131 Analyzer

length, of either a random or deterministic signal, then the averaging time is virtually decided by the record length. Note that the result may not then be valid at all frequencies, since for short averaging times, at least, the BT_A product may be too low.

Fig.3.55 indicates the BT_A product achieved with the various averaging times over the frequency range of the analyzer. The invalid range where BT_A is less than 1 is not included and a dotted line demarcates the values of BT_A less than 10 for which the result would not normally be valid for random signals. The values of BT_A product may be inserted in Eqn.(3.9) ($\epsilon = 1/(2 \sqrt{BT_A})$) or Table 3.4 to obtain the relative standard deviation of the error for gaussian random signals.

4. FAST FOURIER TRANSFORM (FFT)

As mentioned in Section 2.2.4 the Fast Fourier Transform (FFT) is an algorithm or calculation procedure for obtaining the Discrete Fourier Transform (DFT) with a greatly reduced number of arithmetic operations compared with a direct evaluation. Since its first publication in 1965 (Ref.4.1) it has revolutionized the field of signal analysis, and it is still probably the most important single analysis technique available. At first the algorithm was implemented on large computers in a high-level language such as FORTRAN, later in assembler language on mini-computers. For several years there has been a dominance of

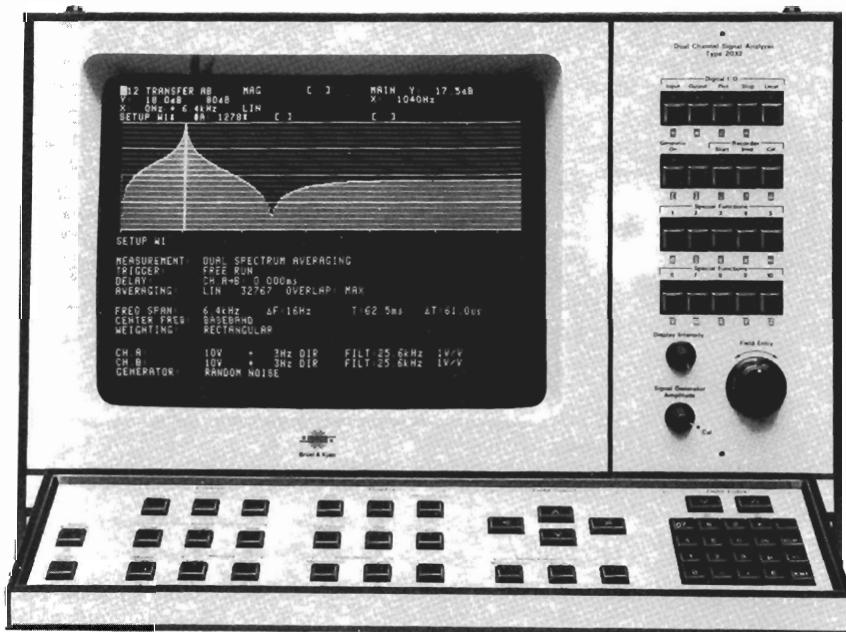


Fig. 4.1. Dual Channel Signal Analyzer Type 2032

dedicated analyzers which have the FFT algorithm implemented in a combination of hardware and firmware. The B & K Analyzer Type 2032 (Fig.4.1) is one of the most advanced of this kind, and despite being a standalone analyzer, is very flexible in its operation, and can even be programmed by the user for specialised tasks.

One of the major advantages of the FFT over other types of frequency analysis, e.g. using filters, is that retention of phase information makes transformation in either direction possible and in fact relatively simple. It also permits the evaluation of a large number of functions applicable to multi-channel measurements and system analysis, e.g. correlation, coherence, frequency response functions, etc. In this chapter, the discussion is limited to frequency analysis of single channel (stationary) signals, but the other topics are taken up in Chapter 7.

4.1. THE FFT ALGORITHM

For convenience, the basic equations of the DFT (from Section 2.2.4) will be repeated here. In particular, the DFT (forward and inverse transforms) are represented by:

$$G(k) = \frac{1}{N} \sum_{n=0}^{N-1} g(n)e^{-j\frac{2\pi kn}{N}} \quad (2.18)$$

$$g(n) = \sum_{k=0}^{N-1} G(k)e^{j\frac{2\pi kn}{N}} \quad (2.19)$$

while a matrix version of Eqn. (2.18) for $N = 8$ is represented by:

$$\begin{bmatrix} G_0 \\ G_1 \\ G_2 \\ G_3 \\ G_4 \\ G_5 \\ G_6 \\ G_7 \end{bmatrix} = \frac{1}{8} \begin{bmatrix} \uparrow & \uparrow \\ \uparrow & \nearrow & \rightarrow & \swarrow & \downarrow & \leftarrow & \uparrow & \nearrow \\ \uparrow & \rightarrow & \downarrow & \leftarrow & \uparrow & \rightarrow & \downarrow & \leftarrow \\ \uparrow & \swarrow & \leftarrow & \rightarrow & \uparrow & \swarrow & \rightarrow & \swarrow \\ \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \downarrow \\ \uparrow & \leftarrow & \rightarrow & \nearrow & \uparrow & \leftarrow & \downarrow & \rightarrow \\ \uparrow & \nearrow & \leftarrow & \downarrow & \uparrow & \nearrow & \downarrow & \rightarrow \\ \uparrow & \downarrow & \nearrow & \leftarrow & \uparrow & \downarrow & \nearrow & \rightarrow \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ g_2 \\ g_3 \\ g_4 \\ g_5 \\ g_6 \\ g_7 \end{bmatrix} \quad (2.20a)$$

The advantages of the FFT can be achieved in a variety of ways (Refs.4.2, 3.3) but we will first limit the discussion to a particular version of a Radix 2 algorithm, where N is a power of 2. The differences between the different versions are of a secondary nature and of most interest to an instrument designer rather than user. However, it is probably of interest to a user to know roughly how the FFT algorithm functions, so as to be able to appreciate any

restrictions in its applicability. Users who wish to modify the operation of the analyzer, which is possible with the Type 2032 already mentioned, using so-called SPL (Signal Processing Language) will also benefit from a detailed knowledge of the operation.

Basically, the savings of the FFT algorithm result from factorizing the **A** matrix of Eqn.(2.20a) into a number ($\log_2 N$) of individual matrices. This does not immediately appear to give any benefit, but it will be shown that the factor matrices contain only two non-zero elements in each row, of which one is always unity, so that multiplication by each of the factor matrices requires only complex multiplications. The total number of multiplications is thus of the order of $N \log_2 N$ instead of the N^2 required to multiply by the **A** matrix in one step. As mentioned in Section 2.2.4 this represents a saving by a factor of 100 for the typical case of $N = 1024$.

Row number in B		Row number in A
0 0 0 (0)	↑ ↑ ↑ ↑ ↑ ↑ ↑ ↑	0 0 0 (0)
0 0 1 (1)	↑ ↓ ↑ ↓ ↑ ↓ ↑ ↓	1 0 0 (4)
0 1 0 (2)	↑ → ↓ ← ↑ → ↓ ←	0 1 0 (2)
0 1 1 (3)	↑ ← ↓ → ↑ ← ↓ →	1 1 0 (6)
1 0 0 (4)	↑ ↗ → ↘ ↓ ↗ ← ↘	0 0 1 (1)
1 0 1 (5)	↑ ↗ → ↘ ↓ ↗ ← ↘	1 0 1 (5)
1 1 0 (6)	↑ ↙ ← ↖ ↓ ↙ → ↖	0 1 1 (3)
1 1 1 (7)	↑ ↙ ← ↖ ↓ ↙ → ↖	1 1 1 (7)

770482

Fig. 4.2. Matrix **B** (**A** with rows reshuffled to bit-reversed address)

In fact it is not the matrix **A** directly which is factorized, but a reshuffled version of it which is illustrated in Fig.4.2, and which will be called **B**. Matrix **B** is obtained from **A** by interchanging the rows with those of "bit-reversed address", i.e. the binary representation of the row numbers 0 to 7 (000 to 111) is reversed end-for-end to obtain the new address. Symmetrical numbers such as 101 remain unchanged. In Fig.4.2, on the left-hand-side, the actual row numbers are given in binary form, while on the right-hand side the bit-reversed numbers are given, and it can be checked that these indicate the original row number (in matrix **A**) of the same row. It will be seen that multiplication by matrix **B** instead of **A** will mean that the results will also be in bit-reversed order and will have to

be reshuffled to obtain them in natural order. The reshuffling is a rapid process, however, requiring negligible time in comparison with the matrix multiplications.

Fig.4.3 shows the three matrices **X**, **Y** and **Z** into which **B** can be factorized, and demonstrates that when they are multiplied together the result is the matrix **B**. Even though this demonstration only applies for $N = 8$, it will be appreciated that the same principles can be extended to values of N equal to any power of 2.

X	Y	Z
$\begin{bmatrix} \uparrow & \uparrow & 0 & 0 & 0 & 0 & 0 & 0 \\ \uparrow & \downarrow & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \uparrow & + & 0 & 0 & 0 & 0 \\ 0 & 0 & \uparrow & - & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \uparrow & + & + & 0 \\ 0 & 0 & 0 & 0 & \uparrow & \diagup & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \uparrow & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \uparrow & \downarrow \end{bmatrix}$	$\begin{bmatrix} \uparrow & 0 & \uparrow & 0 & 0 & 0 & 0 & 0 \\ 0 & \uparrow & 0 & \uparrow & 0 & 0 & 0 & 0 \\ \uparrow & 0 & \downarrow & 0 & 0 & 0 & 0 & 0 \\ 0 & \uparrow & 0 & \downarrow & 0 & 0 & 0 & 0 \\ 0 & \uparrow & 0 & \downarrow & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \uparrow & 0 & \rightarrow & 0 \\ 0 & 0 & 0 & 0 & 0 & \uparrow & 0 & \rightarrow \\ 0 & 0 & 0 & 0 & 0 & 0 & \uparrow & 0 & \leftarrow \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \uparrow & 0 & \leftarrow \end{bmatrix}$	$\begin{bmatrix} \uparrow & 0 & 0 & 0 & \uparrow & 0 & 0 & 0 \\ 0 & \uparrow & 0 & 0 & 0 & \uparrow & 0 & 0 \\ 0 & 0 & \uparrow & 0 & 0 & 0 & \uparrow & 0 \\ 0 & 0 & 0 & \uparrow & 0 & 0 & 0 & \uparrow \\ 0 & 0 & 0 & 0 & \uparrow & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \uparrow & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \uparrow & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \uparrow \end{bmatrix}$

Multiplication **Y** . **Z** can be considered as:

$$\begin{bmatrix} Y_1 & 0 \\ 0 & Y_2 \end{bmatrix} \begin{bmatrix} I & I \\ I & -I \end{bmatrix} = \begin{bmatrix} Y_1 & Y_1 \\ Y_2 & -Y_2 \end{bmatrix} = C$$

Matrix **C** can thus be written down and multiplied on the left by **X** to give the result $\mathbf{X} \cdot \mathbf{C} = \mathbf{X} \cdot \mathbf{Y} \cdot \mathbf{Z}$

$\begin{bmatrix} \uparrow & \uparrow & 0 & 0 & 0 & 0 & 0 & 0 \\ \uparrow & \downarrow & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \uparrow & + & 0 & 0 & 0 & 0 \\ 0 & 0 & \uparrow & - & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \uparrow & \diagup & 0 & 0 \\ 0 & 0 & 0 & 0 & \uparrow & \diagup & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \uparrow & 0 & \downarrow \end{bmatrix} \times \begin{bmatrix} \uparrow & 0 & \uparrow & 0 & 0 & \uparrow & 0 & 0 \\ 0 & \uparrow & 0 & \uparrow & 0 & \uparrow & 0 & 0 \\ \uparrow & 0 & \downarrow & 0 & \uparrow & 0 & \downarrow & 0 \\ 0 & \uparrow & 0 & \uparrow & 0 & \uparrow & 0 & 0 \\ 0 & \uparrow & 0 & \uparrow & 0 & \uparrow & 0 & 0 \\ 0 & 0 & \uparrow & - & 0 & \downarrow & 0 & 0 \\ 0 & 0 & \uparrow & - & 0 & \downarrow & 0 & 0 \\ 0 & 0 & 0 & 0 & \uparrow & 0 & \leftarrow & 0 \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow \\ \uparrow & \downarrow \\ \uparrow & \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \downarrow & \uparrow \\ \uparrow & \uparrow & \uparrow & \downarrow & \uparrow & \downarrow & \uparrow & \uparrow \\ \uparrow & \uparrow & \uparrow & \uparrow & \downarrow & \uparrow & \uparrow & \uparrow \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \downarrow & \uparrow & \uparrow \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \downarrow & \uparrow \\ \uparrow & \downarrow \end{bmatrix}$
--

Thus $\mathbf{X} \cdot \mathbf{C} = \mathbf{B}$ Q.E.D.

770509

Fig. 4.3. Factorization of **B** into **X**, **Y** and **Z**

As an example of the systematic nature of the factor matrices, it will be seen that the upper left quadrant of matrix **Y** (submatrix Y_1) is a copy of matrix **Z** but of order $N/2$. The upper left submatrix of matrix **X** is the next smaller version of the same matrix, having the general form:

$$\begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & -\mathbf{I} \end{bmatrix}$$

Each successive matrix (in the order **Z**, **Y**, **X**) introduces progressively smaller rotations; **Z** involves only $1/2$ revolutions, **Y** introduces $1/4$ revolutions while it is only at the last stage that $1/8$ revolutions are introduced.

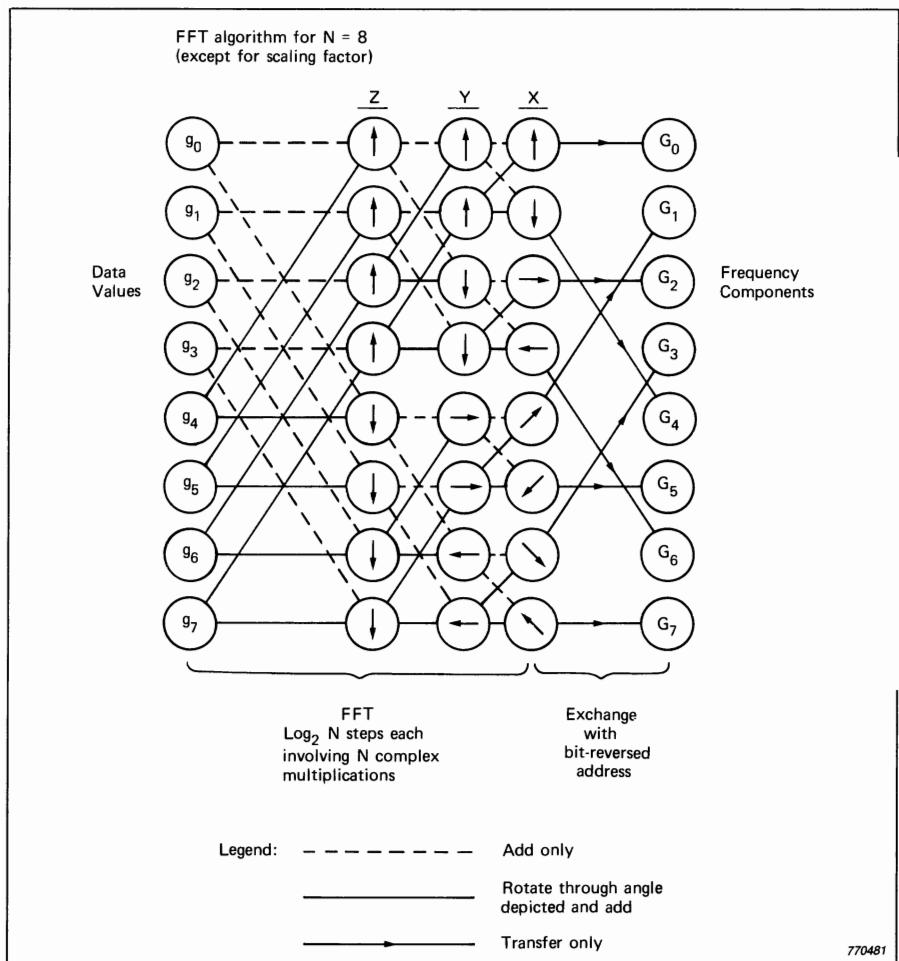


Fig. 4.4. Flow diagram for FFT algorithm

Fig.4.4 shows another representation of the same algorithm, this time indicating the way in which the successive multiplications by the factor matrices **Z**, **Y** and **X** can actually be carried out in a computer. It will be seen that the whole operation can be done "in place", i.e., the result of each step is stored in the same memory locations as the original data (and thus so is the final result).

At each step the data values can always be operated on pairwise, the results of the operation being placed back in the memory locations of the 2 data values from which they were obtained. It can easily be checked that the first three steps in the diagram of Fig.4.4 correspond to multiplication by the matrices **Z**, **Y** and **X**, respectively, while the final step represents the reshuffling to bit-reversed addresses. Note that the reshuffling can alternatively be done as a first step, and this is for example the case in the 2032 Analyzer.

Each of the operations on a pair of data values is called a "butterfly" operation, and by examination of the butterflies in the diagram of Fig.4.4, it can be seen that the actual number of complex multiplications to be performed can be halved again, because one unit vector in each butterfly is always simply the negative (180° phase shift) of the other. Thus, only one actual multiplication by the components of the vector is necessary in each butterfly operation. This concept can be extended to obtain a further speedup by 2:1 using a so-called Radix 4 algorithm, where sequence length N is a power of 4.

The Radix 4 algorithm corresponds to combining the factor matrices pairwise (cf. matrix **C** in Fig.4.3). There will now be half the number of factor matrices, but twice as many non-zero elements (viz. 4) in each row. Each butterfly thus operates on four elements, and places the results in the same four memory locations. However, the unit vectors in each butterfly are either in-line, at 180° to each other or at 90° to each other (see Fig.4.5). This can be seen in fact from matrix **C** of Fig.4.3 (even though this is of order 8 instead of 16) where each column (representing operations on the same element) has four non-zero elements complying with this rule.

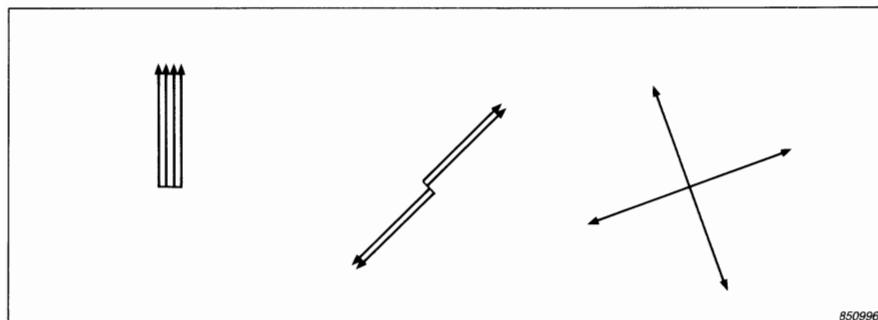


Fig. 4.5. Vector patterns for Radix 4 algorithm

Thus, once again only one actual complex multiplication is involved in each butterfly operation, all other results being obtained by changing sign, exchange between real and imaginary components etc. These simple operations are much faster than multiplications, so the overall effect is a speedup by almost 2:1.

The only disadvantages are the limitation on transform size to a power of four, and a slightly more complex algorithm.

A slight further increase in speed could be achieved with a radix 8 algorithm, because the sine and cosine components of 45° are the same, but this is normally not considered to compensate for the extra complexity of the algorithm, and the restriction to powers of 8 in sequence length.

A practical point is that multiplication by a unit vector with a given orientation involves multiplications by the sine and cosine coefficients of that angle. Considerable time can be saved by having these sine values tabulated, rather than generating them repeatedly, although this of course requires more memory space. However, for a transform of size N it is only necessary to store $N/4$ sine coefficients since all values can be simply generated from those for the first quarter period. Moreover, both sine and cosine coefficients can be generated from the same table.

4.2. THE FFT FOR REAL-VALUED SERIES

Thus far, no distinction has been made between the time samples $g(n)$ and the frequency spectrum values $G(k)$, but in the most common practical situation the $g(n)$ values will be real while the $G(k)$ values will be complex. The FFT algorithm so far discussed is equally valid for real or complex data. If applied to real data, however, there will be two redundancies:

1. The imaginary part of each input data value will be zero and thus half the memory will be used for storing zeroes.
2. The second half of the resulting spectrum, i.e. the frequencies from f_N to f_s also represent the negative frequencies from $-f_N$ to zero (Fig.2.6) and since the latter are the complex conjugates of the positive frequency values (Eqn.(2.12)) there is no need to store them separately.

It is possible to remove these redundancies by means of an algorithm which transforms a number N of real values as though they were $N/2$ complex values and then manipulates the result to obtain the first half of the spectrum of the original real data points (Ref.4.3). Most FFT systems operate in this manner. It should be noted that because of the antisymmetry of the imaginary components of the spectra of real functions both the zero frequency (DC) and Nyquist

frequency components are real numbers and in fact appear as the real and imaginary components, respectively, of the first complex number of the output spectrum (Fig.4.6). All the other complex numbers are genuine frequency components distributed linearly with frequency up to $f_N \times (N-2)/N$ i.e. just less than the Nyquist frequency. It is possible to move the Nyquist frequency component to its correct position at the end of the table, but this then requires $(N/2 + 1)$ (complex) storage locations. It is in any case not usually necessary, since to avoid aliasing it is desirable for the Nyquist frequency component to be zero, or at least so small that it is not usable.

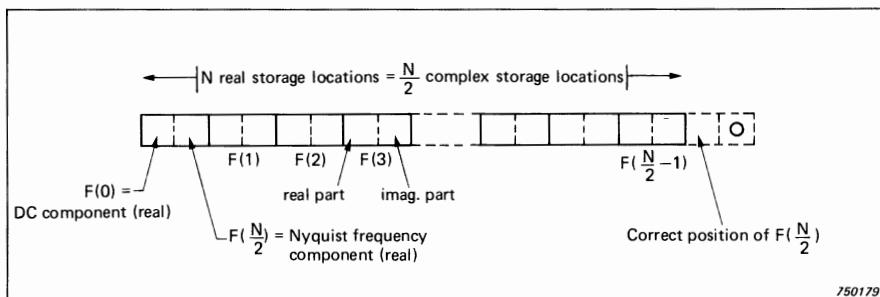


Fig. 4.6. Arrangement of output array from FFT algorithm for N real-valued samples

Some FFT analyzers, e.g. the B & K Type 2033, only have an algorithm implemented for real-valued time series. However, with the aid of an external computer to manipulate the data, it is possible to perform inverse transforms of complex data, as explained in detail in Ref.4.4.

4.3. LIMITATIONS OF FFT PROCESSING

As has just been shown, the FFT algorithm produces an identical result to direct application of the DFT. Thus, any limitations of the FFT process are those of the DFT. These have been briefly discussed in Section 2.2.4, and are basically due to the finite (circular) and discrete nature of the DFT algorithm. Thus, regardless of the actual nature of the input signal, the analyzed record and results are a finite number, N , of discrete digital samples, in theory representing one period of an infinitely long periodic signal.

The effects of this are illustrated in Fig.4.7, which makes use of the Convolution Theorem in a graphical way, as developed in Chapter 2. Detailed discussions will also be found in Refs.4.5 and 4.6.

Fig.4.7(a) shows a time signal $g(t)$ and its Fourier spectrum $G(f)$, both assumed to be infinitely long. The first step in a digital analysis is discrete

sampling of the time signal, which, as illustrated in Fig.4.7(b) can be considered as equivalent to multiplication by an impulse train $\Delta_1(t)$ with sampling interval Δt . Its Fourier transform $\Delta_1(f)$ is likewise an impulse train, with impulses at all multiples of the sampling frequency $f_s = 1/\Delta t$. Fig.4.7(c) shows the result of this

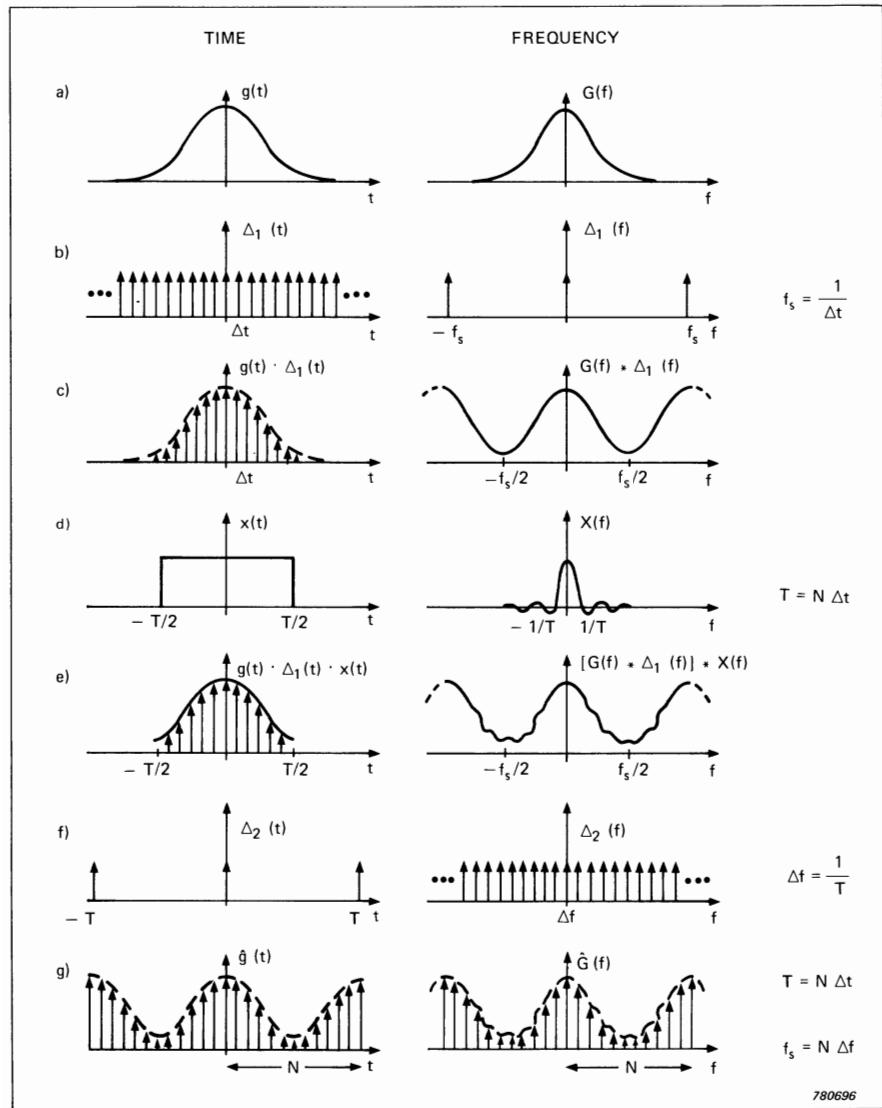


Fig. 4.7. Derivation of the Discrete Fourier Transform from the Integral Transform

multiplication in the time domain, which in the frequency domain corresponds to a convolution of the two spectra. Thus, it is seen that any frequency components in the original spectrum $G(f)$ which extend over half the sampling frequency, overlap in that region and cause aliasing, as discussed in Chapter 2. This is the first problem to be dealt with in using the DFT.

The next step in the procedure, represented by Figs.4.7(d) and (e), shows the effect of the time limitation necessary to fit the time signal into a finite record length. This is the equivalent of multiplication by a time window function (rectangular in the illustration) so that once again there is a convolution in the frequency domain, this time with the Fourier transform of the time window function. For the rectangular time window illustrated, the spectral function is a $\sin x/x$ function with zeroes at multiples of $1/T$, where T is the window length. The effect of the window function is not very great on the smooth frequency spectrum illustrated, but for signals containing discrete frequency components (delta functions), these become replaced by the window function as illustrated in Fig.2.29. This effect, variously known as "window effect", "sidelobe generation" and "leakage" (because power from discrete frequency components is "leaked" into adjacent bands), is the second problem to be dealt with in using the DFT.

The final effect is illustrated in Figs.4.7(f) and (g) and is known as the "picket fence effect". It is not possible to represent the continuous spectrum of Fig.4(e) in a digital memory and so it must also be discretely sampled. Hence the name, as it can be considered as viewed through the slits in a picket fence (Fig.4.8). By analogy with the steps of Figs.4.7(b) and (c), the discrete sampling in the frequency domain corresponds to a periodic repetition of the time record, with a period equal to the record length.

These three effects, and how to deal with them, will now be discussed in a little more detail.

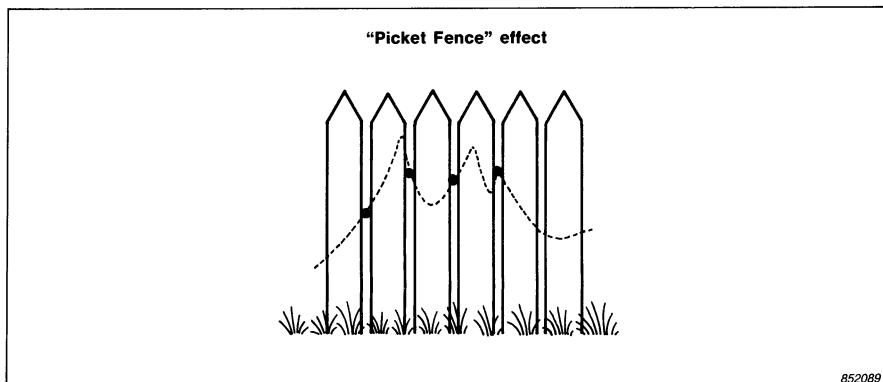


Fig. 4.8. Illustration of Picket Fence Effect

4.3.1. Aliasing

Aliasing is not normally of concern in the analysis of stationary signals, because the analyzers generally have built-in low-pass filters appropriate to each frequency range. Because all values up to the Nyquist frequency are calculated in any case, it is usual to choose a very steep low-pass filter with cutoff frequency at about 80% of the Nyquist frequency, and only display results unaffected by the filter. Typically, for a 1K (1024 point) transform, 512 frequency components are calculated, and 400 displayed. Similarly, for a 2K transform, 800 lines are displayed (or 801 where the DC component is included).

The steepness of filter characteristic required is of the order of 120 dB/octave to ensure that high frequency components folded back into the measurement range (by aliasing) are attenuated sufficiently. In the case of a 2048 point transform, for example, with Nyquist frequency in line number 1024, a component in line number 1248 ($800 + 224 + 224$) would fold back into line number 800, and is only a little more than half an octave higher in frequency. A lowpass filter with 120 dB per octave roll-off would attenuate it by approx. 75 dB and thus take it outside the normal dynamic range of a typical analyzer.

The lowpass filters can either be analog or digital, but in the latter case the first lowpass filter applied before digitising must of course be analog. As examples, the B & K single channel analyzer Type 2033 has a separate analog filter for each of its 11 frequency ranges, while the dual channel analyzer Type 2032 (Fig.4.1) has a single analog filter at 25,6kHz (for each channel) and all lower frequency bands are achieved by digital filtering and resampling in an appropriate number of octave steps.

Even so, as previously stated, aliasing is not usually a problem to the user when analyzing stationary or other signals in a fixed frequency band. It is a problem to be dealt with when performing a “tracking” analysis, where for example the sampling frequency is tied to the speed of a machine, and this question is taken up in more detail in Chapter 6, on the analysis of non-stationary signals.

4.3.2. Window Effects

As illustrated in Fig.4.7(d) and (e), the time window applied to the data effectively determines the filter characteristic associated with the analysis. Applying no special window is the same as applying a rectangular window, as illustrated in Fig.4.7, but for stationary signals, and in particular those containing discrete frequency components, this is in general a poor choice of window because of the potential discontinuity where the ends of the record are effectively joined into a loop.

A better choice of window function (for stationary signals) is one which is equal to zero at each end, and whose amplitude varies smoothly along the record length. An excellent general purpose window is known as "Hanning", the name being derived from von Hann, who applied an equivalent process to meteorological data. In the time domain the Hanning window is equivalent to one period of a raised cosine (i.e. cosine squared) function, as illustrated in Fig.4.9 where it is compared with a rectangular window. Their frequency spectra are also compared there, (on log/log scales) from which it is evident that even though the main lobe (and bandwidth) of the Hanning function is greater, the sidelobes fall off at 60 dB/decade rather than 20 dB/decade.

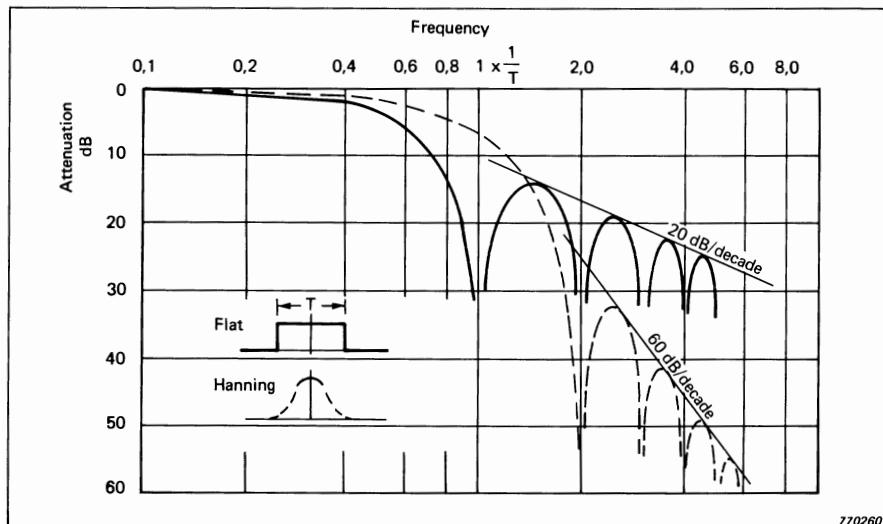


Fig. 4.9. Comparison of the spectra (filter characteristics) of the Flat (rectangular) and Hanning time weighting functions

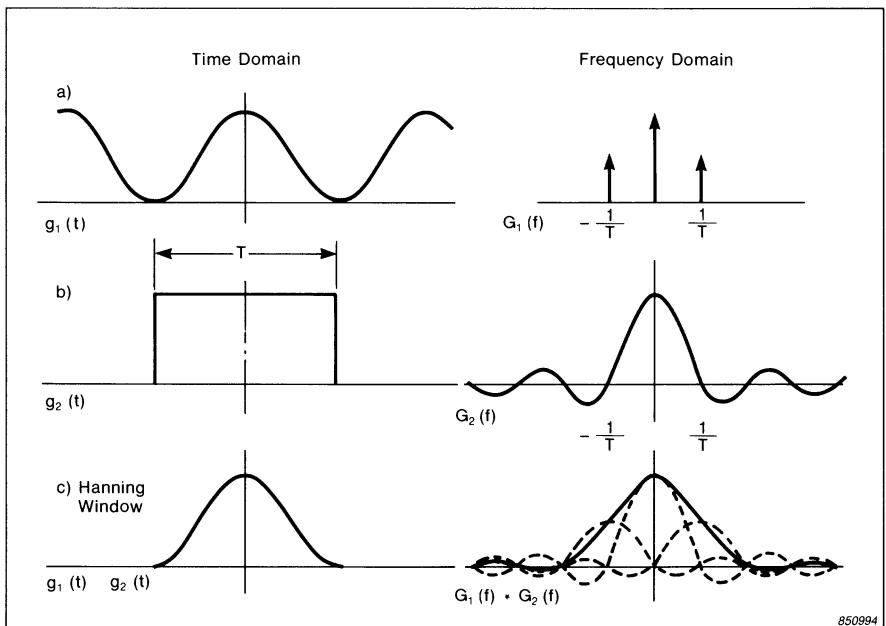


Fig. 4.10. Derivation of the spectrum of the Hanning weighting function

The convolution theorem can be used to explain the reduced sidelobes of the Hanning window as illustrated in Fig.4.10. The Hanning function is considered as being formed by limiting the length of a raised continuous cosine by multiplication by a rectangular window of length equal to one period. The convolution of the two spectra results in three overlapping $\sin x/x$ functions, where the central one has twice the scaling of the other two. The displacement of the two smaller functions is such that their sidelobes are out-of-phase with the sidelobes of the main function, and almost cancel them. The general situation is that one sidelobe of order "n" is counteracted by the sum of two sidelobes (one of order "n + 1", the other of order "n - 1") of half scale. It will be appreciated that the larger the order "n", the more complete is the cancellation, and hence the more rapid rate of fall-off of the resulting sidelobes.

Other window functions may be found useful for special purposes. One is named "Hamming" (directly after its originator) and should not be confused with "Hanning". It consists of a Hanning window on a small rectangular pedestal, scaled so that the first few sidelobes cancel. It has the advantage over Hanning, that the highest sidelobe is at -42 dB (compared with -32 dB) but on the other hand the remaining sidelobes are dominated by the rectangular function and still fall off at 20 dB/decade. The Hamming function probably found

most use in the days when dynamic range was typically 50 dB, and it has now been superseded by a number of others.

The best of these is probably the so-called "Kaiser-Bessel" window, whose highest sidelobes are at -69 dB, and bandwidth 1,8 times the line spacing (compared with 1,0 for rectangular, and 1,5 for Hanning). In theory a Gaussian function (e^{-x^2}) transforms into another Gaussian function with no sidelobes at all, but the Gaussian function is infinitely long and must be truncated in practice. Truncation giving a highest sidelobe of -69 dB gives a bandwidth 1,9 times the line spacing and is thus slightly inferior to the Kaiser-Bessel.

Another window which is useful for a different purpose is the "Flat-top", which is designed specifically to minimise the picket fence effect, and thus facilitate calibration, using a calibration tone which may lie anywhere between two lines of the analyzer.

Comparative properties of the most important windows are given in Table 4.1. It should perhaps be mentioned that where digital zoom is available, it is rarely necessary to resort to exotic windows, at least with stationary signals, since separation of closely spaced components can be achieved by zooming.

Window Type	Highest Sidelobe (dB)	Sidelobe Falloff (dB/decade)	Noise Bandwidth*	Maximum Amplitude Error (dB)
Rectangular	-13	-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
Flat-top	-93	0	3,77	<0,01

* Relative to line spacing

T01254GB0

Table 4.1. Properties of various data windows

The window type is not the only factor affecting the final filter characteristic; it is also influenced by the discrete sampling of the windows in the frequency domain, and by the phase of the signal.

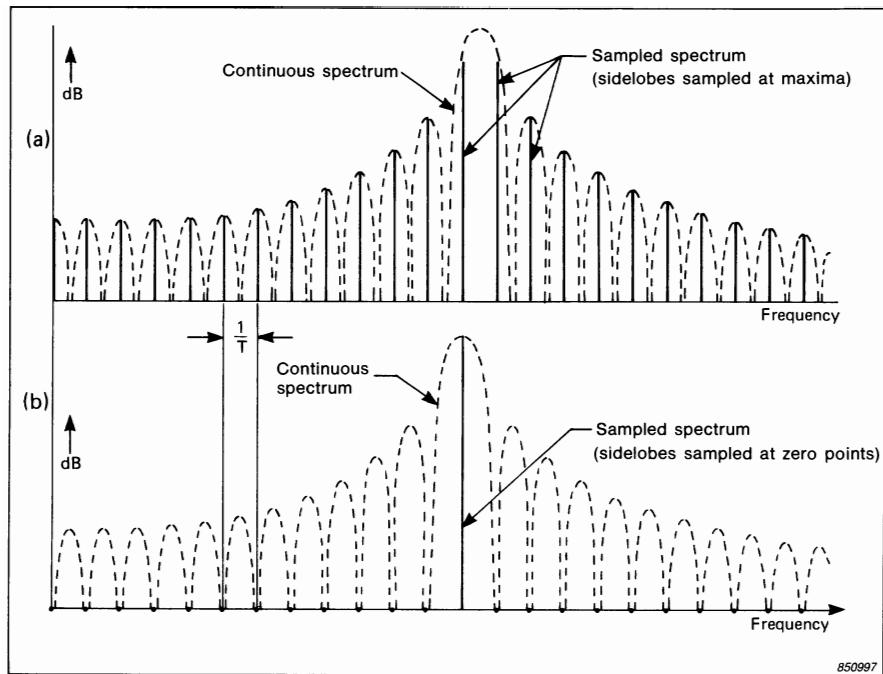


Fig. 4.11. Frequency sampling of the continuous spectrum of a time-limited sinusoid. Number of periods within the time window:
a) half-integer b) integer

The first point is illustrated in Fig.4.11 which shows two extremes for rectangular weighting of a single sinusoid. In Fig.4.11(a) there is a half integer number of periods in the record length, meaning that the peak of the window function falls between two lines of the analysis, and all sidelobes are sampled at their maxima. The other extreme is shown in Fig.4.11(b), which corresponds to an integer number of periods along the record length, so that the peak of the window characteristic coincides with an analysis line, and all other samples fall at the zeroes between the sidelobes. Of course, when an integer number of periods is repeated periodically, the result is an infinitely long sinusoid with only one frequency component.

The second point is illustrated by Fig.2.29, where it is seen that the different phase relationships between a sine and a cosine cause different interactions of the sidelobes from the positive and negative frequency sides, giving reinforcement of some, and partial cancellation of others.

Fig.4.12 (from Ref.4.5) illustrates these two points with actual measurements, and shows the effect of both phase relationships and number of periods.

It should be mentioned that the case illustrated in Fig.4.12(a) is extremely difficult to achieve in practice, except by design, and thus normally a Hanning window is chosen for continuous signals containing discrete frequency components. An exception is where tracking analysis is used, as already mentioned, where the digital sampling is tied to the speed of a machine (or process) and where the number of periods in the memory can be forced to be an integer. The same will apply to all harmonics of the tracking frequency, but it is almost impossible to arrange for other discrete frequencies (for example stemming from a different speed shaft) to satisfy the requirement at the same time.

Before leaving window functions it may be worth pointing out that they can be applied by direct multiplication in the time domain, but it is often more efficient to apply them by convolution in the frequency domain. The convolution coefficients for Hanning weighting, for example, are $-1/4$, $1/2$, $-1/4$, multiplication by which can be achieved by lateral shifts of the binary representations of the numbers, and therefore the process does not involve actual multiplications, only additions. In particular in the case of non-destructive zoom (see section 4.4.2) where the time record is very long, it is more efficient to apply the window function by convolution in the frequency domain rather than multiplication in the time domain.

4.3.3. Picket Fence Effect

The picket fence effect is by no means limited to FFT analysis; it is also found in other situations where discrete fixed filters (as opposed to swept filters) are used, such as in normal $1/3$ -octave analysis.

In general, unless a frequency component coincides exactly with an analysis line, there will be an error in both the indicated amplitude and frequency (where the highest line is taken as representing the frequency component). This can be compensated for, provided it is known (or assumed) that one is dealing with a single stable frequency component. The discussion here is limited to Hanning weighting, but it would be possible to derive the same kind of information for any window function.

For single stable frequency components the frequency and amplitude errors are defined by the difference in dB between the two highest samples around a peak, with the actual frequency obviously lying in between them. Fig.4.13 gives the amplitude correction factor in dB, and frequency error as a proportion of the line spacing, based on Δ dB, the difference in dB between the two highest lines around a peak. For Hanning weighting, Δ dB is a maximum of 6 dB when the frequency coincides exactly with an analysis line, and a minimum of zero dB

Time Signals

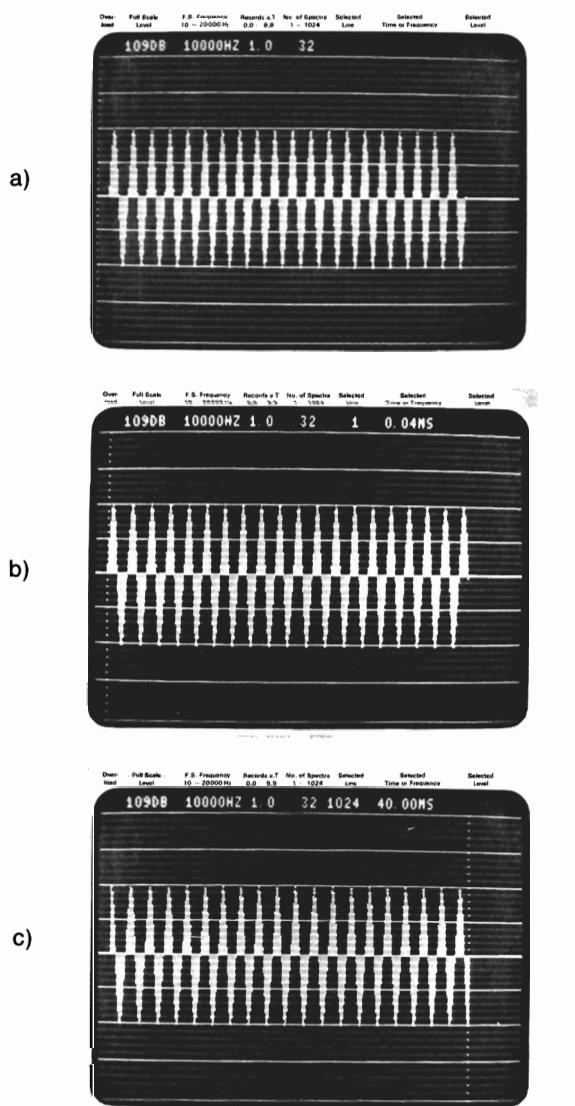
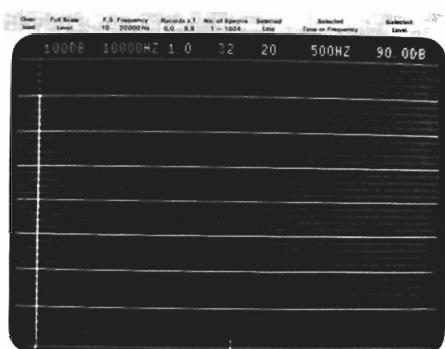


Fig. 4.12. FFT analysis of sinusoidal time signals using different window lengths and phase relationships

(a) Integer number of periods

(b) & (c) Half integer number of periods with different phase relationships

Rectangular Weighting



Hanning Weighting

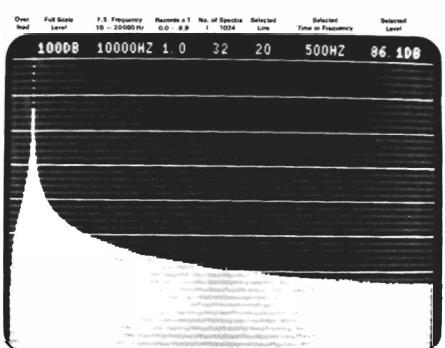
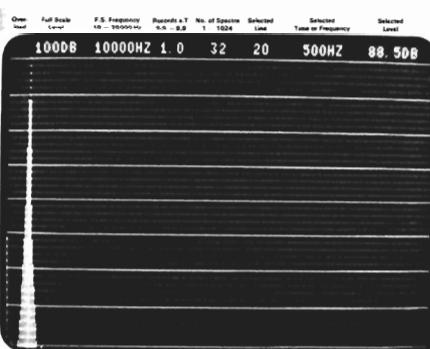
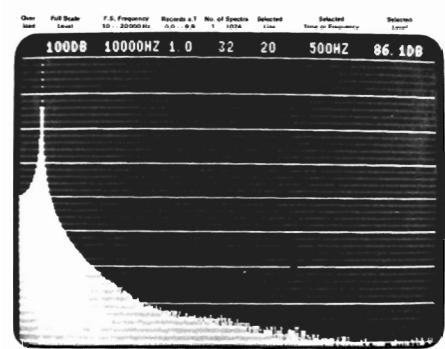
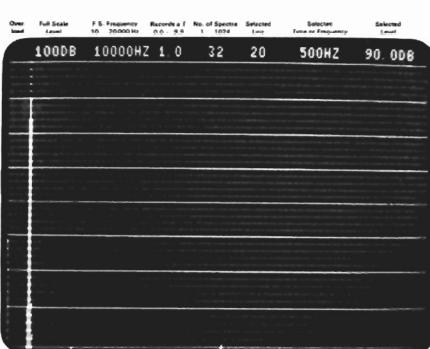
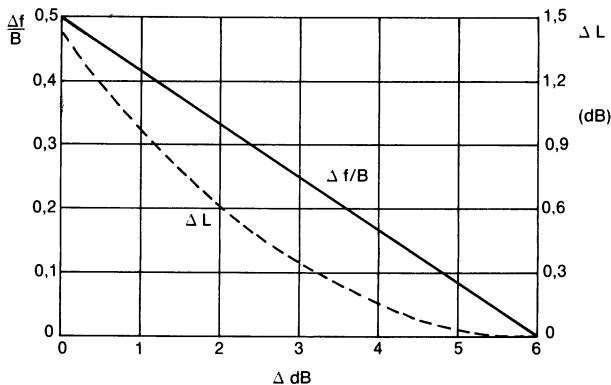
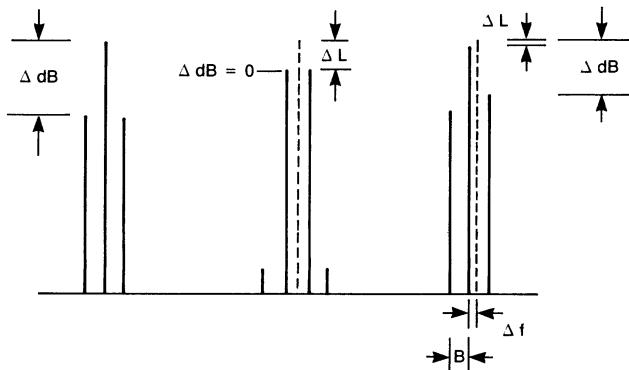


Fig. 4.12. cont.

**Frequency and Level Corrections
for Picket Fence Effect
(Hanning Weighting)**



852090

Fig. 4.13. Amplitude and frequency compensation for picket fence effect with Hanning weighting

when it falls exactly between two lines. A ΔdB of 6 dB occurring anywhere in the upper end of the spectrum is a good indication that interrelated frequency components (e.g. tied to machine speed) are sufficiently stable for these corrections to be valid.

It is also possible to curve fit the Hanning function to the three highest samples around a peak (as in some Brüel & Kjær software for "Harmonic and

Sideband Cursors") and this will even cope with some frequency instability. Note that this technique is used in the 2515 Analyzer to compensate amplitude and phase automatically for picket fence errors, when the Harmonic Cursor is selected.

Using these picket fence correction techniques it is possible to achieve a frequency accuracy approximately 10 times finer than the line spacing, and in a non-destructive zoom analysis (typically 4000 line) this can give a frequency accuracy of approx. 1:20 000.

Where a "Flat-top" weighting or automatic compensation is not available, the amplitude correction factors of Fig.4.13 will be found extremely valuable for calibration with a calibration tone.

4.4. ZOOM FFT

The FFT algorithms discussed so far result in a so-called "Baseband analysis", where the frequency range extends from zero up to the Nyquist frequency f_N , and the frequency resolution is determined by the number of frequency lines up to f_N (normally half the number of original data samples). In certain situations it is desirable to obtain a considerably finer resolution over a limited portion of the spectrum, and the so-called "Zoom-FFT" procedure permits this. It can be considered as "zooming in" on a limited portion of the spectrum with a resolution power corresponding to the number of lines normally used for the whole spectrum (Fig.4.14).

In fact there are two main procedures used for digital zoom, each having certain advantages and disadvantages, so it is interesting to compare the two methods.

To understand the fundamental difference, it is necessary to examine the factors determining the resolution of an FFT analysis:

$$\begin{aligned} \text{The sampling interval } & \Delta t = 1/f_s \\ \text{and record length } & T = N\Delta t = N/f_s \\ \text{Thus, analysis resolution } & \Delta f = 1/T = f_s/N \end{aligned} \tag{4.1}$$

Consequently, the two ways of reducing the resolution Δf , are either:

- (1) Reduce the sampling frequency f_s . This corresponds to so-called "Real-time zoom".
- (2) Increase the record length N . This corresponds to so-called "Non-destructive zoom".

A detailed discussion follows.

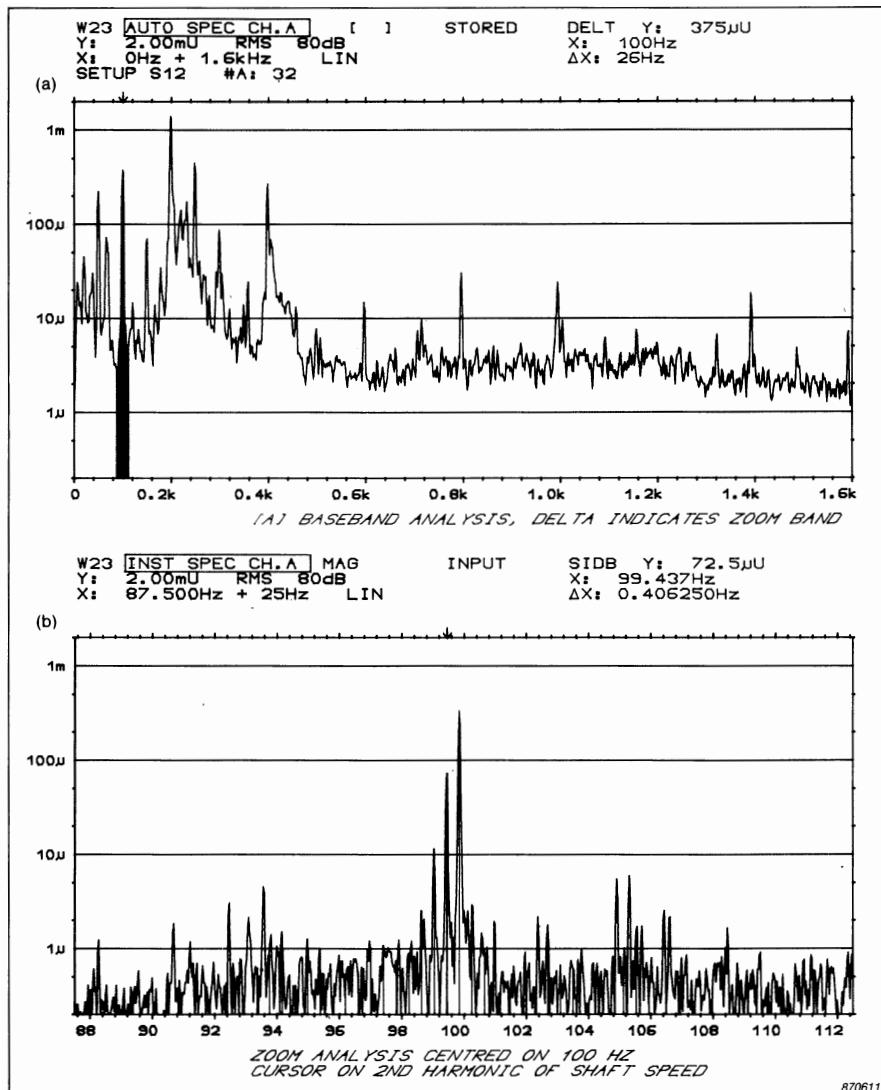


Fig. 4.14. (a) Original baseband spectrum
 (b) Shaded section of (a) "zoomed" by factor 64:1

4.4.1. Real-time Zoom

This technique derives its name from the fact that the signal must be processed in real-time by a zoom processor, in order to shift the frequency origin

to the centre of the zoom band, low-pass filter the signal, and resample as for the digital filtering described in Section 3.5. The final FFT operation does not have to be in real-time.

The basic principles can be understood by analogy with the discussion of Section 2.2.1, on the way in which the Fourier integral (Eqn.(2.10)) functions. As described there, multiplication by a rotating unit vector $e^{-j2\pi f_k t}$ effectively shifts the frequency origin to frequency f_k . The component at frequency f_k is stopped in the position it occupied at time zero, and virtually becomes a new DC component (although in general it is complex). The positive and negative sampling frequencies $\pm f_s$ are likewise moved by an amount f_k , as illustrated in Fig.4.15. (This may introduce aliasing in the negative frequency region, as the new negative Nyquist frequency $(-f_N + f_k)$ may lie higher in frequency than the lowest frequency component.) Note that even if the original time signal $g(t)$ were real, the modified signal would be a sequence of complex values.

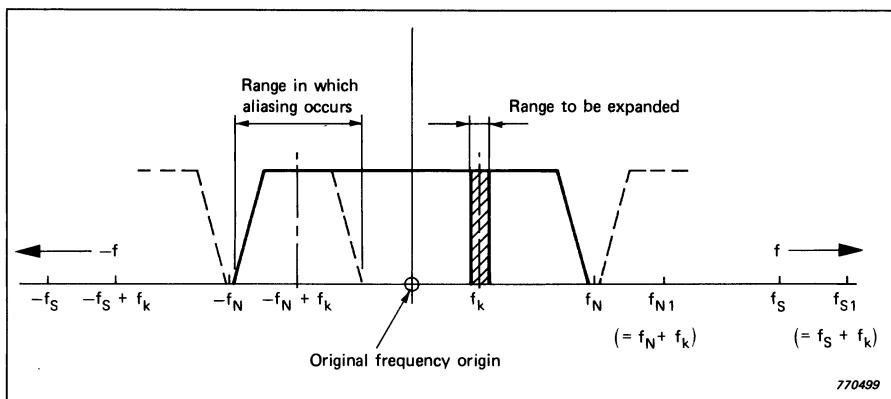


Fig. 4.15. Frequency shift caused by multiplying signal by unit vector rotating at $-f_k$

It is thus possible by multiplying any time signal $g(t)$ by a unit vector rotating at $-f_k$ to change its frequency origin to frequency f_k . The complex signal, thus modified, can then be low-pass filtered (using a digital filter) to remove all frequency components except for a narrow band around f_k , as illustrated in Fig.4.15. Note that at the same time this lowpass filtration would generally remove the portion of the spectrum where aliasing may have occurred. The narrow frequency band remaining after lowpass filtration (the shaded area in Fig.4.15) is shown to a larger scale in Fig.4.16 where it is also made apparent that it is now possible to reduce the sampling frequency while still complying with the sampling theorem. For example, if the total bandwidth after filtering is less than 1/10 of the sampling frequency, it is possible to reduce the sampling

frequency to 1/10 without overlapping in the vicinity of the new Nyquist frequency. In a similar manner to the digital filters discussed in Section 3.5, the reduction in sampling frequency is achieved simply by retaining a reduced number of samples, in this case every tenth, the rest being discarded. The resampled sequence of lowpass-filtered complex samples can be transformed by a (complex) FFT transformation to give the required "zoomed" spectrum.

The above discussion applies strictly to the general complex FFT transform. If the baseband analysis system is designed to produce $N/2$ spectral values from N real data values (Section 4.2), the data memory will only hold $N/2$ complex data values. On the other hand, the complex forward transform (Eqn.(2.18)) gives $N/2$ complex results which are now all valid, because in general there is no symmetry about the new zero frequency (the original frequency f_k). Thus the number of lines resolution in the zoomed spectrum is unchanged. Note that as mentioned in the footnote on p.28 the second half of the frequency spectrum obtained represents the negative frequencies (i.e. the original frequencies below f_k) which should be moved to their correct position before the first half prior to display.

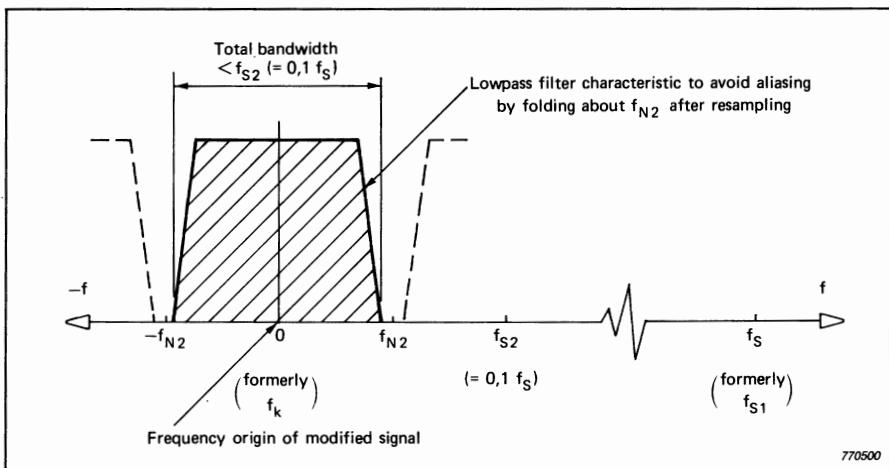


Fig. 4.16. Detail of range to be expanded after resampling

In this case, in order to achieve a zoom factor of 10, the sampling frequency would have to be changed by 20:1, the first 2:1 zoom virtually achieving the same effect as the procedure described in Section 4.2. Note that in the Type 2032 Analyzer (Fig.4.1) both procedures can be used to obtain the same spectral data (i.e. "Baseband" analysis, or "Zoom" in the centre of the baseband range). Even though the spectral result is the same, the data treatment is different, and in particular the analytic time function will be different in the two cases.

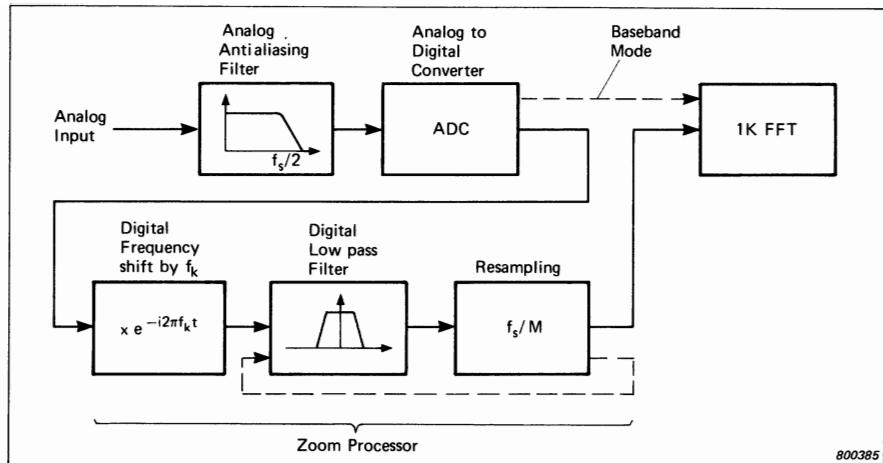


Fig. 4.17. Flow diagram for real-time zoom

800385

In practice, as with the digital filters of Section 3.5.3, it is usually most efficient to low-pass filter and resample in cascaded octave (2:1) steps, meaning that the zoom factors would normally be powers of 2. Fig.4.17 shows a block diagram of the procedure for real-time zoom. Two features should be noted:

- (1) In order to zoom around a new centre frequency f_k , it is necessary to re-process the signal through the zoom processor. This may not be possible unless the signal has been stored.
- (2) The memory buffer required for the time signal is no longer than for baseband analysis. The extra length of time signal required to obtain the finer resolution is achieved by reducing the sampling rate.

4.4.2. Non-destructive Zoom

Non-destructive zoom is simply a way of achieving a large transform size (e.g. 10K or 10240 samples) by repeated application of a smaller transform (e.g. 1K or 1024 samples). For a zoom factor M , it requires a data buffer M times longer than the transform size N .

Fig.4.18 illustrates the basic principles, for a 10K transform. The samples in the data record are numbered 0 through 10239. Ten 1K transforms are performed on records obtained by taking every 10th sample from the original 10K record, first Nos. 0, 10, 20, ..., 10230, then Nos. 1, 11, 21, ..., 10231 etc. until all data values have been transformed. Because of the linearity of the DFT, the

sum of the transforms of the undersampled records must be equal to the transform of the sum of the records. This sum is only equal to the original 10K record when compensation is made for the small time displacements of each of the undersampled records (except for the first). In the Fourier spectra this represents a simple linear change of phase proportional to both the time displacement and the frequency of the component in question (Ref.4.7).

Fig.4.19 is a block diagram of the process, for comparison with Fig.4.17. Note that the zoom accumulator is of limited size, and thus only a selected 400 line section of the resulting high resolution spectrum is generated at any time. If a 4000 line buffer were available it would be possible to generate all 4000 lines from one set of transforms, because even though only 512 complex spectrum results are produced by each partial transform, the entire 5120 lines can be deduced by periodic repetition (the undersampling of the part records leads to this periodicity by aliasing as illustrated in Fig.4.18). It should be noted that any errors introduced by this aliasing cancel out in the final summing operation (by the linearity properties of the DFT).

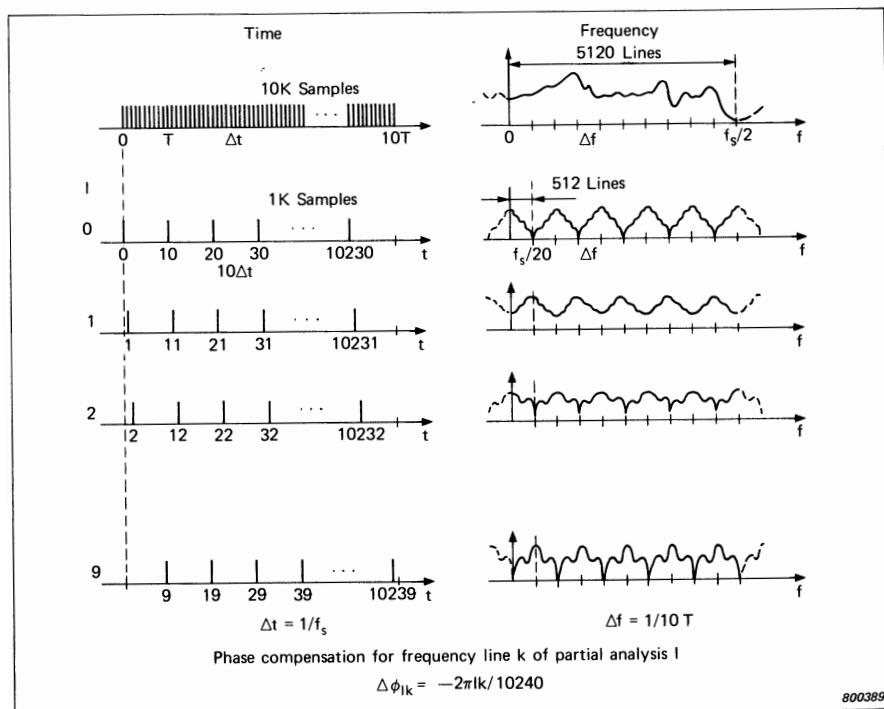


Fig. 4.18. Implementation of a 10 K transform using a 1 K transform ten times

The block diagram of Fig.4.19 also makes the following two features evident:

- (1) Zoom in different frequency regions is based on exactly the same data record.
- (2) A long data buffer is required and the zoom factor is limited by the length of memory.

4.4.3. Comparison of Zoom Techniques

The main advantages of non-destructive zoom accrue from the fact that it is based on exactly the same data record. Thus, it is most valuable for (single channel) signal analysis because even very stable signals vary slightly from one record to another. A typical example would be in gearbox analysis, where one might first want to zoom around the first harmonic (i.e. fundamental) of the toothmeshing frequency and then around the second and third harmonics, knowing that because the same data record is used, there is an exact integer 1,

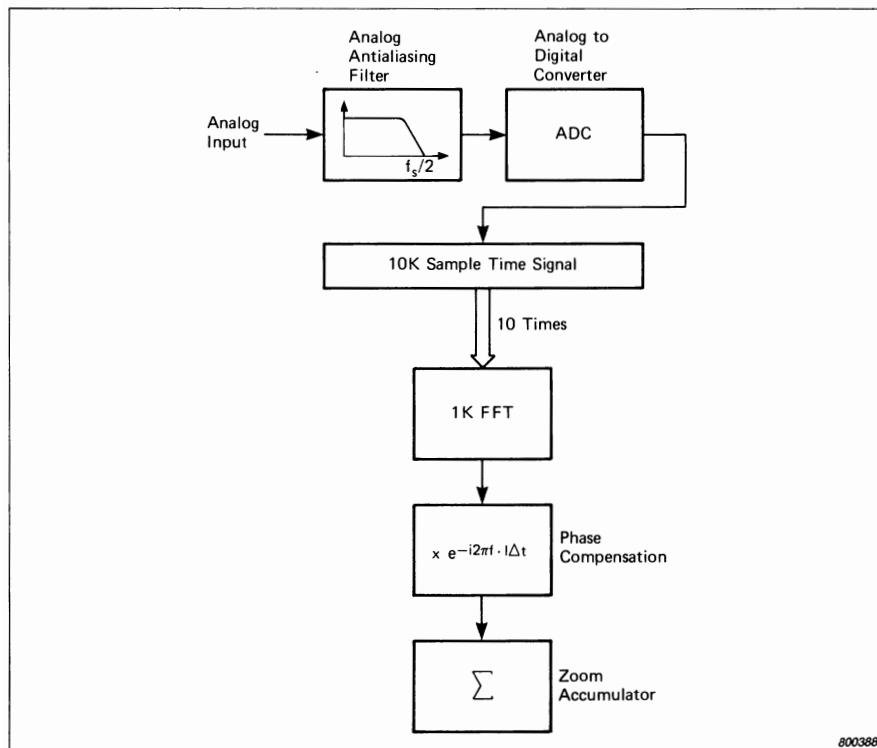


Fig. 4.19. Flow diagram for non-destructive zoom

2, 3 relationship between the frequencies of these components. The same cannot be said for real-time zoom, because for zooming around each harmonic, a different data record is processed.

It is only with non-destructive zoom that frequency component definitions to an accuracy of 1:20 000 (as described at the end of Section 4.3.3) can be achieved.

The disadvantage, that zoom factors are limited by the record length, is rarely a restriction in signal analysis, because it is rarely that individual components are so stable in frequency as to justify a resolution better than 1:4000 or so, for example when tied to machine speed.

A possible exception is equipment with servo-controlled speeds such as Hi-Fi audio equipment.

Larger zoom factors are more relevant for (dual channel) system analysis, where for example frequency response functions can be expected to be stable even if the individual excitation and response signals vary somewhat. Thus, even the results of real-time zoom in contiguous (or overlapping) bands can be expected to match up at the intersections. Zoom is often required in frequency response measurements to achieve adequate resolution of lightly damped resonance peaks. (See Chapter 7.)

4.5. PRACTICAL ANALYSIS OF STATIONARY SIGNALS

4.5.1. Analysis Parameters

Most of the parameters associated with an FFT analysis have already been mentioned, but will be summarised here for ease of reference.

The **frequency range** for baseband analysis is from zero (DC) to the Nyquist frequency f_N (i.e. one half of the sampling frequency f_s), independent of the number of samples in the data record. The actual **useful frequency range** is limited by the anti-aliasing filters, typically to about 80% of the Nyquist frequency (for example 800 lines out of 1024 calculated).

For zoom analysis with factor M , the **frequency span** is M times smaller than the useful frequency range for the corresponding baseband analysis.

The **number of lines** for baseband analysis is related to the transform size N , and is usually $N/2$ up to the Nyquist frequency, and approximately 80% of this within the display range. For real-time zoom, the number of lines is usually the same as for the corresponding baseband analysis, while for non-destructive zoom it is usual to talk about a number of lines corresponding to the larger

data record (e.g. 4000 lines, of which 400 are calculated at a time, corresponding to a record length of 10K).

The **line number** of a spectral line corresponds to the number of periods of that frequency in the data record. Thus, line number 1 corresponds to one period along the record length, and line number 400 to 400 periods.

The **resolution** β , or line spacing of the analysis is simply the useful frequency range or frequency span divided by the number of display lines. It is always equal to the reciprocal of the record length transformed i.e.

$$\beta = \frac{1}{T} = \frac{1}{N\Delta t} = \frac{f_s}{N} \quad (4.1)$$

The **bandwidth** of the analysis is generally greater than the resolution by an amount determined by the time window used. Table 4.1 indicates the bandwidth B in proportion to the resolution β for a number of time windows.

Note that so-called “zero-padding”, which sets the second half of a data record to zero, effectively halves the record length, and thus doubles the bandwidth, if the same type of weighting function is used on the reduced record length. Zero-padding is mainly used in the calculation of time functions such as cross correlation, as discussed in Chapter 7.

4.5.2. Calibration and Scaling

Since this chapter deals with the analysis of stationary signals, it is only necessary to distinguish in scaling between stationary deterministic signals, made up of discrete frequency components, and stationary random signals whose power is distributed in frequency.

Usually, discrete frequency components are scaled such that the correct amplitude of a particular frequency component is read at the centre frequency of a peak, independent of the weighting function used. If the frequency coincides exactly with an analysis line, the value is read directly, otherwise it will be necessary to apply picket fence corrections as described in Section 4.3.3. The “power spectrum” value can be obtained by squaring the amplitude. It is meaningless to talk of power spectral density for discrete frequency components, because in theory they have zero bandwidth.

For stationary random signals the reverse is the case, since it is most relevant to speak in terms of “power spectral density” (PSD) which should give a consistent result, provided the analysis bandwidth is less than the bandwidth of the peaks in the signal itself. One way of checking the latter is to zoom at the peak and confirm that the result does not change.

To calculate PSD, the power spectrum value, obtained as described above, should be divided by the analysis bandwidth B defined as in Section 4.5.1.

Note that the bandwidth associated with each line must be taken into account when integrating power spectra over a frequency band, e.g. for conversion to constant percentage bandwidth or to obtain the total power in the entire spectrum. The same result will be obtained by integrating PSD directly with frequency, or by adding the power values of the individual lines and then compensating for the bandwidth associated with each line. As an example, analyzing a white noise signal with Hanning weighting will produce power spectrum values 1.5 times greater than with flat (rectangular) weighting, but compensating for the bandwidth (Table 4.1) the PSD results will be the same.

Even a spectrum consisting of discrete frequency components must be compensated for the bandwidth factor when integrating over a frequency band. Comparing Hanning with flat weighting, for the case where frequencies coincide exactly with analysis lines, each discrete component, in the case of Hanning weighting, will have two adjacent lines at -6 dB ($\Delta\text{ dB} = 6$) which sum up to give 50% of the power in the central line. It can be shown from Parseval's theorem that the correct result is obtained by summation over all lines even when the frequencies do not coincide with the analysis lines.

4.5.3. Averaging

It has already been shown (Eqn.(4.1)) that the resolution β is always equal to the reciprocal of the record length T . For flat weighting, bandwidth $B = \beta$, and therefore the BT product is always equal to unity for a single FFT transform result. It can be shown (Ref.4.8) that the same applies even when a time weighting function is used, as the increased bandwidth is exactly compensated by the reduced effective length of the time record.

Thus, a higher BT product can be achieved by averaging a number of individual spectra together. The total BT product is simply the number of independent records averaged together, and this value can be inserted (instead of BT_A) in Eqn.(3.9).

Fig.4.20 shows the results of analyzing a stationary random signal with different numbers of averages, and illustrates at the same time the meaning of the standard error ϵ , as calculated by Eqn.(3.9).

For rectangular weighting, independent records simply means non-overlapping, but where a time-weighting function is used, as is normally the case, the situation is a little more complex. The following discussion is based on Hanning weighting, but could be modified for any time-weighting function.

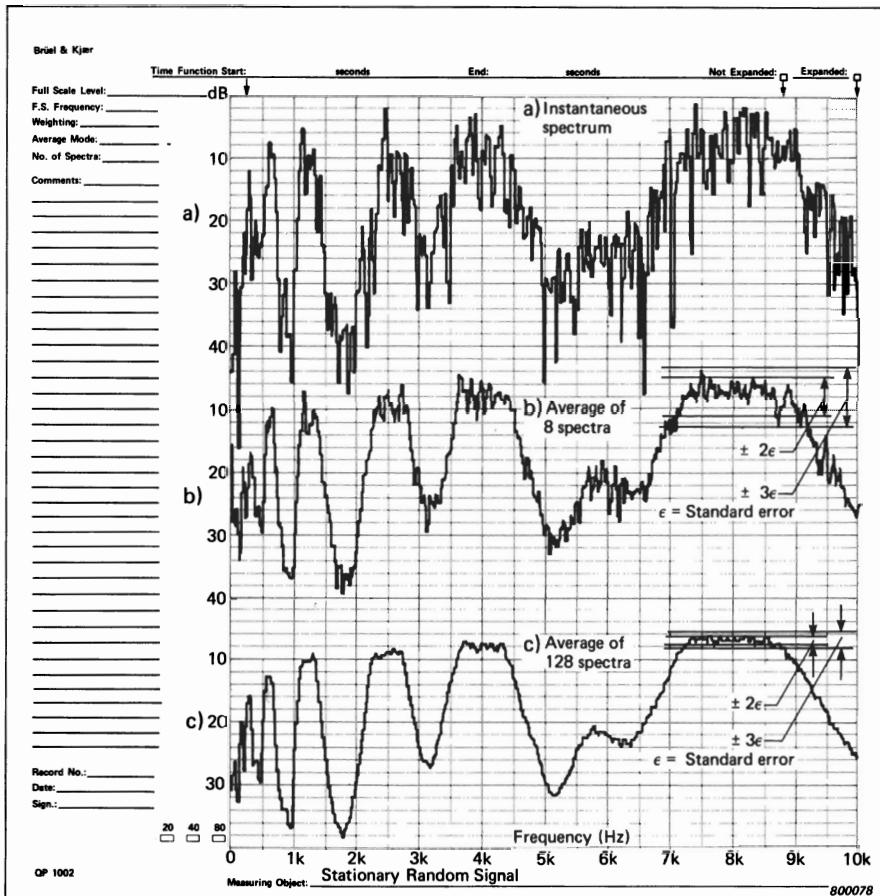


Fig. 4.20. Effect of number of averages with a stationary random signal

Fig.4.21 illustrates that when Hanning weighting is applied to non-overlapping records, virtually half the available data is disregarded. In a statistical sense, most of this information can be retrieved by averaging 50% overlapping records, but the overall (power) weighting function will still be non-uniform, with a 3dB ripple. A calculation method detailed in Ref.4.9 gives a general method for arbitrary window functions, and for 50% overlapping Hanning functions indicates that there is a slight reduction in statistical reliability, but for practical purposes this can be disregarded.

Overlaps greater than 50% generally give a more uniform overall weighting function, but no appreciable improvement in statistical error. It can be shown in fact that the overall weighting function is completely uniform for overlapping

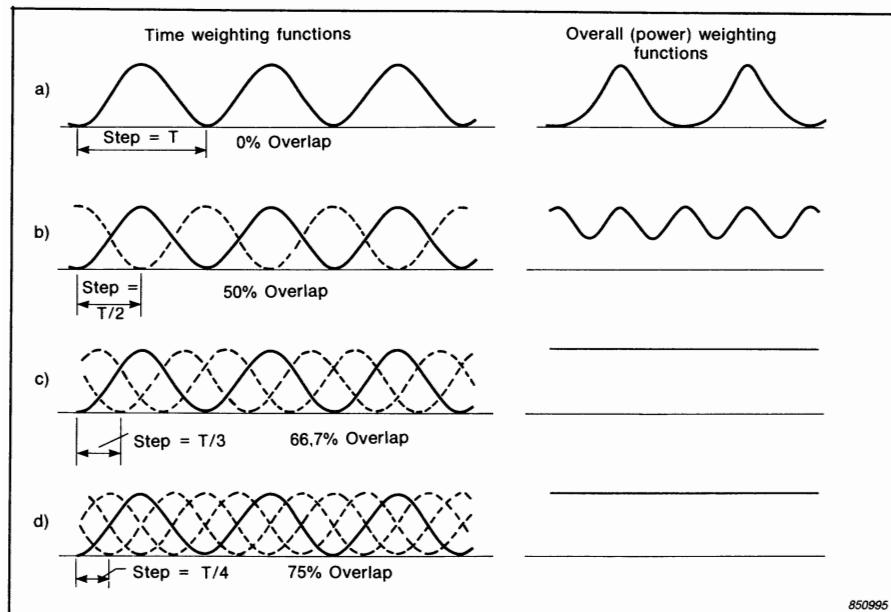


Fig. 4.21. Overall weighting functions for overlapping Hanning windows

Hanning weightings with step lengths equal to $1/3$, $1/4$, $1/5 \dots$ of the record length. The first two of these are illustrated in Fig.4.21. For overlaps greater than 50%, the BT_A product can be taken as twice the number of independent contiguous records in the total signal, i.e. the same as the actual number averaged with 50% overlap.

4.5.4. Real-Time Analysis

The so-called "real-time frequency" of an analyzer is a parameter which is often quoted in specifications, so it is as well to define it, and discuss its importance.

An FFT analyzer is said to be operating in "real-time" when the time taken to Fourier transform and otherwise process each record is less than or equal to the time taken to collect the data transformed. It also implies that data is always being recorded in one memory buffer, even when it is being transformed in another, so that none of the incoming data goes unprocessed.

From the discussion of Section 4.5.3, however, it will be realised that where weighting is used, as is normally the case, then effectively half the data will not be taken into account when processing non-overlapping records. To overcome

this problem, it is necessary to process 50% overlapping records, which involves twice as much processing for the same data collection time, and thus the actual real-time frequency is closer to half that usually stated.

In fact, to obtain results the equivalent of a true real-time analysis with parallel filters (analogue or digital), the overall weighting function must be uniform, and thus the overlap must be at least $\frac{2}{3}$, giving a real-time frequency $\frac{1}{3}$ of that usually stated.

On the other hand, it is only very rarely that a true real-time analysis is required. For stationary signals, for example, any data which is missed is statistically the same as that processed, and so the only advantage of a real-time analysis is that it achieves a result with a given accuracy in the minimum possible time.

Even for non-stationary signals, it is not always necessary to analyze in real-time; it depends on how quickly the signal itself changes. For machine run-ups and run-downs, for example, the inertia of the rotor would often limit the rate at which the signal itself can change, and it is only necessary that successively analyzed records are not greatly different from each other.

It is only in the case of rapidly changing non-stationary signals, such as speech, that it may be necessary to analyze in real-time, so as not to lose any information. It should be emphasized that this implies that the results must be further processed at the same rate, at the very least by averaging them together. Note that such averaging would smear out the variability of the non-stationary signal, which was presumably of interest. Storing all the results in a digital memory does not give much advantage over storing the original signal, and then post-processing it.

The main use of the real-time frequency specification is as a comparative indicator of analysis speed, and to give an indication of the time required to obtain a result with a given number of averages (for frequency ranges above the real-time frequency).

4.6. Hilbert Transform Techniques

As described in Section 2.6, the Hilbert Transform of a time signal can be carried out by Fourier transformation into the frequency domain, modification of the phase, and inverse transformation back to the time domain (Fig.2.32). These operations can readily be carried out in an FFT analyzer, and for example in Type 2032 (Fig.4.1) the Hilbert Transform is generated for all time functions. One way of generating an analytic signal, is by making use of the fact that its spectrum is one-sided, and of double amplitude with respect to that of the real part alone (Fig.2.36) and this operation can also be carried out

in an FFT analyzer. For single channel analysis the major applications of the Hilbert Transform are to do with demodulation (Section 2.6.2), both amplitude demodulation and phase (or frequency) demodulation. As given in Eqn. (2.55), a generally modulated signal can be represented as the real part of the analytic signal $A(t) e^{j\phi(t)}$, where $A(t)$ represents the amplitude modulation function (plus DC offset) and $\phi(t)$ represents the phase modulation signal (plus carrier frequency component).

A given measured signal $a(t)$ can be demodulated by the following process:

- (1) Calculate its Hilbert transform to give $\tilde{a}(t)$
 - (2) Form the analytic signal $a(t) + j\tilde{a}(t)$
 - (3) Decompose this into its amplitude and phase components $A(t) e^{j\phi(t)}$.
- $A(t)$ is then the amplitude modulating signal (including DC offset).
- (4) Multiply $e^{j\phi(t)}$ by $e^{-j2\pi f_o t}$ to remove the carrier frequency component f_o . The resulting phase function $\phi(t) - 2\pi f_o t$ is the required phase modulation signal. If it is expressed modulo 2π it may require "unwrapping" to give a continuous signal, $\phi_m(t)$.
 - (5) Calculate $\frac{1}{2\pi} \frac{d\phi_m(t)}{dt}$ to give the frequency modulation signal $f_m(t)$.

If the demodulation is to be carried out on a band-pass filtered signal, the procedure can be modified slightly, in that zoom FFT can be used for the bandpass filtration. The complex time signal produced by the zoom processor (Fig.4.17) is obtained from the positive frequency components only, and thus automatically represents an analytic signal, frequency shifted by subtraction of f_k , the centre frequency of the zoom band. It is thus already equivalent to the results of step (4) above, in particular if the zoom centre frequency f_k is made equal to the carrier frequency f_o . Even in the case of amplitude demodulation, where the carrier signal is not of interest, the frequency shift has no effect on the amplitude function $A(t)$.

Some examples will now be given of the practical application of amplitude and phase demodulation, as carried out using an FFT analyzer.

4.6.1. Amplitude Demodulation

A situation where amplitude demodulation is useful is in the analysis of signals containing a near-periodic series of high frequency bursts. Such signals arise from the excitation of high frequency resonances, for example by repeti-

tive impulsive forces in a reciprocating machine, or faulty rolling element bearings. The interesting diagnostic information is often contained in the repetition frequency of the bursts, rather than in their frequency content, which would usually be a composite of all resonance frequencies excited.

Fig.4.22 illustrates for a simple model (where only one resonance is excited) the results of frequency analyzing such a signal (vibration acceleration assumed). The dotted line represents the energy spectrum of one pulse (representing the frequency response of a single-degree-of-freedom system). If the pulses were identical, and perfectly uniformly spaced, the spectrum of the periodically repeated bursts would be a line spectrum comprising all harmonics of the repetition frequency $1/T$, with the largest values in the vicinity of the resonance frequency. The repetition frequency could in principle be obtained by zooming in this range and measuring the harmonic separation. In practice, however, there are slight variations between the pulses and also in their spacing, so that the higher order harmonics tend to broaden, and eventually merge. As an example, a speed fluctuation of 0,1%, would cause merging in the vicinity of harmonic no. 1000.

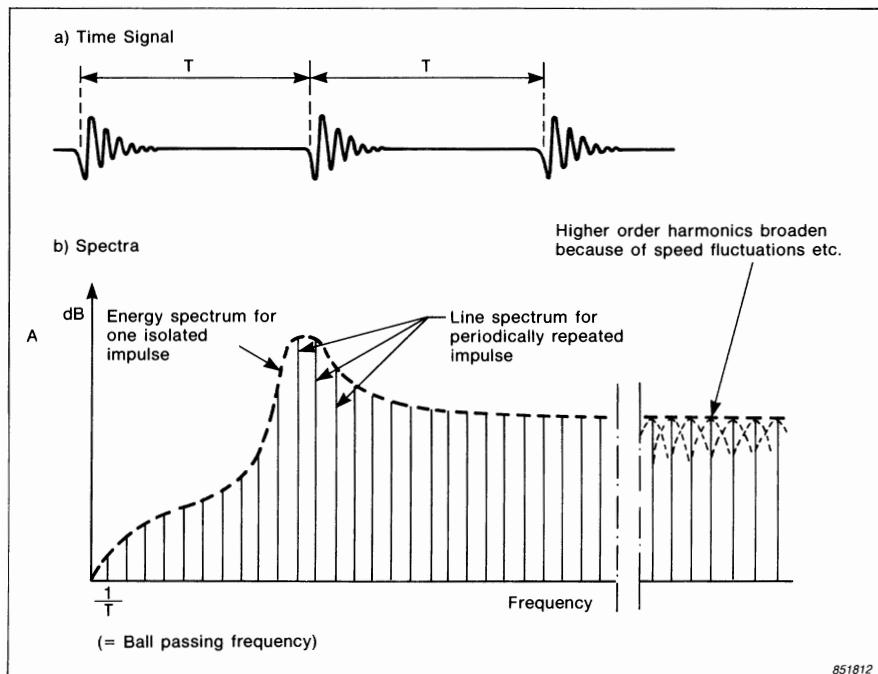


Fig. 4.22. Frequency spectra for an idealised case with repetitive impulses

If, however, the envelope of the original signal is formed by amplitude demodulation, then frequency analysis of the envelope signal will reveal the repetition frequency directly. A speed fluctuation of 0,1%, for example, would only result in a broadening of this component to 0,1% of the centre frequency.

Figs.4.23 to 4.25 show the application of this procedure to the diagnosis of a fault in a roller bearing in a paper mill. Fig.4.23 compares baseband spectra (to 12,8 kHz) for a good bearing and one with a localised fault in the outer race. In the baseband spectra, the fault is revealed by a moderate increase in spectrum levels in the vicinity of 5,4 kHz, evidently a resonance frequency excited by the fault. Figure 4.24 shows, however, that a zoom analysis in this range does not reveal any harmonic structure (harmonic order > 300). The envelope analyses in Fig.4.25, obtained by zooming in the same frequency range, do however reveal a periodic burst structure. In the signal from the good bearing, signals from pneumatic lubricators (at approx. 5,5 Hz) are apparent, while in the signal from the faulty bearing an additional harmonic series corresponding to the ball-pass frequency (15,4 Hz) is also present. This is a typical example where the required diagnostic information cannot be obtained by frequency analysis of the raw signal.

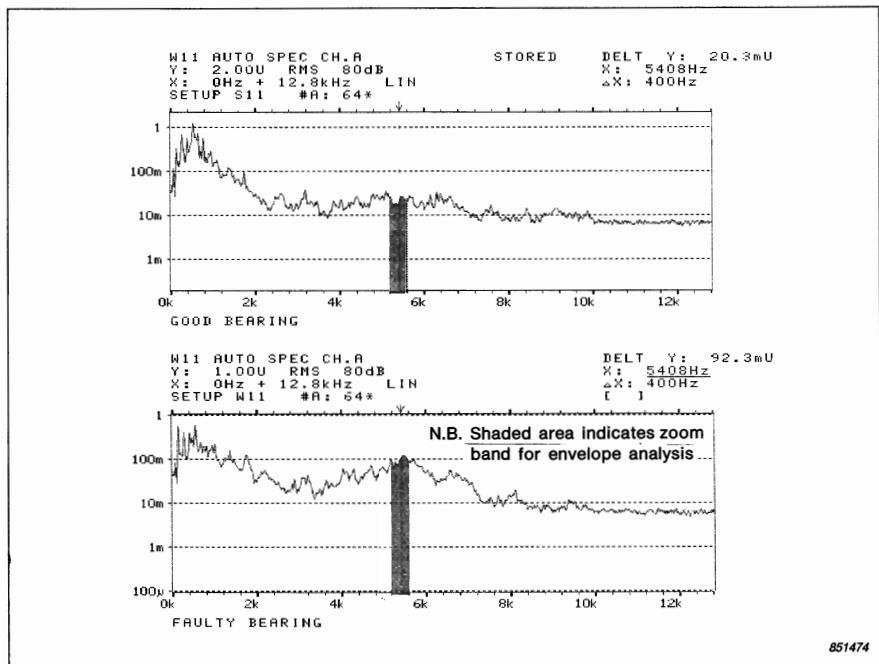


Fig. 4.23. Baseband spectra with and without a bearing fault

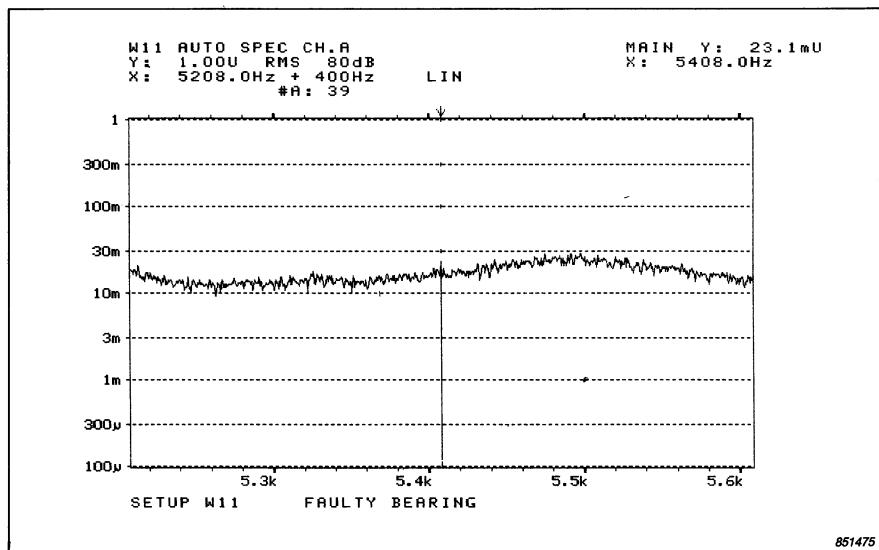


Fig. 4.24. Zoom analysis in same band as used for envelope analysis

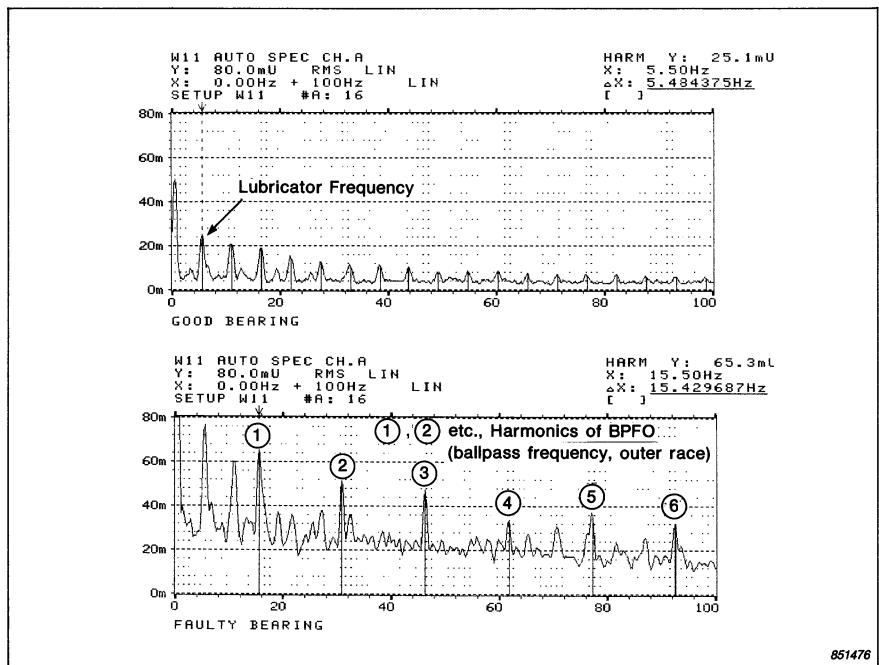


Fig. 4.25. Envelope analysis revealing a bearing frequency

4.6.2. Phase Demodulation

A situation where phase demodulation is of interest is in the analysis of torsional vibrations. If an accurately machined toothed wheel is mounted on the shaft in question, and a magnetic or other proximity probe arranged to give a pulse for the passage of each tooth, then phase demodulation of the tooth passage signal gives a direct measure of torsional oscillations.

In this case, the modulated function is not a sinusoid, but a series of pulses, having a large number of harmonics of the fundamental tooth-pass frequency. Each harmonic is modulated by exactly the same phase modulation signal (in terms of shaft phase angle), although the relative modulation is higher, the higher the order of harmonic demodulated.

In Fig. 4.26, it will be seen that it is only possible to demodulate the signal around one harmonic, when the sidebands do not overlap with those around the other harmonics. In practice this means that the maximum zoom band would be equal to the carrier frequency f_c , and that the modulation sidebands do not extend to more than $f_c/2$ on either side. Where the phase deviation β is less than 1 radian, the phase modulated signal can be described adequately with two pairs of sidebands, and thus the maximum number of harmonics which could be represented would be half the number of sidebands, or $1/4$ of the number of teeth on the toothed wheel.

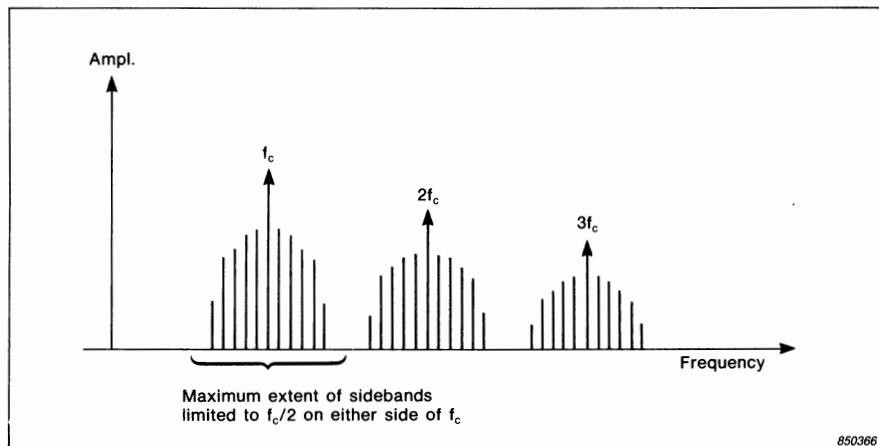


Fig. 4.26. Sideband limitations in demodulating a series of pulses with fundamental frequency f_c

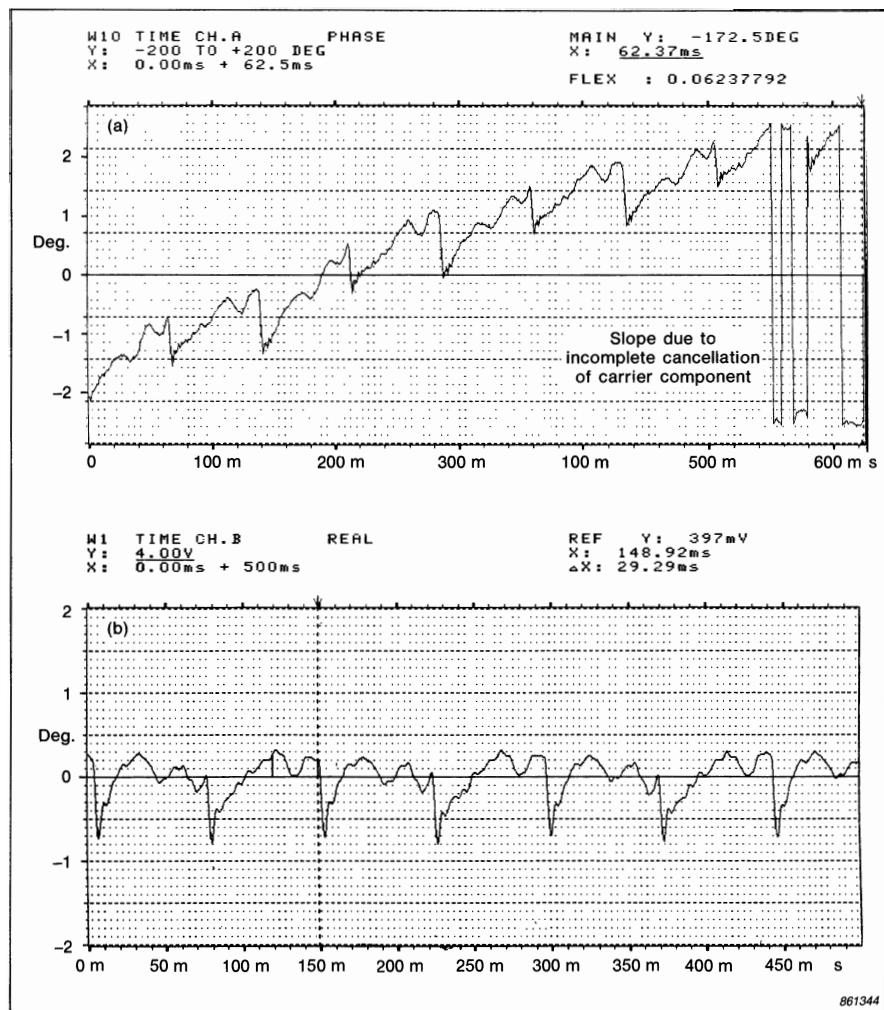


Fig. 4.27. Torsional vibration signal for a diesel engine
 (a) Phase of the analytic signal
 (b) Measured by frequency demodulation and integration

The phase deviation β represents the phase deviation of the carrier component, so that in terms of the torsional oscillations of the shaft, it must be divided by the harmonic order of this carrier with respect to the shaft speed. Thus, for demodulation of the toothpass fundamental, with a 60-tooth wheel, the $\pm 180^\circ$ scale on which the phase would normally be displayed represents $\pm 3^\circ$ of shaft oscillation. It would represent $\pm 1.5^\circ$ if the second harmonic were demodulated,

and so on. This provides a means of changing the scaling of the result, though it should be kept in mind that if β becomes larger than 1 radian, more sidebands are required to represent the modulation, and the number of shaft harmonics which can be accommodated reduced accordingly.

Fig.4.27(a) shows an example where the phase of the analytic signal obtained from passage of a 63-tooth wheel is displayed. In this case it was not possible to make the zoom centre frequency f_k coincide exactly with the tooth-pass frequency f_o , and the torsional oscillation signal is superimposed on a small slope represented by $2\pi(f_o - f_k)t$. Fig.4.27(b) shows the results obtained from the same signal by an instrument designed specifically to produce such torsional vibration signals (by frequency demodulation and integration). The two results are very similar, confirming the validity of the procedure given here.

One way of removing the slope of the phase signal such as that given in Fig.4.27(a) would be to use "tracking" (see Chapter 6). In this case, the analyzer sampling frequency is generated from the rotational speed of the shaft in question and it is possible to force the tooth-pass frequency to coincide exactly with an analysis line, which then can be used as a zoom centre frequency. The mean slope can alternatively be removed numerically.

5. TRANSIENT ANALYSIS

Transients, in particular short transients, are one of the classes of non-stationary signals described in Section 2.4. There is no clear distinction between transients and other non-stationary signals, but here they are defined as signals of finite duration which are to be analyzed as an entity, with no regard to how they vary with time. Hence, they are considered to have a finite amount of “energy” (amplitude squared integrated over time) and the spectrum is most readily expressed in terms of “energy spectral density”. Their length would generally be of the order of, or even shorter than, the impulse response of the physical systems to which they are applied or from which they result.

Before the advent of FFT analyzers, transients were analyzed by analog techniques, and these are treated in Section 5.2, but nowadays FFT techniques are by far the most powerful, and commonly used, and are thus treated first. Analysis using digital filters is quite efficient, and sometimes has advantages over FFT techniques; digital filters are treated along with analog filters in Section 5.2.

5.1. FFT TECHNIQUES

5.1.1. Short transients

FFT analysis of transients is particularly straightforward when the entire transient fits into the transform size T without loss of significant high frequency information (T is inversely proportional to the frequency range selected for the analysis, see Eqn. (4.1)).

Figure 5.1 shows a typical example (the acceleration response of a simple structure to a hammer blow) analyzed using the 2032 Analyzer (record length 2 K, or 2048 samples). The signal has effectively died away to zero by the end of the record (Fig.5.1(a)), while the frequency range chosen (3.2 kHz) includes all relevant frequency information (Fig.5.1(b)). The units of the time signal (here denoted U) are m/s^2 , and the spectrum is depicted as an RMS spectrum with the same units. This scaling is however dependent on the analysis parameters,

and should be converted to energy spectral density (ESD) as follows:

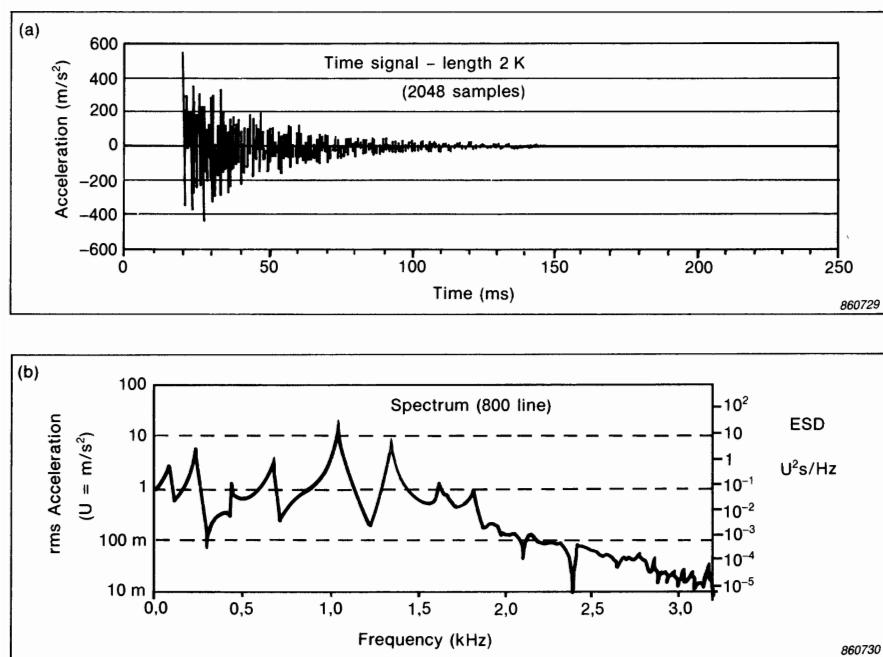


Fig. 5.1. FFT analysis of a short transient

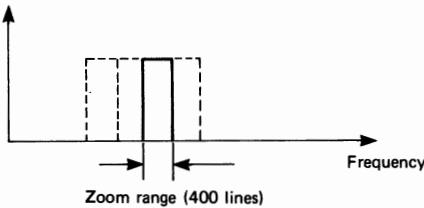
- a) Time signal – length 2 K (2048 samples)
- b) Spectrum (800 line)

Because rectangular weighting has been used, the bandwidth B is equal to the line spacing, or resolution, β (Table 4.1), which in this case is 4 Hz (3,200 Hz/800). The record length $T (= 1/\beta) = 250$ ms in this case. The squared RMS values of the left-hand scale must thus be multiplied by the record length T to convert them from "power" to "energy", and divided by the bandwidth B to obtain ESD in the units $U^2\text{s}/\text{Hz}$. This corresponds to a multiplication by T^2 , in this case $1/16\text{s}^2$, and this scale is inserted on the right of Fig.5.1(b). The 2032 analyzer can in fact scale in terms of ESD directly, but since this possibility is not included in all analyzers, the detailed procedure has been given here. Rectangular weighting should normally be used in cases such as this, provided the signal within the record length starts and finishes with zero (thus eliminating any discontinuity where the ends effectively are joined into a loop, an inherent property of the FFT process). When making impact measurements on lightly damped structures, it is common to apply two special windows to the signals for the following reasons:

1. The force signal is known to be very short, and a short rectangular window can be applied to it. This will eliminate noise from the rest of the record, thus improving the signal/noise ratio of the result.
2. For lightly damped responses, an exponential window can be applied to force the signal to near zero by the end of the record. It must be recognized, however, that the results of the analysis will be modified by the extra damping which this represents; the added damping is very precisely known, and can be subtracted from the measured results.

5.1.2. Longer Transients

- a) Frequency window method (ZOOM-FFT)
 e.g. a 10K record length allows generation of a 4000 line spectrum from ten 400-line zoomed spectra. (A 10 times smaller bandwidth requires a 10 times longer record).



- b) Time window method
 "Scan" average with overlapping Hanning windows

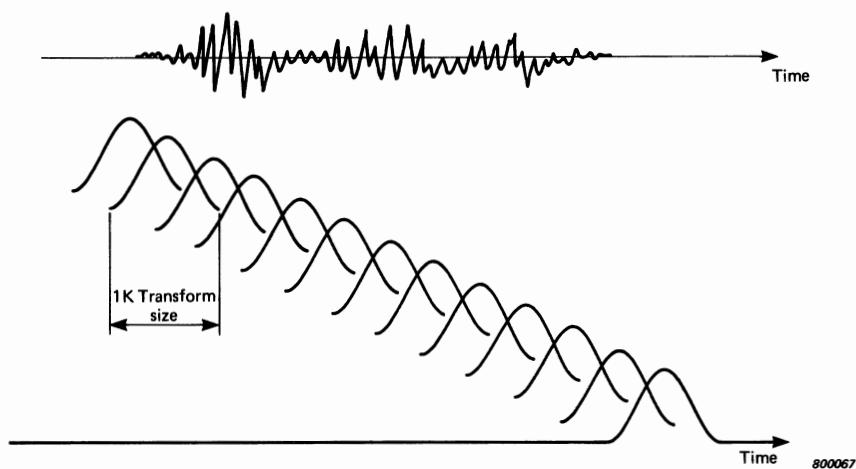


Fig. 5.2. Two methods of analyzing a long transient with an FFT analyzer

When the transient is longer than the normal transform size for the analyzer, the analysis can still be performed in one of the ways illustrated in Fig.5.2. It is simplest if the transient is first recorded in a long memory in the analyzer, such as would be required for non-destructive zoom (Section 4.4.2).

Using the **frequency window** method (Fig.5.2(a)), zoom is used to increase the record length T to greater than the length of the transient.

To obtain the complete analysis, a number of zoom analyses in contiguous frequency bands must be made, for which reason non-destructive zoom is most appropriate. Where only real-time zoom is available, the signal would have to be recorded in an external medium and played back for each zoom analysis. Because the signal starts and finishes with zero in the (extended) record length, rectangular weighting should be used. Conversion of the results from mean square values to ESD is the same as in the preceding section (multiplication by T^2) using a value of T equal to the extended record length ($1/\beta$, where β is the line spacing for the zoom analysis).

Figure 5.3 shows the results of analyzing such a signal (obtained by impacting a lightly damped bell-like object). The results were obtained using repeated non-destructive zoom analysis with the 2033 analyzer. This analyzer has a 10 K (10240 sample) memory, giving a zoom factor of 10 in conjunction with the normal 1 K transform size. Figure 5.3(a) shows the (envelope of the) 10 K time record of length 2 s, as analyzed in the 2 kHz frequency range. This was plotted using the 7509 Controller for the 2308 X-Y Recorder. It is seen that the transient is considerably longer than the 1 K normal transform size, and thus the method of Section 5.1.1 could not be used. Fig.5.3(b) is a composite of 10 separate 400-line zoom analyses in contiguous 200 Hz bands (indicated by the dotted lines). Note that the results match up perfectly at the junctions of the frequency bands, because exactly the same data record was used for each. The left-hand scale shows RMS values of acceleration ($U = \text{m/s}^2$) while the right-hand scale has been converted to ESD in $\text{U}^2/\text{s}/\text{Hz}$ by multiplying mean square values (U^2) by T^2 (in this case 4s^2).

Using the **time window** method (Fig.5.2(b)), a time window such as Hanning is moved in overlapping steps along the signal (so-called scan analysis), and the results of each part analysis averaged together. As shown in Section 4.5.3 (Fig.4.21) the results of such a "scan average" will have uniform weighting (at least along the centre portion of the record) if the step length is made equal to T/n where T is the transform size (defining the length of the Hanning window) and n is an integer ≥ 3 . It is often most practical to use $n = 4$ (75% overlap) as the transform size is always a power of 2, and therefore not divisible exactly by 3. Even though the analysis can be correct in the sense that it has uniform weighting, the analysis bandwidth will be determined by the Hanning window, and may therefore be greater than that of the signal itself (now longer than the window). It is thus not certain that the result can be validly converted to ESD,

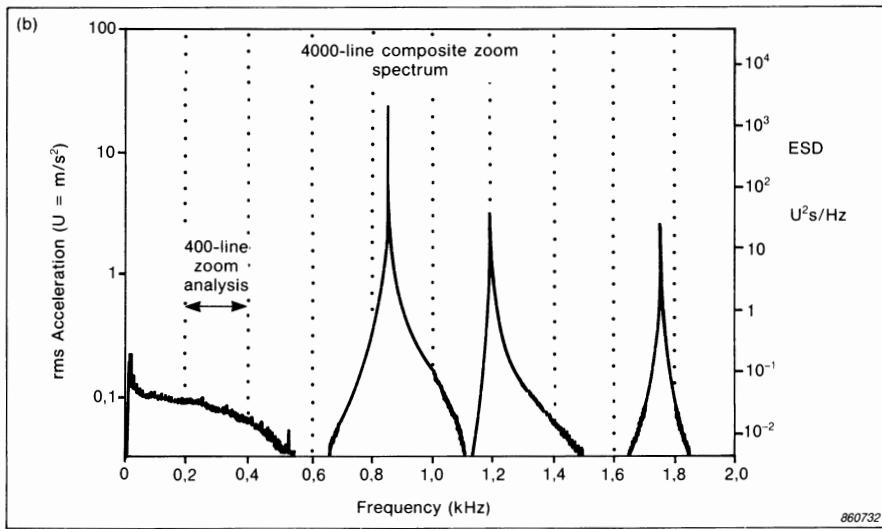
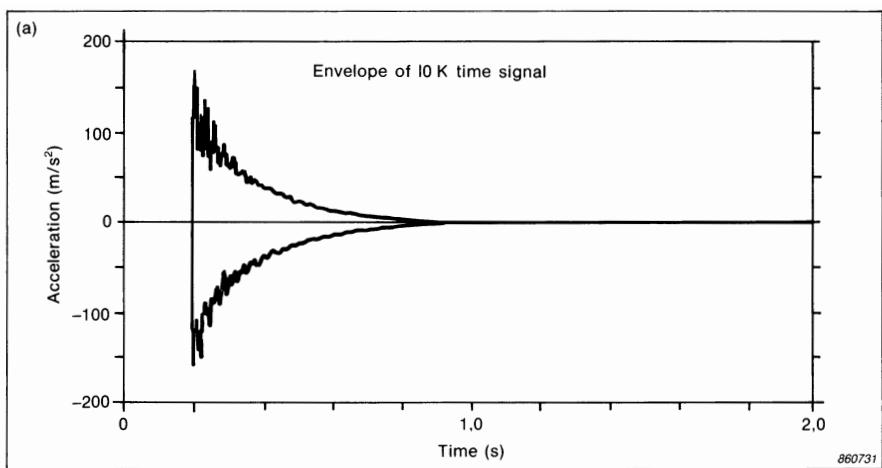


Fig. 5.3. Analysis of a long transient using repeated (non-destructive) zoom
a) Envelope of 10 K time signal
b) 4000-line composite zoom spectrum

although the broader bandwidth does increase the dynamic range of the result in comparison with the results obtained by zoom.

Figure 5.4 shows the results of a scan average analysis on exactly the same signal as in Fig.5.3, and illustrates these points. For most of the spectrum, the RMS values (left-hand scale) are 12 dB higher than the equivalent values obtained by zoom. The ESD scale on the right is correspondingly offset by this amount (details later). The peaks, however, differ by less than 12 dB, indicating that the ESD scaling is invalid where the intrinsic bandwidth of the signal is less than the analysis bandwidth. (The results could, however, be interpreted as "mean ESD in the band").

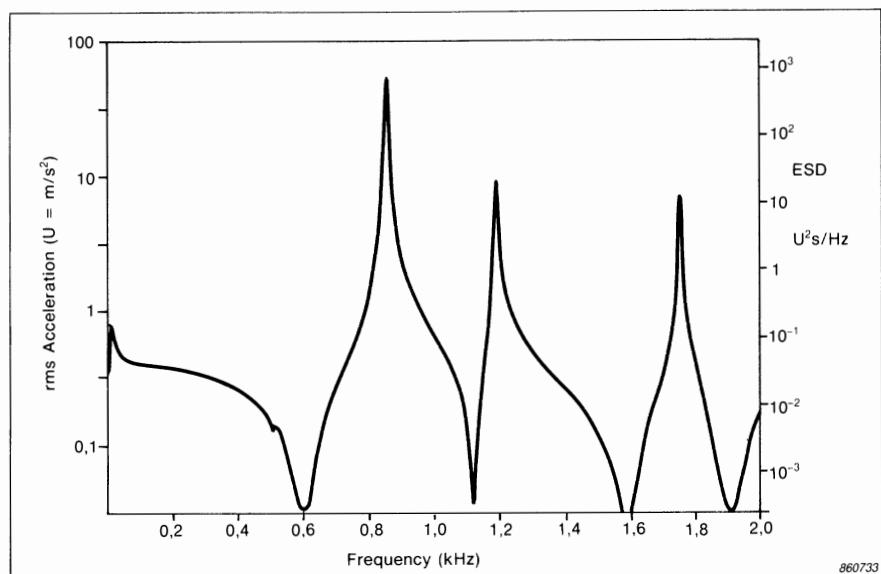


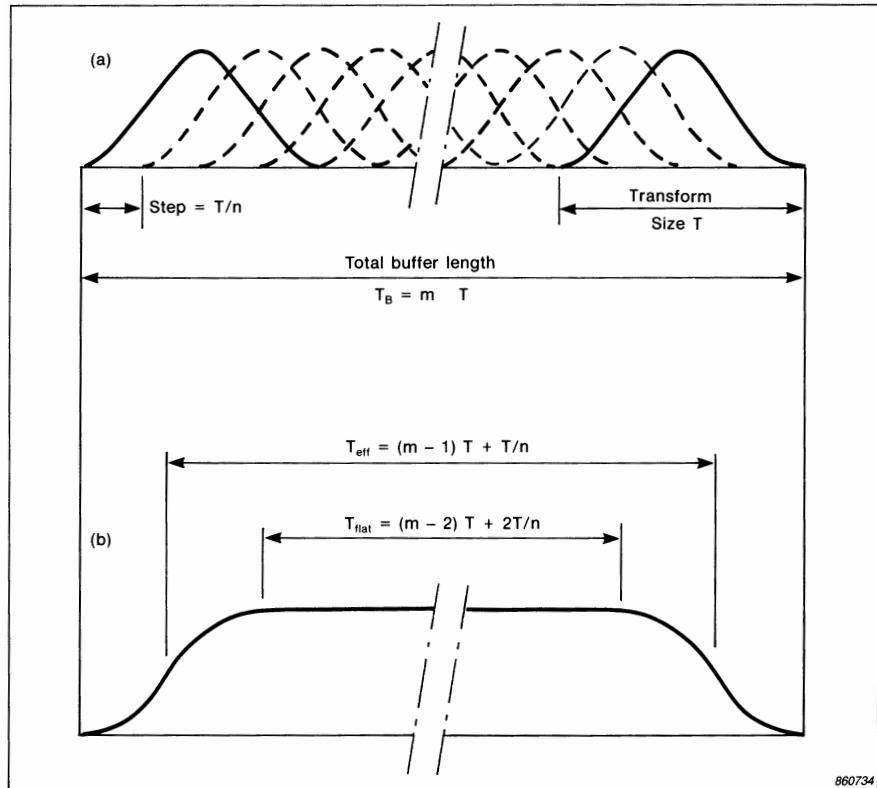
Fig. 5.4. Analysis of a long transient using scan average on the same signal as in Fig.5.3(a)

The ESD scaling is achieved as follows:

First, the mean square values should be multiplied by the "effective record length" to convert them from "power" to "energy". In this case, it is somewhat less than the total record length, because of the non-uniform weighting at the ends of the record. Figure 5.5 illustrates the general overall weighting function which varies according to the amount of overlap used in the scan average (Ref.5.1). In this case, 75% overlap (step length $T/4$) was used, in which case the effective record length is 9,25 K. Finally, the "energy" results should be divided by the analysis bandwidth (1,5 line spacings for Hanning weighting) to convert from energy to energy spectral density.

In this case the overall difference in scaling factor between zoom (with factor 10) and scan average is:

$$10 \log_{10} (10 \times 1,5 \times \frac{10}{9,25}) = 12,1 \text{ dB}$$



860734

Fig. 5.5. 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 & 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

Note that it can be seen by inspection whether the ESD scaling is likely to be valid. Spectrum peaks should be roughly 3 times broader than the analysis bandwidth, and in this case that means a 3 dB bandwidth of 5 lines or more. It can be seen that this condition is not satisfied here.

Another point to be noted is that the results will only be fully correct if the signal is entirely located in the section with uniform weighting. From Fig.5.5 it will be seen that this section is shorter than the "effective record length", and it may be necessary to adjust trigger delay times so that the signal is zero for the first portion of the record (this has been done in Fig.5.3(a) even though it was not strictly necessary for the zoom analysis).

Specialised analysis of transients is also considered in later chapters, for example cepstrum analysis in Chapter 8, and both correlation and system response analysis in Chapter 7.

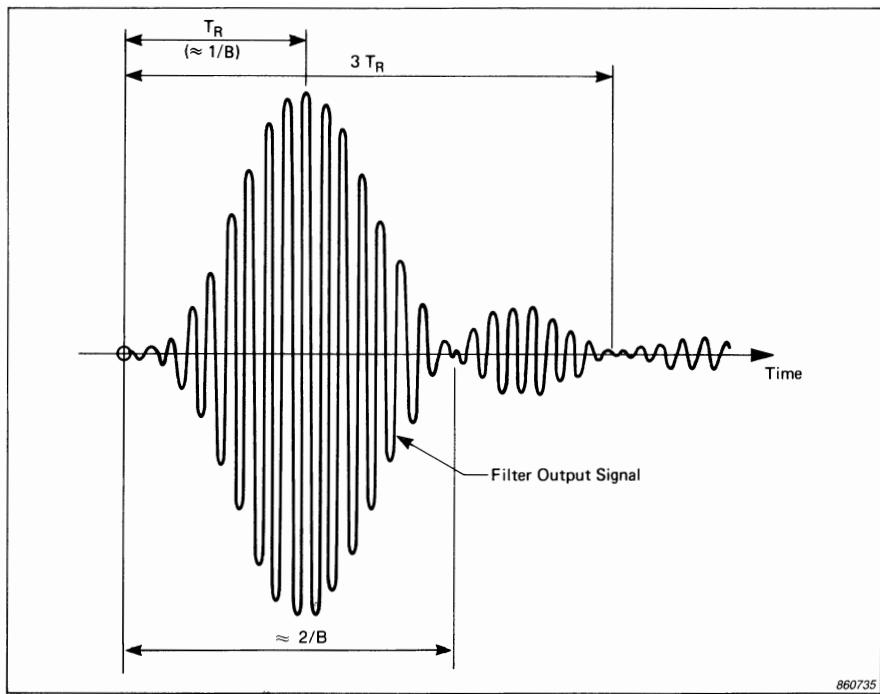
5.2. FILTER ANALYSIS TECHNIQUES

When a transient signal is applied to the input of a filter, the filter output signal represents the convolution of the input signal with the filter impulse response. A typical impulse response for a filter of bandwidth B is shown in Fig.5.6. The "response time" T_R is approximately $1/B$, and the length of the main pulse $2/B$. To include all of the energy in it, it is desirable to integrate over a length of at least $3/B$.

An actual filter output pulse would be longer than its impulse response (by an amount up to the length of the input transient) but the energy contained in it would always be equal to the mean ESD in the filter band multiplied by its bandwidth, and is thus related directly to the required spectral component of the applied signal. Thus, one way of obtaining the complete spectrum is to measure the energy of the filter output pulses for a series of filter positions covering the desired frequency range. An alternative procedure, somewhat akin to FFT analysis, is to record the transient on a loop (tape loop, or circulating digital memory) from which it can be played back repeatedly and analyzed as a periodic signal.

The signal can be analyzed in one pass by a bank of parallel filters, (such as in a digital filter analyzer) or one filter can be moved sequentially to each frequency to be covered. This is the most common procedure using analog filters.

Transient analysis with a real-time digital filter analyzer such as the B & K Type 2131 has the advantage compared with FFT analysis that it is better adapted to constant percentage bandwidth analysis, in particular $1/3$ -octave analysis. It is thus treated first, although sequential analysis using analog filters



860735

Fig. 5.6. Typical filter impulse response

(which has now largely been superseded by FFT analysis) is included in a later section for the sake of completeness.

5.2.1. Digital Filter Analysis of Transients

Assuming an analyzer such as the B & K Type 2131 which has constant percentage bandwidth filters covering a 4-decade frequency range, the filter bandwidth varies widely with frequency. For a given input signal, the filter output signal thus also varies considerably with frequency. Ideally, the energy in the output pulse from each filter would best be measured using running linear integration with an averaging time longer than the longest pulse. The correct result would be the maximum output value of the averager, which would be available the whole time that the entire pulse were contained within the averaging time, and which could be retained by a "Max.-hold" circuit after the averager. Running linear averaging is extremely difficult to achieve in practice, however, even by digital means, and with the Analyzer Type 2131 it is necessary to choose between running exponential averaging and fixed linear averaging. Even so, it is useful to study the concept of analysis using running linear averaging, in order to compare it with exponential averaging.

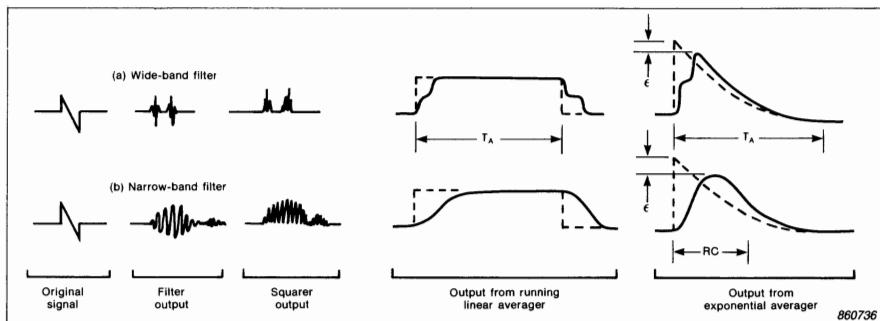


Fig. 5.7. Passage of a transient signal through an analyzer comprising a filter, squarer, and averager (alternatively running linear averaging and exponential averaging). The dotted curves represent the averager impulse responses. T_A = averaging time. RC is the time constant for exponential averaging. ϵ is the error in peak response

Figure 5.7 traces the passage of a typical transient signal (an N-wave) of length T_i through an analysis system consisting of filter, squaring circuit, and alternatively a running linear averager, or exponential averager. The situation is depicted both for a wideband filter (where $T_R \ll T_i$) and for a narrow band filter (where $T_R \gg T_i$). For the wideband filter, the filter output signal follows the input signal more closely, and its length is dominated by T_i ; for the narrow-band filter the output signal approaches the filter impulse response, with a length dominated by T_R . The averaging time T_A is made longer than the total length of the filter output pulse in both cases.

Note that the impulse response of the equivalent exponential averager (with averaging time T_A) has a peak value twice that of the linear averager (as shown in Fig.3.15). On the other hand, because the exponential averager "leaks" energy at a (maximum) rate of $8,7 \text{ dB}/T_A$, the peak output from the averager is somewhat less than the true peak output corresponding to the impulse response. However, the error ϵ will be less than $0,5 \text{ dB}$ if T_A is made at least 10 times greater than the effective length of the filter output pulse ($T_i + T_R$).

In the case of the 2131 (or similar) Analyzers, it is possible to use fixed linear averaging, but this must be initiated just prior to the arrival of the transient. This could in principle be achieved by passing the signal through a delay line at the analyzer input, or if the signal is recorded in an external medium (for example a tape recorder) a trigger signal can be inserted (for example on another channel). This was the procedure used in the following analyses.

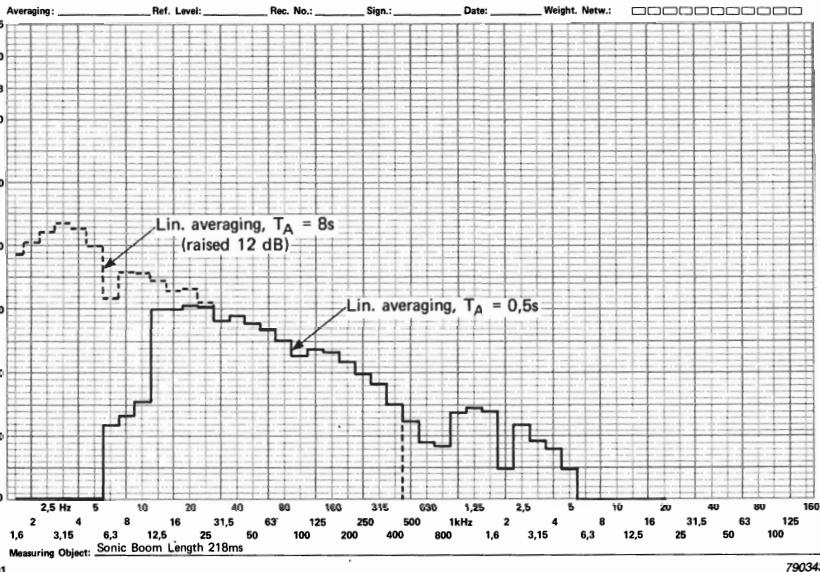


Fig. 5.8. 1/3-octave transient analysis with different averaging times

It will be found that it can be an advantage to have the signal recorded externally, as it may be necessary to analyze it in two passes, in order to extract all information from it. Fig.5.8 (Ref.5.2) shows for example the analysis of a 220ms N-wave (sonic boom) using averaging time $T_A = 0.5\text{s}$. This is only valid down to about 50 Hz ($T_I + 3T_R = 0.48\text{s}$) but on the other hand includes frequencies up to 5 kHz. Fig 5.9(a) shows an analysis of the same signal with $T_A = 8\text{s}$ so as to include all frequencies down to 1,6 Hz ($T_I + 3T_R = 8.3\text{s}$). Because of the 12 dB loss of dynamic range with this longer averaging time, all the frequency components above 500 Hz have been lost. This result (with scaling adjusted by 12 dB) is given as a dotted line in Fig.5.8, and shows that the two results are identical over the mutually valid range. Fig.5.9(c) shows a 1/12-octave analysis obtained by four passes of the same signal, with $T_A = 8\text{s}$. This gives a 6 dB loss of dynamic range, and is only valid down to 6,3 Hz ($T_I + 3T_R = 8.5\text{s}$) but in the frequency range from 6,3 to 250 Hz it gives more detail of the spectrum than the 1/3-octave analysis. Finally, Fig.5.9(b) shows a 1/3-octave analysis made with exponential averaging ($T_A = 8\text{s}$) and "Max. Hold", and this confirms that the result is approx. 3 dB higher than for linear averaging over most of its range, but the difference reduces down to 2,5 dB at 8 Hz, the lowest valid frequency ($T_I + T_R = 0.76\text{s}$).

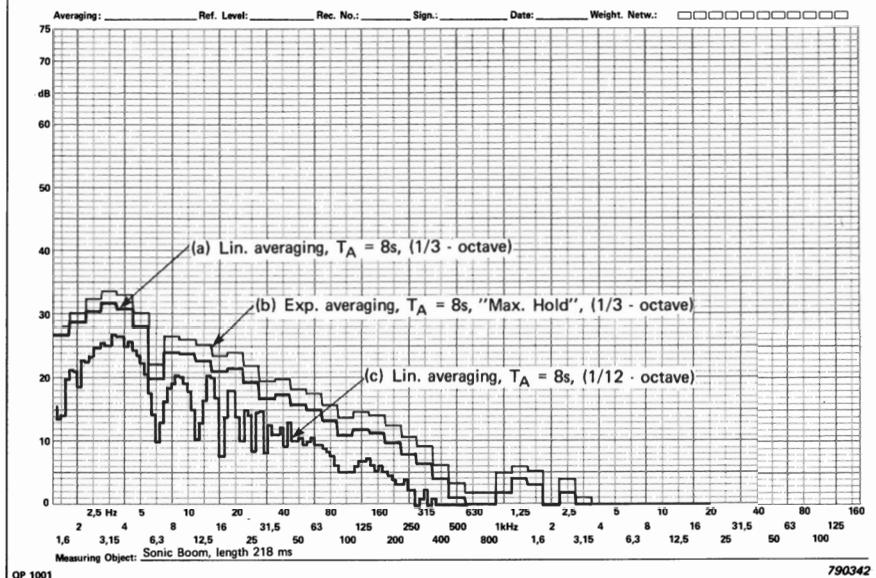


Fig. 5.9. Comparison of analysis methods

Another possibility, which will sometimes be of advantage, is described in the Addendum to Ref.5.2. It makes use of the "constant confidence" mode of the 2131 analyzer, where the averaging time is varied in inverse proportion to the frequency (thus giving an approximately uniform $B T_A$ product). This makes it possible to ensure that the averaging time is always greater than $10 T_R$, but there will be an upper limiting frequency beyond which it becomes less than T_R . Fig.5.10 compares such an analysis with that for $T_A = 8\text{s}$, and illustrates that an additional two valid frequency bands are obtained in this case. The scaling is, however, different for each octave, because of the varying averaging time.

It will rarely be valid to scale constant percentage bandwidth analyses of this kind in terms of ESD (except in the limited frequency range where $B \ll B_{eff}$) but the results can be converted to an "energy" spectrum (per filter bandwidth). The conversion factor to be applied to the measured "power" is equal to T_A for linear averaging, and $T_A/2$ for exponential averaging.

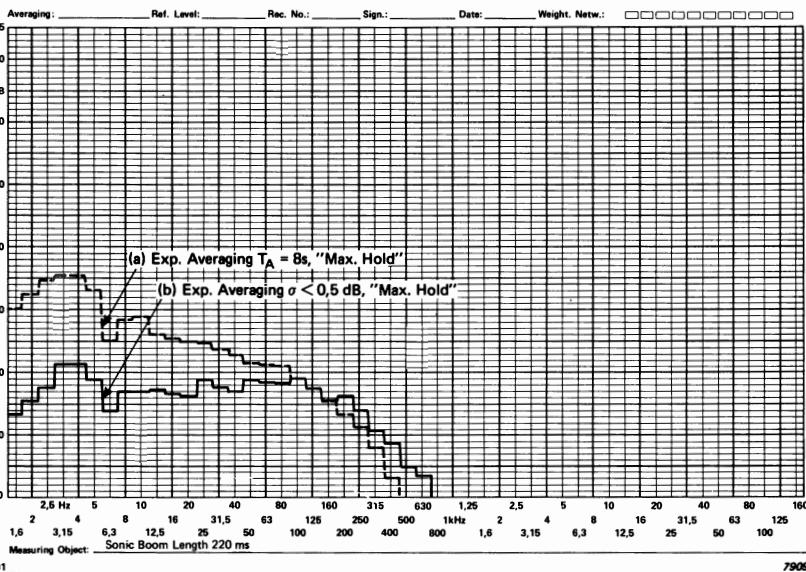


Fig. 5.10. Comparison of results for constant T_A vs. constant BT_A product

5.2.2. Sequential Filter Analysis

There are three main techniques which can be used (here denoted Methods A, B, C) and Method B has two variants depending on the type of detector used. Figure 5.11 illustrates the differences between them, in particular as regards the choice of analysis parameters such as bandwidth, averaging time and sweep speed. The optimum choice in a given situation depends on a number of factors, and should hopefully be clarified in the following discussion:

5.2.2.1. Method A

In this technique, the signal is played back repeatedly and analyzed as a periodic signal. As with the FFT process, the spectrum obtained will then theoretically be a line spectrum, but the individual lines (correctly scaled) represent the ESD spectrum at the frequency of the line. In this case, it is possible (and desirable) to choose a filter bandwidth B equal to or greater than the line spacing $1/T_{rep}$ in order to avoid resolving the individual lines, which only confuse the issue. The line spacing must in any case always be less than the intrinsic bandwidth of the transient (B_{eff}) because its effective length T_{eff} is obviously less than the loop length T_{rep} , which defines the periodic time.

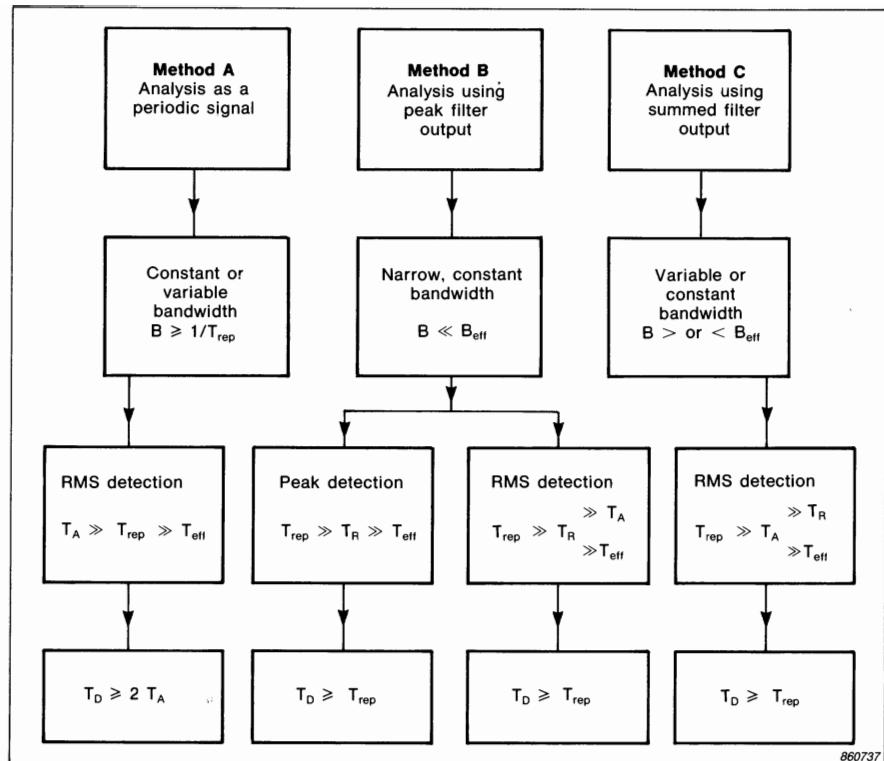


Fig. 5.11. Analog methods of transient analysis. Choice of analysis method and parameters.

B = filter bandwidth. B_{eff} = effective bandwidth of transient.

T_{rep} = loop repetition time. T_A = averaging time

T_R = filter response time ($\approx 1/B$)

T_{eff} = effective length of transient ($\approx 1/B_{eff}$)

T_D = Dwell time per bandwidth. i.e. sweep speed = B/T_D

In general, the sign \gg means "at least 3 times greater than"

Just as for FFT analysis, the loop length T_{rep} should ideally be chosen to be somewhat longer than the transient itself, so that the signal value is zero at the loop junction. Note that with analog tape loops, the following problems have to be taken into account:

1. The tape splice itself may generate a noise pulse; this is particularly the case with FM recording, because the carrier frequency component will almost certainly have a discontinuity at the splice, even if the signal value is zero on either side. It is best to use a splice noise suppression circuit to window it out.

- There may be physical limitations on the minimum loop length, which make the analysis time considerably longer than it need be.

The averaging time T_A should be chosen so as to treat the (periodic) signal as stationary, in other words to obtain a stable output from the detector. If the filter bandwidth is of the same order as the line spacing, T_A can be as little as 3 times the loop repetition time T_{rep} (because the filter impulse response will be of the same order as T_{rep}). Where the filter bandwidth is considerably greater than the line spacing, the signal to the detector will be a series of pulses, which in the limit (Fig.3.17) may require the ratio to be as high as 16. It is suggested that a factor of 3 be tried at first, and then increased as necessary to limit fluctuations. The dwell time, T_D , and sweep speed can then be chosen as for other stationary signals (Section 3.6.2).

Calibration of the results is very similar to the case of FFT analysis. The measured "power" must be multiplied by the loop repetition time T_{rep} to convert it to "energy" (per filter bandwidth) after which it can be divided by the filter bandwidth B to convert it to ESD (provided $B \ll B_{eff}$).

The advantage of this method is that familiar techniques for analysis of stationary signals can be used, and that the results are not very sensitive to the choice of averaging time and filter bandwidth, provided the minimum requirements specified here are adhered to.

5.2.2.2. Method B

This method is probably the most efficient, but is limited to analysis with narrow constant bandwidth filters. It is based on the fact that if the filter bandwidth B is chosen to be appreciably smaller than the effective bandwidth of the transient, B_{eff} , the filter output will always resemble its impulse response, with oscillation frequency corresponding to the tuned centre frequency, and peak amplitude corresponding to the required spectral component (Ref.5.3). Fig.5.12 (which is similar to Fig.5.6) shows a typical filter impulse response, and indicates the three parameters which are of interest in this analysis method. V_{peak} is the maximum short-term peak value of the output pulse, V_{RMS} is the maximum short-term RMS value, and V_{eff} represents the RMS value assuming that the total energy in the output pulse is distributed over the length of the main pulse ($2/B$). Because the oscillations inside the envelope are approximately sinusoidal, $V_{RMS} \approx V_{peak}/\sqrt{2}$, and assuming that most of the energy is in the main pulse, and that its envelope is roughly sinusoidal:

$$V_{eff} \approx V_{RMS}/\sqrt{2} \approx V_{peak}/2 \quad (5.1)$$

The requirement that $B \ll B_{\text{eff}}$ will be satisfied if:

$$B \leq \frac{1}{5 T_{\text{eff}}} \quad (5.2)$$

where T_{eff} is the effective length of the transient. With experience this can be estimated by eye with sufficient accuracy, but a conservative estimate will always be obtained by setting it equal to the total length (the limiting case corresponding to a rectangular pulse or tone burst).

The loop repetition time T_{rep} should then be selected from the expression:

$$T_{\text{rep}} \geq 3/B \quad (5.3)$$

to ensure that successive filter output pulses are sufficiently well separated.

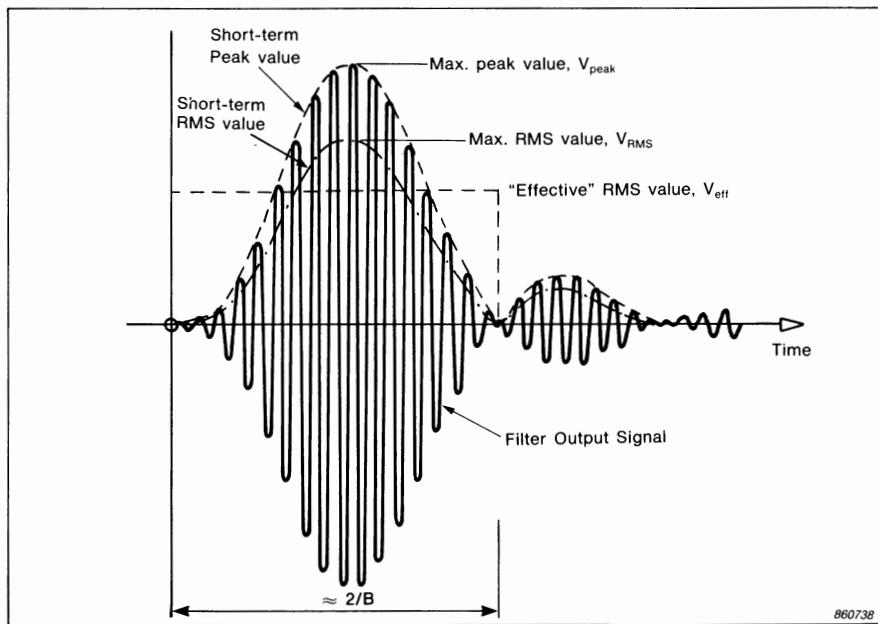


Fig. 5.12. Filter impulse response with definition of the terms V_{peak} , V_{rms} and V_{eff} .

A peak detector, reset for every playback of the loop (for example triggered by the loop junction) can be used to measure V_{peak} . Alternatively, an RMS detector can be used to measure V_{RMS} , provided that the averaging time fulfills the following requirements:

- The averaging time must be large enough to reduce ripple due to the filter ringing frequency $f_o^{\$}$, and from Fig.3.16 this will be achieved if:

$$T_A \geq 3/f_o \quad (5.4)$$

- The averaging time must be appreciably shorter than the filter response time T_R , so that the short-term RMS value follows the short-term peak value (though smaller by a factor of $\sqrt{2}$, or 3 dB). This will be the case within about 1 dB for:

$$T_A \leq \frac{1}{3B} \quad (5.5)$$

From Expressions (5.4) and (5.5) it will be seen that RMS detection can only be used for frequencies greater than 9 or 10 times the bandwidth. This limitation does not apply to the Analyzer Type 2010, or others with a fixed filter output frequency (see footnote).

For very small values of the repetition time T_{rep} it should be checked that the recorder writing speed does not limit the registration of the spectrum. The pen will be able to rise and fall over at least a 20 dB range if the following expression is satisfied.

$$T_{rep} \geq \frac{100}{W} \quad (5.6)$$

where W is the writing speed corresponding to 100 mm paper. As an example, for $W = 1000 \text{ mm/s}$, the minimum value of T_{rep} is 0,1 s.

The dwell time T_D (the time required to sweep a frequency range corresponding to one bandwidth) should not be less than the repetition time T_{rep} in order that samples of the spectrum are not separated by more than one filter bandwidth. The filter bandwidth is in any case constrained by Eqn. (5.2) to be considerably less than the bandwidth of the transient being analyzed. The sweep speed can then be calculated by definition as B/T_D .

In order to scale the results, use can be made of the following relationship between the frequency characteristic and impulse response of the filter (Parseval's theorem):

total energy in frequency = total energy in time

Thus,
$$\text{ESD} \times B = V_{eff}^2 \cdot \frac{2}{B}$$

or,
$$\text{ESD} = V_{eff}^2 \cdot \frac{2}{B^2} \quad (5.7)$$

[§] Note that with the B & K Analyzer Type 2010, the filter output frequency is constant (750 Hz or 30 kHz) independent of the filter centre frequency. The shortest averaging time (0,1 s) will therefore always damp out ripple.

Making use of Expression (5.1) this can be related to the measured parameters V_{peak} and V_{RMS} as:

$$ESD \approx \frac{V_{peak}^2}{2B^2} = \frac{V_{RMS}^2}{B^2} \quad (5.8)$$

This expression applies exactly for an ideal filter (Ref.5.3) but may require modification for a practical filter. The ratio between V_{peak} , V_{RMS} and V_{eff} can be established by analyzing a known signal such as a rectangular pulse or tone burst. The relationship between V_{peak} and V_{eff} should be constant for a given filter characteristic, while the relationship between the measured value of V_{RMS} and V_{peak} will be influenced to a small extent by the ratio of T_A to T_R (i.e. the BT_A product) but should be constant for a given ratio (See Expression (5.5)).

5.2.2.3. Method C

Like Method B, this method obtains a separate estimate of the spectrum for each playback of the transient, but there is no restriction on bandwidth; the total energy in the filter output pulse is measured independent of the pulse shape just as for digital filter analysis (Section 5.2.1). Thus, this method can be used for constant percentage bandwidth analysis, e.g. 1/3-octave analysis.

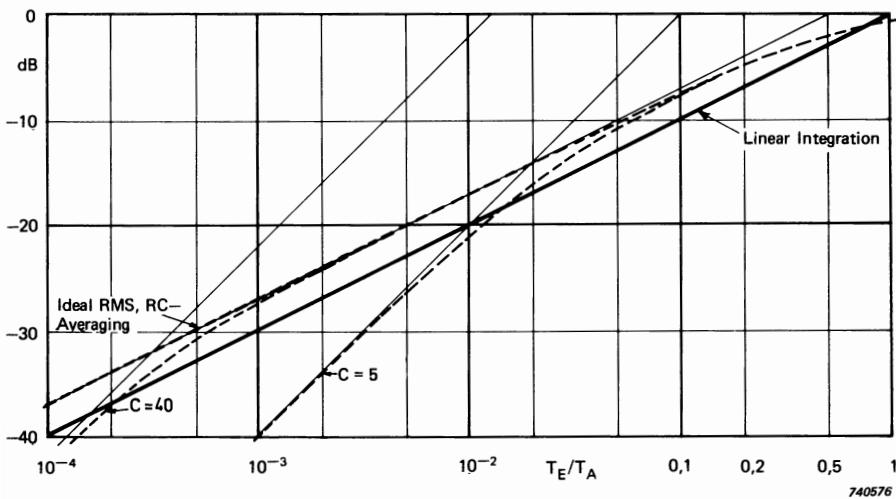


Fig. 5.13. Valid range for the factor impulse length: averaging time (T_E/T_A) for RMS detectors with different crest factors C

The analysis is very similar in principle to that illustrated in Fig.5.7, using exponential averaging, with the additional complication that it may be necessary to take account of the crest factor capability of the detector (if a Wahrman detector is used). Figure 5.13 (from Ref. 5.4) illustrates the departure of measured results from the theoretical values against the ratio T_E/T_A , for a range of values of crest factor C . In the derivation of the diagram, T_E was the length of square pulses applied to the detector, but in practice the effective length of the filter output pulse ($T_i + T_R$) can be used. The dotted line in the upper part of Fig.5.13 ($0,1 < T_E/T_A < 1$) shows the effect of error ϵ in Fig.5.7, and confirms that the error is less than 0,5 dB for $T_E/T_A \leq 0,1$. A larger value of this ratio could be used, however, provided compensation were made in accordance with the diagram.

It is perhaps worth noting that the (now obsolete) real-time $1/3$ -octave Analyzer Type 3347 had detectors with a crest factor capability of 5. From Fig.5.13 it can be seen that it can still be used for impulse analysis over a range of the ratio T_E/T_A between about 0,03 and 0,3 with a scaling adjustment of about 1 dB.

For swept frequency analysis, a suitable value of T_A can be selected using Fig.5.13, after which the loop length T_{rep} can be selected according to the following expression:

$$T_{rep} \geq 3 T_A \quad (5.11)$$

This will ensure that the detector can fall at least 20 dB between successive playbacks, keeping in mind that the maximum rate of fall is $8,7 \text{ dB}/T_A$. Note that this requirement is automatically satisfied in Method B.

5.2.2.4. Examples

To assist in assessing the influence of the various factors which play a role and to compare the pros and cons of the various analysis methods, a transient signal was analyzed in all of the four ways illustrated in Fig.5.11.

The signal used was a tone burst of length approx. 10 ms, recorded in a 7502 Digital Event Recorder with 10 K memory. The circulation time of the memory (for input sample rate 100 kS/s) was approximately 100 ms. The frequency of the tone was approximately 1 kHz, but in fact adjusted to 976,6 Hz to compensate for the actual pulse length 10,24 ms (and repetition time 102,4 ms) thus ensuring that there were exactly 10 periods in the tone burst. The RMS level of the sinusoidal part was 3,16 V (see Fig.5.14). The analysis setup is shown in Fig.5.15, but it should be noted that the Measuring Amplifier Type 2607 was not always used; in some cases the DC output of the 2010 was taken directly to the Level Recorder Type 2307.

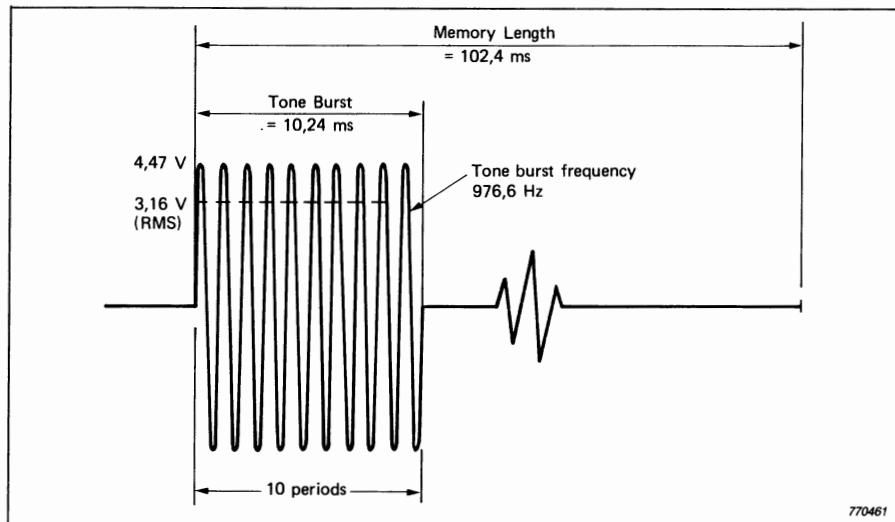


Fig. 5.14. Test signal

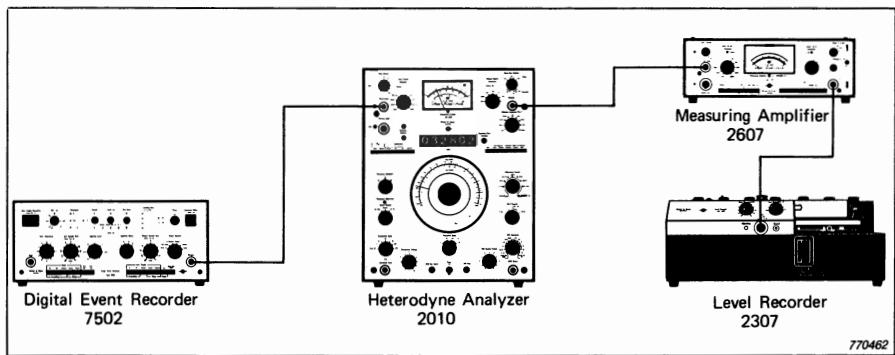


Fig. 5.15. Instrument set-up for example analyses

The four different methods will now be discussed in some detail.

1) Method A (cf. Section 5.2.2.1)

The tone burst occupied 10% of the memory of the Digital Event Recorder, so that there was a good chance that there would be no problems with crest factor. The total RMS level of the periodic signal was 10 dB below that of the sinusoidal section (i.e. 1 V) and peak value was $3.16 \sqrt{2} \approx 4.5$ V, and therefore the crest factor of 4.5 for the unfiltered signal was within the limit of 5 for the 2010 and

2607. The crest factor of the filtered signal would be even less, since as mentioned in the next section, the bandwidth was so chosen that the filter response time was equal to the repetition time and thus the filter output signal almost stationary.

The choice of the various analysis parameters was made as follows:

1. $T_{rep} = 100\text{ms}$. This corresponded to the memory length of the 7502, and gave a spectral line spacing ($1/T_{rep}$) of 10 Hz, which was suitably narrow with respect to B_{eff} ($= 1/T_E = 1/0,010\text{s} = 100\text{Hz}$) for the tone burst.
2. $B = 10\text{Hz}$. This just satisfied the requirement that $B \geq 1/T_{rep}$ ($= 1/100\text{ms} = 10\text{Hz}$). As already mentioned, it also satisfied the other requirement, viz. $B < B_{eff}$, so that expression of the results as Energy Spectral Density was valid.
3. $T_A = 0,3\text{s}$. This represents a ratio to T_{rep} of 3, but was found to be adequate to damp out fluctuations because the filter response time was of the order of the repetition time.
4. $T_D = 0,6\text{s}$. This was chosen as twice the averaging time ($K_A = 2$ in Table 3.5) as this is suitable for the 2010).
5. $B_{eq} = 0,75\text{mm}$. This resulted in a suitable total length for the total analysis ($\approx 200\text{mm}$) and could be obtained with the 2010 and 2307. The resulting paper speed (Eqn.(3.21)) was thus given by

$$P \leq \frac{B_{eq}}{T_D} = \frac{0,75\text{mm}}{0,6\text{s}} \therefore \text{Select } 1\text{mm/s}$$

The resulting analysis is given in Fig.5.16.

6. Scaling — The voltage corresponding to Full Scale Deflection (FSD) on the recording paper was 0,32 V, corresponding to a power of 0,1 V². The analyzer bandwidth was 10 Hz, and therefore the maximum PSD of the result was:

$$\frac{0,1\text{V}^2}{10\text{Hz}} = 0,01\text{V}^2/\text{Hz}$$

The repetition time T_{rep} was 0,1 s and thus the final result as an Energy Spectral Density (ESD) was:

$$0,01\text{V}^2/\text{Hz} \cdot 0,1\text{s} = 0,001\text{V}^2\text{s}/\text{Hz}$$

This agrees with the theoretical calculation as follows:

$$\text{Energy in pulse} = 3,16^2 \text{V}^2 \cdot 0,01 \text{s} = 0,1 \text{ V}^2\text{s}$$

$$\text{Effective bandwidth } B_{\text{eff}} = \frac{1}{T_E} = \frac{1}{0,01 \text{ s}} = 100 \text{ Hz}$$

$$\therefore \text{ESD} = \frac{0,1 \text{ V}^2\text{s}}{100 \text{ Hz}} = 0,001 \text{ V}^2\text{s/Hz}$$

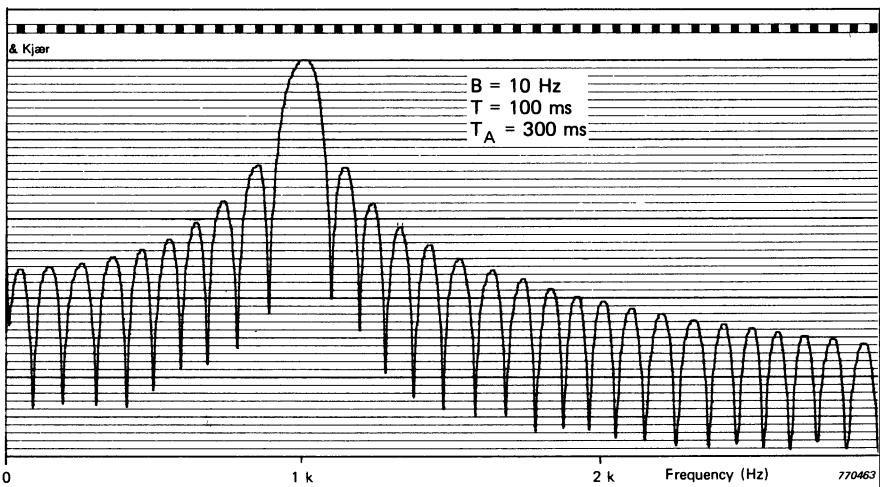


Fig. 5.16. Analysis as a periodic signal (Method A)

(2) Analysis using Method B with Peak Detection (cf. Section 5.2.2.2.)

The various analysis parameters were chosen as follows:

1. $B = 3,16 \text{ Hz}$. This more than satisfied Eqn.(5.2) which requires $B < 20 \text{ Hz}$, but it was desired to compare the results with those for short-term RMS averaging (next section) and this required that the filter response time T_R (320ms) was somewhat longer than the averaging time (100ms minimum). Thus, this analysis was considerably less efficient than it could have been. On the other hand, it was possible to allow the detector to fall with a time constant corresponding to $T_A = 0,1 \text{ s}$ rather than triggering the "reset" function, which would have required a more complicated setup.
2. $T_{\text{rep}} = 1 \text{ s}$. This satisfied the requirement of Eqn.(5.3) of $T_{\text{rep}} \geq 3/B = 3/3,16 = 0,95 \text{ s}$ and, incidentally, at the sweep speed used, gave successive peaks

separated by less than the pen thickness, so that the envelope of the peaks was very well demarcated.

- With the narrower bandwidth, $B_{eq} = 0,24\text{ mm}$ and thus in order to satisfy the requirement $T_D \geq T_{rep}$ the paper speed would have to be less than $0,24\text{ mm/s}$. However, in this case it was decided to make the paper speed $0,3\text{ mm/s}$, resulting in $T_D = 0,8\text{ s}$

The justification for this was that the analyzer bandwidth was so much smaller than that of the function being measured (3 vs 100 Hz) and the steps between successive peaks ($0,3\text{ mm}$) were less than the line thickness. The resulting analysis is depicted in Fig.5.17.

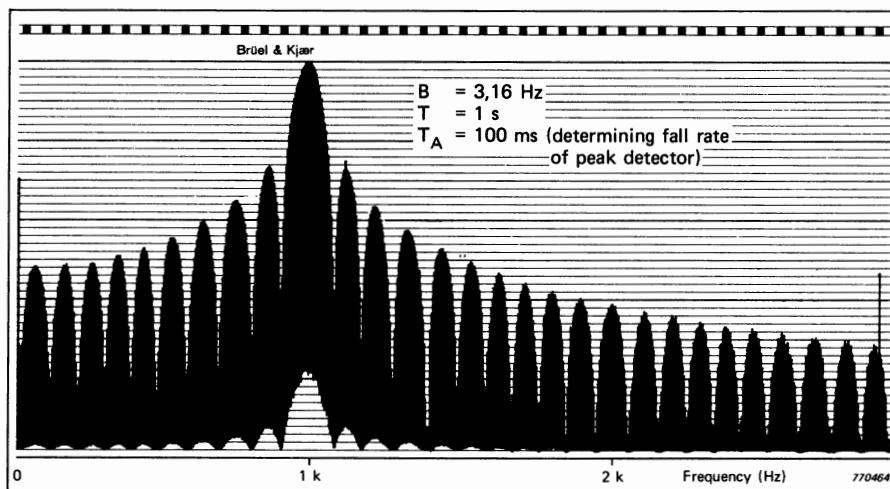


Fig. 5.17. Analysis by Method B using peak detection

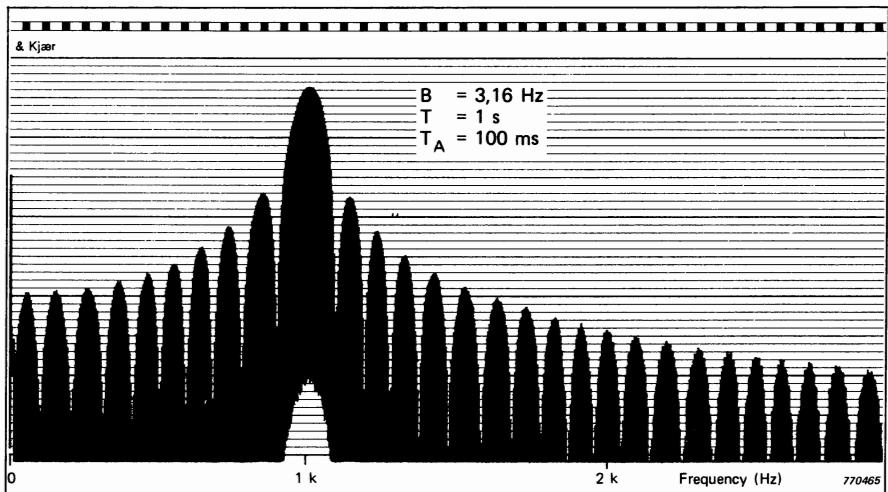
- Scaling — The measured peak voltage corresponding to FSD was $0,14\text{ V}$. Using Eqn.(5.8) the corresponding ESD is

$$\frac{V_{peak}^2}{2B^2} = \frac{0,14^2\text{V}^2}{2 \cdot 3,16^2\text{Hz}^2} = \frac{0,02\text{ V}^2\text{s}}{20\text{ Hz}} = 0,001\text{ V}^2\text{s/Hz}$$

as before.

- (3) Analysis using **Method B** with Short-term RMS Detection (cf. Section 5.2.2.2)

- Analysis parameters were exactly the same as for peak detection as described in the previous section with the exception that the peak detection was replaced by RMS detection with $T_A = 0,1$ s. The resulting analysis (with unchanged amplification) is depicted in Fig.5.18.



*Fig. 5.18. Analysis using short-term RMS detection
(same amplification as Fig.5.17)*

- Scaling — As discussed in connection with Eqn. (5.5), because the filter response time T_R (≈ 300 ms) is only 3 times the averaging time T_A (100 ms) the recorded result is 3,5 – 4 dB below that for peak detection compared with the expected 3 dB. If it had not been possible to calibrate the system, the result obtained by adjusting by 3 dB would have been very close to correct.

(4) Method C (cf. Section 5.2.2.3)

Parameters were selected as follows:

- $B = 10$ Hz. In contrast to the analysis just discussed, it was no longer necessary to have the filter response time longer than the averaging time, and in fact an advantage to have it as short as possible. 10 Hz was the largest standard bandwidth satisfying the condition that it was less than 1/5 the bandwidth of the function itself. It also allows a more direct comparison with the results of Fig.5.16.

- $T_A = 0,3\text{s}$. This is at the lower limit of the allowable range indicated in Fig.5.13 for a crest factor of 5, but is acceptable in this case where it is known that the effective length of the filter output pulse is the same for all frequencies (determined in all cases by the filter response time because of the constant bandwidth.) In a physically generated transient this is not certain (the high frequencies have a tendency to die out more rapidly) and moreover for constant percentage bandwidth the filter response time varies with frequency, so that it is advisable to choose the ratio T_E/T_A closer to 0,1.
- $T_{rep} = 1\text{s}$. Because of the relatively narrow bandwidth it was sufficient to satisfy Eqn.(5.11), i.e. $T_{rep} \geq 3T_A$.
- Taking $T_D = 1\text{s}$ to make it at least equal to T_{rep} , the maximum sweep speed is given by $B_{eq}/T_D = 0,75\text{ mm/s}$ and thus $0,3\text{ mm/s}$ was again selected.

The resulting analysis is shown in Fig.5.19. The amplification was exactly the same as for the analysis by Method A (Fig.5.16).

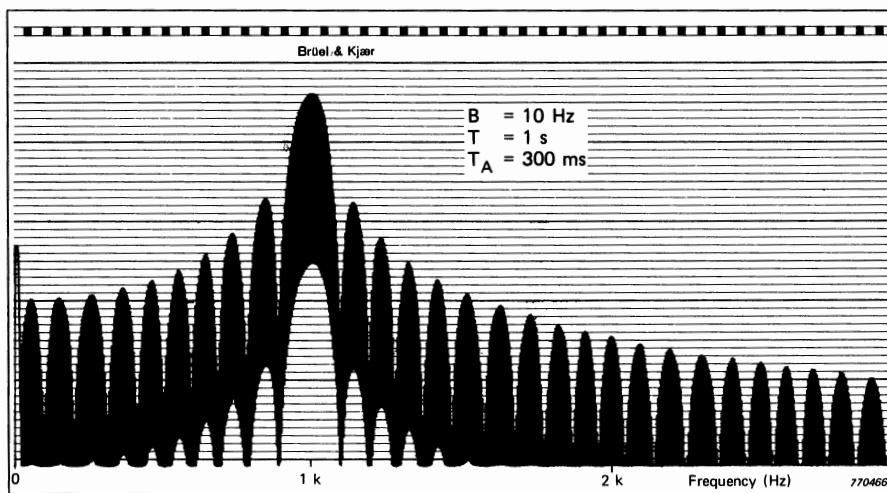


Fig. 5.19. Analysis using Method C
 (Same amplification as Fig.5.16)

- Scaling — In comparison with the analysis of Fig.5.16, for which the bandwidth was the same, the difference in scaling would arise from the fact that the measured PSD units should here be multiplied by $T_A/2$ (i.e. 150ms) whereas in that case the multiplying factor was the repetition time T_{rep} (100ms). From this, it would be expected that the results should be $10 \log_{10} (1,5) = 2 \text{ dB}$ below those of Fig.5.16, instead of the actual 4 dB. The major

part of the difference can be explained by reference to Fig.5.13, where for $T_E/T_A = 0,3$ (as applies here) the actual detector characteristic is approx. 1,5 dB below the theoretical line for "Ideal RMS, RC-averaging". It should be remembered that in a practical case the ratio T_E/T_A should have been chosen smaller, but it was difficult to find an example on which all four methods could be used, keeping in mind that they are best adapted to different situations.

6. NON-STATIONARY SIGNALS

The type of non-stationary signals treated in this chapter, in contrast to the short transients of Chapter 5, are those where one is interested in the change of spectral information with time. One example would be speech, another the vibration signals from a machine during run-up or run-down in speed. The only technique considered here is the use of window functions to window out short sections of the overall signal which are near stationary or which contain isolated events. By moving such a window along a longer record, in overlapping steps, the variations of spectral information with time can be determined.

Cyclic signals, for example the vibration and sound from reciprocating machines such as diesel engines, represent a special case, where in the long term they may be considered stationary, but where short-term variations within the cycle are of interest. With such signals it is possible to average the results over a number of cycles, which often would be necessary to obtain a stable, repeatable result. This type of analysis is considered in Section 6.1.2.

With machine run-ups and run-downs, it is sometimes desirable to express the results as an "Order Analysis", where the frequency axis is changed to one of harmonic order, independent of shaft speed. The various ways of achieving this are discussed in Section 6.2.

Time windowing techniques are generally suitable for the types of applications mentioned here. In some specialised cases, the variations may be too rapid to be resolved in this way, and an alternative technique based on the "Wigner Distribution" (Ref.6.1, 6.2) may be of advantage. The reader is referred to the cited literature for further information on the use of the Wigner distribution

6.1. SCAN ANALYSIS

Scan analysis is the term given to the technique whereby a window function such as Hanning is scanned along a record in overlapping steps, and a short-term spectrum obtained for each position. An example was given in Fig.2.25,

showing how a short section of a speech signal (the vowel "i" in the word "this") could be isolated by a window in this fashion, and analyzed separately. The results are somewhat dependent on the choice of the analysis parameters such as frequency range, window type, window length and step length, and so these matters will be discussed first.

6.1.1. Choice of analysis parameters

As regards choice of **window type**, any smoothly rounded window function, such as Hanning, could be used, although Hanning is probably the best choice for the following reasons:

1. It is simply generated, and available in most analyzers.
2. It has a relatively narrow bandwidth, compared with most other specialised windows (Table 4.1).
3. The selectivity of the filter characteristic is less likely to be important (as compared with stationary signals) as it is likely to be limited by the signal itself (i.e., minor non-stationarity within the window).
4. With overlapping Hanning windows, the individual spectra can be averaged together to obtain the spectrum of a larger portion of the signal with uniform weighting (Fig.4.21).

Consequently, most of the remaining discussion will assume the use of Hanning windows.

The **length** of window should be chosen with respect to the following criteria:

The window should be sufficiently short that the signal within it does not vary greatly. For continuous signals such as machine run-downs or vowels in speech, this means that the windowed portion should be quasi-stationary. For impulsive signals such as reciprocating machine vibrations, or plosive consonants in speech, the window length should be short enough to isolate the various impulses from each other, or from the continuous sections. On the other hand the window should not be so short as to restrict the resolution of the analysis (beyond that determined by variations in the signal itself) and should be at least twice as long as individual impulses. The effective length of the window (for Hanning weighting) can be taken as half the total length T .

To take the example of speech, where vowels typically have a duration of 100–200 ms, and plosive consonants < 10 ms, a suitable choice of effective window length is 30–50 ms. Appropriate choice of window length can often best be made by visual inspection of the signal.

Note that where the internal Hanning window of the FFT analyzer is used, its length is selected indirectly by choice of the frequency range. On occasion it will be found that the latter has to be selected higher than otherwise required in order to obtain a sufficiently short window; in such cases the portion of the spectrum outside the range of interest can simply be dropped from the display. In the case of the analyzers Types 2032/2034 the length of window can additionally be varied:

- (a) by the use of "zero padding", which sets it to half the length for the same frequency range.
- (b) arbitrarily, by using the "Special Parameters" to tailor an appropriate window (shorter than the record length T).

With regard to choice of **step length**, the following factors should be taken into account:

- (1) If it is made equal to the effective length of the window (i.e. 50% overlap) the successive analyses will be effectively uncorrelated, and would constitute the minimum number of spectra to represent the entire signal with no loss of information.
- (2) For 3-dimensional (i.e. "waterfall") displays of arbitrarily varying signals (e.g. speech) it will normally be of advantage to reduce the step length to less than half the maximum suggested in (1) (i.e. overlap $> 75\%$). Thus, successive analyses will be partially correlated, which aids visual interpretation of 3-dimensional diagrams (see later for examples).
- (3) Where the signal itself is fairly predictable, e.g. machine run-downs, it is not necessary to satisfy the requirements of (2), and in fact the step length may even be greater than the window length. The individual spectra in a waterfall diagram, for example, are often separated by uniform increments in shaft speed, which may represent time intervals considerably longer than the window length. In fact, for slowly varying machine signals, it may be of advantage to use averaging (normally exponential averaging over, say, 8 spectra) in order to smooth random components in the spectrum. This must be balanced against the smearing of discrete harmonic components due to the effectively longer window obtained.

6.1.2. Cyclic signals

Figure 6.1 shows a typical cyclic signal, the vibration signal from a diesel engine, and indicates how it can be analyzed to determine the way in which the spectrum changes throughout the cycle. The original signal (Fig.6.1(a)) is seen to be dominated by a series of impulses (the dominant ones corresponding to

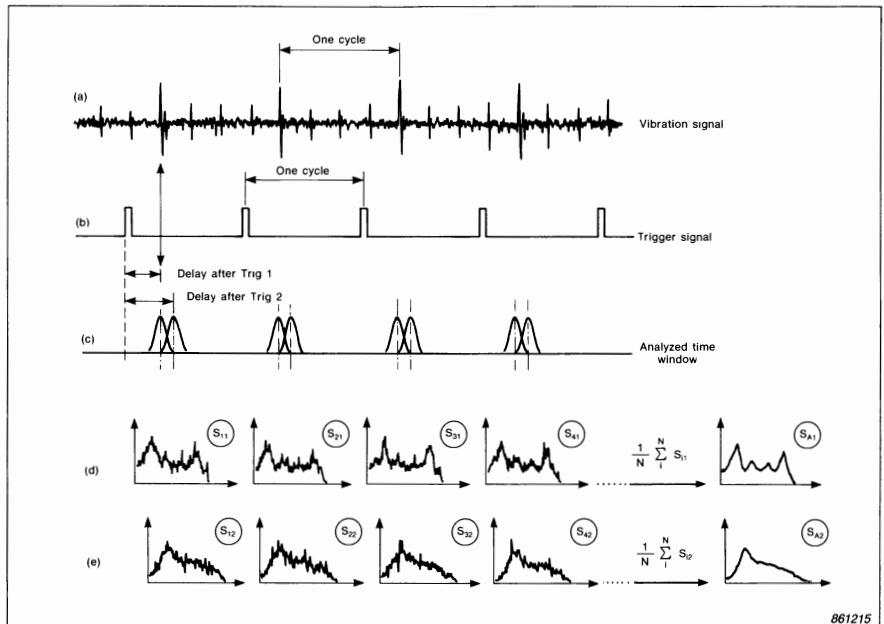


Fig. 6.1. Principle of analysis technique for cyclic signals

the firing of each cylinder) but these can be separated from each other by suitable choice of a Hanning window as indicated. A one-per-cycle trigger pulse (Fig.6.1(b)) is used to position the Hanning window in a particular part of the cycle, using a variable delay after trigger to vary its position (Fig.6.1.(c)). Even in one position, it is found that the spectra from individual cycles vary somewhat, but if a number are averaged together (e.g. over 32 cycles) the results become stable, and typical of that part of the cycle (Fig.6.1(d) & (e)).

Figure 6.2 shows the results of such an analysis (averaged over 64 cycles), displayed in a 3-dimensional “waterfall” diagram. Variations in time can clearly be seen (for example, the firing of the individual cylinders), but in addition the breakdown in frequency gives the possibility of separating events (e.g. combustion and piston slap) which occur at roughly the same time, but which have different frequency contents.

The analysis parameters for Fig.6.2 were selected as follows: The engine was a 4-cylinder, 4-stroke diesel engine running at 1500 rpm (750 cpm or 80 ms cycle time). Thus, the individual firing strokes were separated by about 20 ms, and by viewing the time signal on the analyzer screen it was determined that the individual pulses had a length of ≈ 2 ms. The transform size was therefore chosen to be 8 ms (effective length 4 ms, representing 5% of the total cycle).

This in fact involved playing back the signal (which was recorded on an FM tape recorder) 10 times slower than recorded, and analyzing it in the 5 kHz frequency range (corresponding to 50 kHz in the original signal). Because the signal only contained information up to 20 kHz, only the first 160 lines of the 400 obtained (from the 2031 Analyzer) were used for displaying the results. The step length was set to 0,2 T giving a total of 50 spectra for the whole cycle.

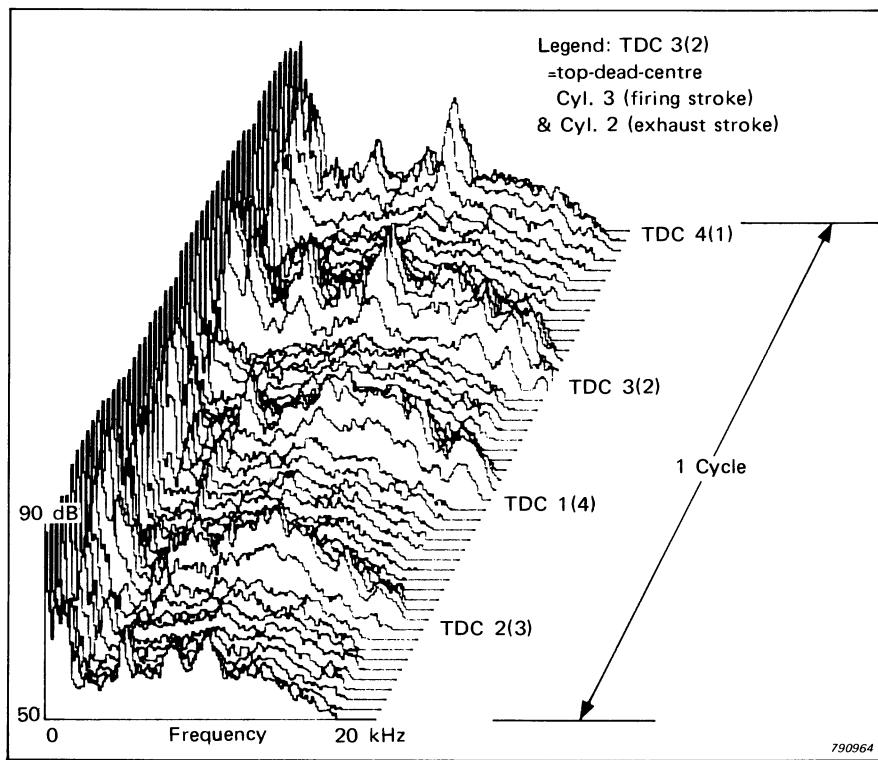


Fig. 6.2. Frequency-time representation of a diesel engine cycle

6.1.3. Spectral Variations with Machine Speed

It is often desired to investigate the way in which machine vibrations (and sound) vary with shaft speed, and one of the best ways of doing this is to present the results in a 3-dimensional spectral map, such as illustrated in Fig.6.3. As indicated, the third axis (which may be inclined as shown, or vertical) can either be time, or shaft speed in RPM; it depends on how the individual spectra are captured. Such a 3-dimensional spectral map is often referred to as a "waterfall" or "cascade" plot, although this term is sometimes reserved for a

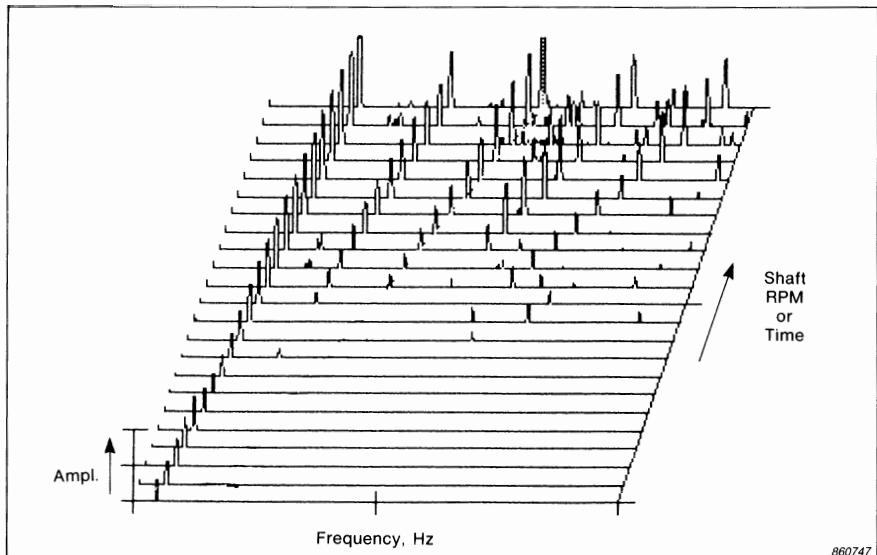


Fig. 6.3. 3-Dimensional spectral map or "Waterfall" plot

live display, continuously updated as new spectra are generated. When the third axis represents shaft RPM, the spectral map is sometimes referred to as a "Campbell diagram", although strictly speaking a Campbell diagram has frequency axis vertical, RPM axis horizontal, and spectral amplitude indicated by the diameter of a circle (or square) at each point in the diagram (Fig.6.4).

Figure 6.5 is a spectral map (with shaft speed as the third axis) which illustrates the advantage of this type of display. It was obtained during the run-up of a small electric motor, with spectra taken at shaft speed intervals of 2,5 Hz between 20 Hz (1200 RPM) and 75 Hz (4500 RPM). This illustrates how the various harmonics fall along radial lines, and can thus be separated from constant frequency components (such as resonances and mains frequency related components) which follow lines parallel with the RPM axis (in this case vertical). This type of display is ideal for determining whether a noise or vibration problem within a particular speed range is primarily due to a large forcing function (and if so, which one) or to excessive amplification by a structural resonance.

The individual spectra for such a plot can basically be captured in three different ways, depending on the rate of acceleration (or deceleration) of shaft speed.

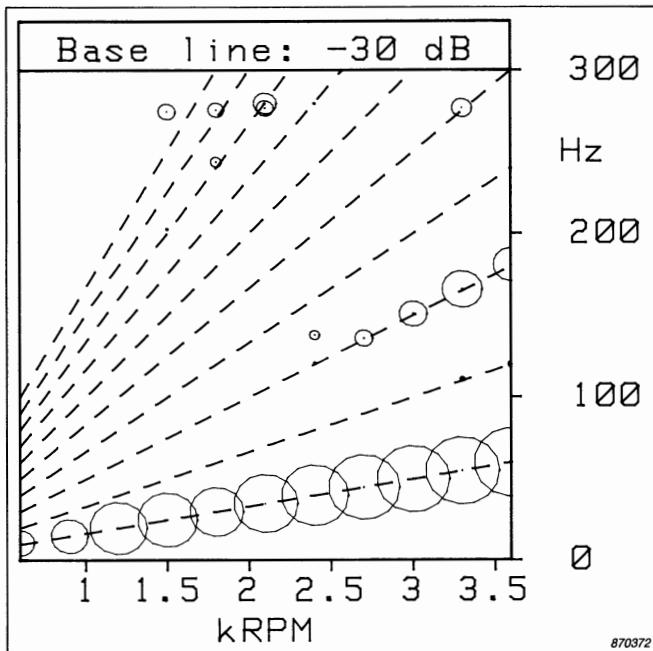


Fig. 6.4. Example of a Campbell Diagram

- (1) Ideally, the shaft speed should be constant at the respective RPM, as the spectrum is taken. Thus, all harmonics will appear as discrete frequency components whose amplitude in the diagram directly indicates the strength of the component. This is sometimes possible during the very slow start-up of such machines as large steam turbines. Linear averaging can even be used to smooth out any random components.
- (2) For intermediate rates of acceleration, the spectra can be captured (by transfer to a digital memory) at predetermined intervals of shaft speed, either automatically, or manually. The shaft speed can either be determined from the signal itself (by selection of the first harmonic), by simultaneous analysis of a tachometer signal (e.g. a pulse once per shaft revolution) or from a separate shaft speed indicator. Note that even where the frequency range is within the real-time capability of the analyzer, so that the analysis speed does not limit the generation of spectra, it may be the time required to transfer the spectra to a digital memory which limits the rate of acceleration of shaft speed.
- (3) For rapid rates of acceleration, even beyond the real-time capability of the analyzer, it is possible to capture the time signal in a long memory, and

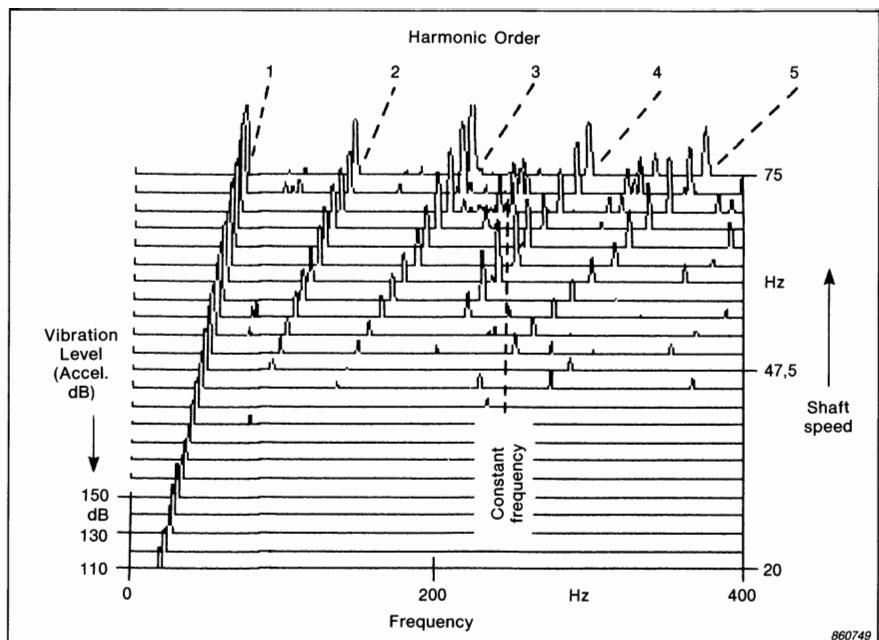


Fig. 6.5. Spectral map with shaft speed as third axis, showing separation of harmonic from constant frequency components

analyze it by scan analysis after the event. As an example, in the 200 Hz frequency range, the 10 K memory of the 2033 Analyzer represents 20 seconds of signal, which would often contain the most important part of a run-down. Where the signal has been recorded on tape, different 20 s sections could in fact be analyzed successively. Figure 6.6 was obtained by analysis of the signal from the run down of a small electric motor. In order to arrange the diagram with the lowest speeds (and signal levels) first, the individual spectra were selected manually, by scanning backwards from the end of the record. Note that because of the rate of acceleration, the higher harmonics (in particular) no longer appear as discrete frequencies, and are smeared over a number of lines which increases with the harmonic order. Thus, the height of the peak does not directly represent the strength of the component; it would be necessary to integrate over the whole of the distributed peak to achieve this (note that with the analyzers Types 2032/2034, this can be done for individual peaks using the "ΔTOTAL" cursor indication). This reduces the utility of this display method to some extent, in cases such as this where machine speed varies along the window length.

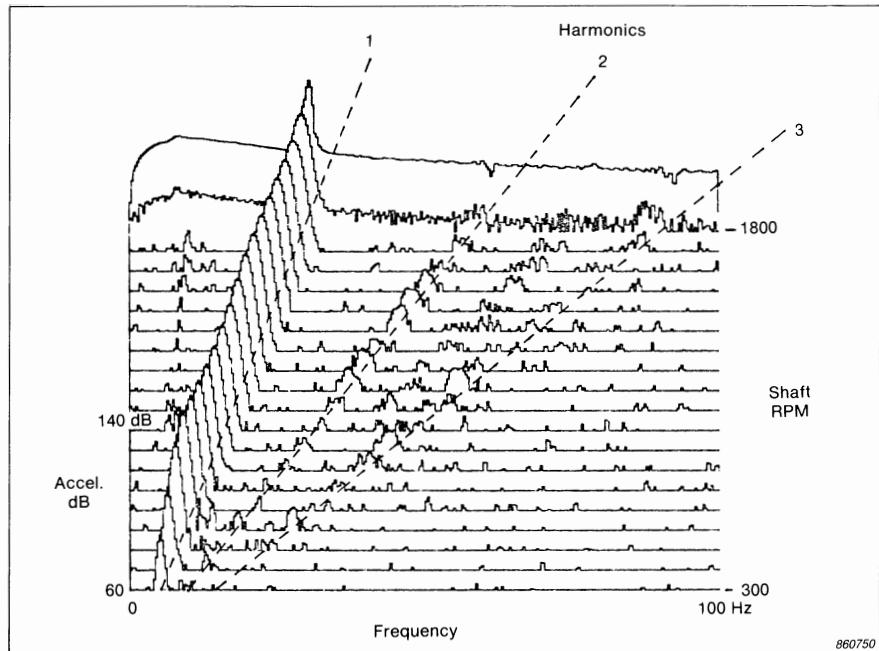


Fig. 6.6. Spectral map obtained by scan analysis in a 10 K memory, representing the run-down of a small electric motor

6.2. ORDER TRACKING ANALYSIS

Where one is primarily interested in the behaviour of harmonic orders of shaft speed (even high orders such as tooth meshing frequencies) then order tracking analysis will often be of advantage. For one thing it eliminates the "smearing" problem, referred to in the previous section, by forcing harmonic components to be located in one analysis line. This is achieved by controlling the sampling frequency of the analog-to-digital (A/D) converter in synchronism with the shaft speed; Fig.6.7 illustrates the basic principles.

Fig.6.7(b) shows a hypothetical signal produced by a rotating shaft during a run-up (in practice, the amplitude would normally also vary with shaft speed). Fig.6.7(a) shows the samples obtained by sampling it with 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. Fig.6.7(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.

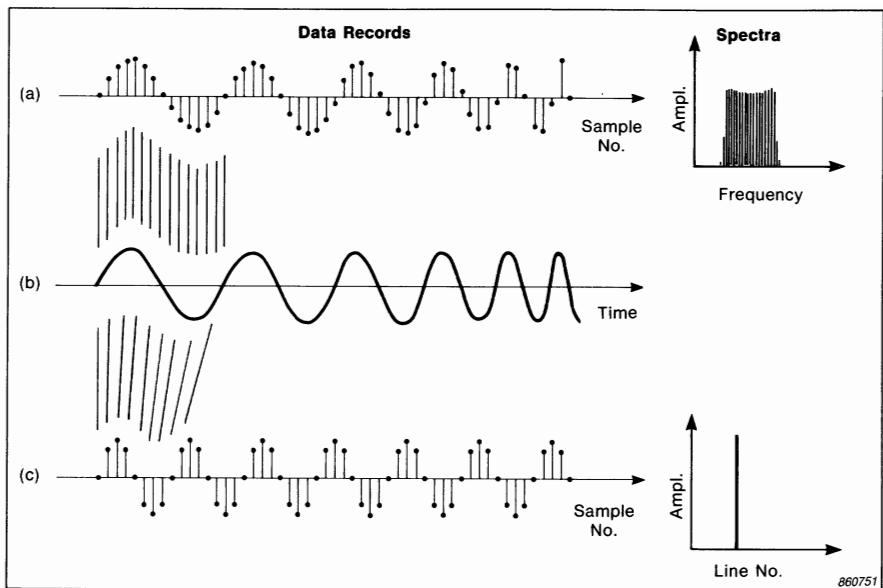


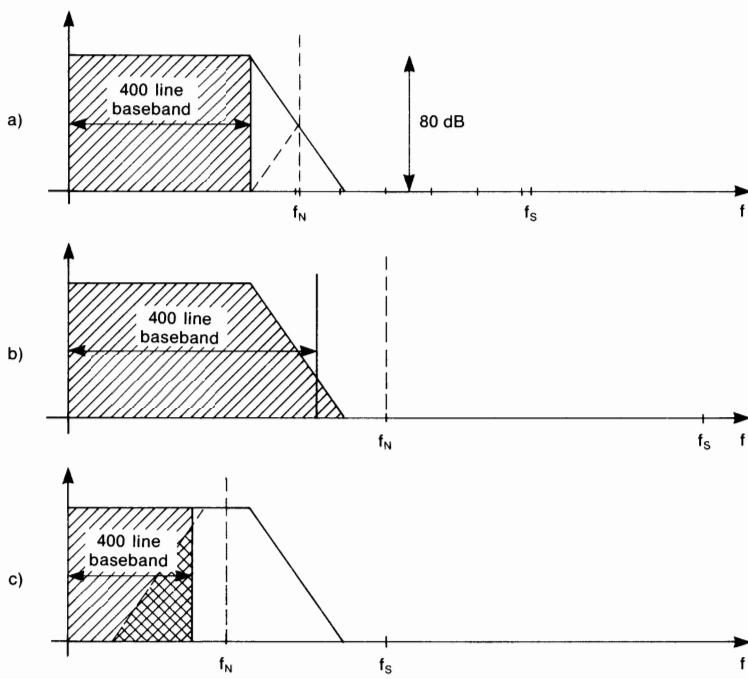
Fig. 6.7. Analysis of a fundamental component which is increasing in frequency
 (a) *Data record resulting from a uniform sampling rate, and its spectrum which spreads over a frequency band corresponding to the speed change*
 (b) *The original time signal*
 (c) *Data record resulting from sampling 8 times per fundamental cycle, and its spectrum which is concentrated in one analysis line*

The signal used to control the sampling of the A/D converter can be generated by a "tracking ratio adaptor", or "frequency multiplier/divider" which takes a synchronizing signal (typically a pulse once per shaft revolution) and produces an output signal whose frequency is multiplied by the set ratio. The choice of multiplying factor can most easily be determined as follows:

As mentioned in Section 4.5.1., the **line number** of a particular spectral component corresponds to the number of periods it represents in the transform size, N samples. Thus if the fundamental (i.e. first harmonic component) is to be located in Line No. n_1 , the multiplying factor, F , by which the fundamental frequency should be multiplied, is given by

$$F = \frac{N}{n_1} \quad (6.1)$$

so that every period of the fundamental frequency will occupy F samples along the memory.



860752

Fig. 6.8. Aliasing problems caused by order tracking (variation of sampling frequency f_s)

(a) *Baseband analysis with correct f_s*

(b) *Baseband analysis with f_s too high. Spectrum affected by lowpass filter*

(c) *Baseband analysis with f_s too low. Spectrum affected by aliasing components (double-hatched area)*

As mentioned in Section 4.3.1, when the sampling frequency varies with shaft speed, problems may be encountered with aliasing, and it is necessary to take special precautions. Fig.6.8(a) illustrates the situation for normal fixed frequency sampling with optimal choice of the anti-aliasing filter. For illustration purposes, the case is taken of a 400-line analyzer such as the B & K Type 2033. The lowpass filter characteristic folds around the Nyquist frequency f_N ($= f_s/2$) but is attenuated by 80 dB before it folds back into the measurement range. If the sampling frequency, f_s , were tied to shaft speed, however, and increased, the upper part of the spectrum would be influenced by the fixed lowpass filter, as illustrated in Fig.6.8(b). If f_s decreased, the aliasing components could fold back into the measurement range (as illustrated by the double cross-hatched area in Fig.6.8(c)) and once again affect the upper part of the spectrum.

The optimum way of solving this problem is to use a tracking low-pass filter whose cutoff frequency is also tied directly to shaft speed, but it is difficult to obtain a tracking filter with such a steep characteristic (120 dB/octave) and it may be necessary to use one of the following alternatives.

6.2.1. Optimum Choice of Internal Filter

If the analyzer has a series of fixed analog filters in a 2,5,10 sequence (for example B & K Type 2033), the largest step is 2,5:1 (from 2 to 5). Even so, it will be found that it is always possible to choose a filter such that at least 60% of the spectrum (in this case 240 lines) is valid. Consequently, if the multiplying factor is chosen (using Eq.(6.1)) such that the highest harmonic of interest is located below Line No.240, the optimum low-pass filter can be selected at any time (manually or automatically) dependent on the current shaft speed, and line numbers above 240 disregarded.

6.2.2. Zoom tracking

For analyzers where it is possible to select a lower frequency range than that determined by the internal analog low-pass filters, e.g. using zoom, another simpler procedure can be used. This technique can be used where the maximum frequency to be included in the analysis (the maximum shaft order at maximum shaft speed) is at least a factor of 4 below the maximum frequency of the analyzer. Figure 6.9 illustrates the principles.

Figure 6.9 can be applied directly to the Analyzers Types 2032/2034 but the same principles can be applied more widely. If a fixed lowpass filter with cutoff frequency 6,4 kHz[†] (1/4 of the maximum frequency 25,6 kHz) is applied to the signal before analysis, and frequency range 6,4 kHz selected, there will be no influence on the measurement results for a sampling frequency of 65536 Hz (the normal sampling frequency corresponding to 25,6 kHz frequency range, as shown in Fig.6.9(a)). The characteristics of the external analog filter, and the internal digital filter will in fact very nearly coincide. As shaft speed (and thus sampling frequency f_s) decrease, the aliasing components (folded around the Nyquist frequency) move downwards towards the displayed part of the spectrum. Note, however, that while the analog filter characteristic remains constant (in this case at 6,4 kHz) the digital lowpass filter characteristic varies in proportion to the sampling frequency. Fig.6.9(b) shows the situation at 1/4 of the original shaft speed. The analog filter characteristics are now overlapping as they would for a normal 6,4 kHz baseband analysis, but are still well removed from the display range (which has shrunk to 1,6 kHz). As illustrated in Fig.6.9(c),

[†] Note that the built-in 6,4 kHz filters in the analyzers Types 2032/2034 are not of anti-aliasing quality, but that special filters can be fitted as a modification.

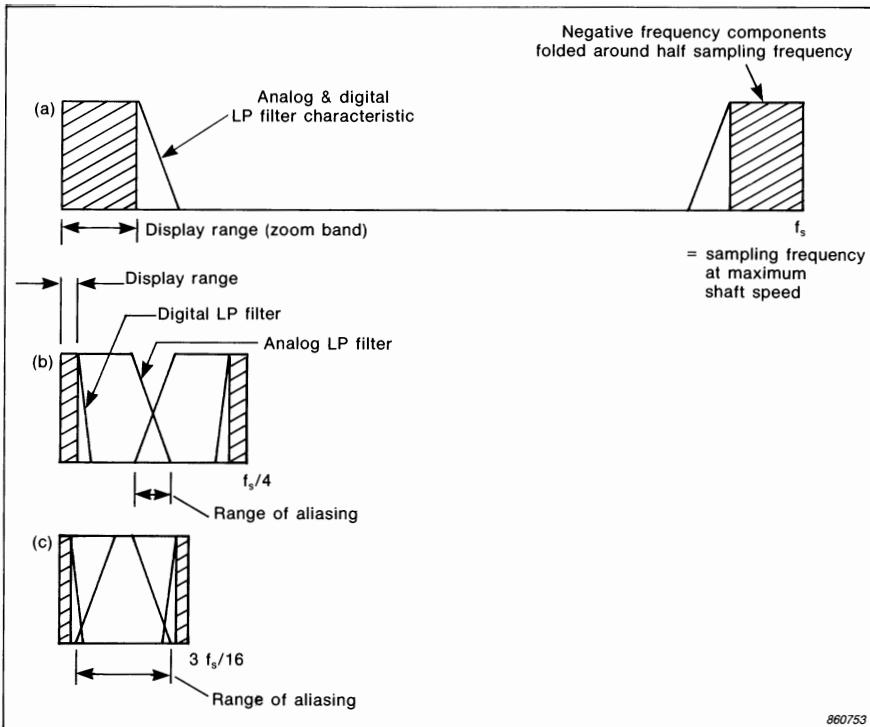


Fig. 6.9. Use of a fixed lowpass (LP) filter to prevent aliasing when tracking with an FFT analyzer employing zoom to analyze in a lower frequency band. For illustration purposes, the sampling frequency at maximum shaft speed has been made 4 times greater than that appropriate to the analog LP filter. 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 1/4 maximum shaft speed. The analog filter characteristics overlap, but are well separated from the display range
 (c) Situation at 3/16 maximum shaft speed. The aliasing range almost intrudes on the display range

the sampling frequency can be reduced to less than 3/16 of the original before there is any danger of aliasing components affecting the display range. The alias-free speed range is in fact given by the formula:

$$\frac{f_{s_{\max}}}{f_{s_{\min}}} = \frac{(2,56 D - 1)}{K} \quad (6.2)$$

where D is the decimation factor (4 in Fig.6.9) and K is the ratio of the frequency where the filter attenuation is adequate (typ. 75–80 dB) to its cutoff frequency ($K = 1,56$ in the example of Fig.6.9), giving a speed range of 5,9 in this case.

This leads to a slightly less simple method, which would allow alias-free tracking over any speed range. The fixed 6,4 kHz analog lowpass filter of Fig.6.9 could be replaced by a tracking lowpass filter with cutoff initially at 12,8 kHz (at maximum shaft speed). The filter steepness would only have to be of the order of 48 dB/octave to avoid aliasing, and this would be maintained over the entire speed range as the tracking filter characteristics would also vary in proportion to shaft speed. The selected display range could be anything less than the maximum by a factor of 2 or more (i.e. 12,8 kHz or lower).

Note that both these techniques have assumed that it is possible to select a multiplying factor such that the maximum sampling rate (at maximum shaft speed) corresponds to 65536 Hz. In practice there can be advantages in limiting the multiplying factors to powers of 2, to ensure both that the multiplying factor is an integer and that there will be an integer number of periods along the record length (the number of samples, N , is a power of 2 and therefore only divisible by powers of 2, see Eqn.(6.1)). Thus, in this situation the maximum frequency corresponding to maximum shaft speed can only be determined to lie between 65536 Hz and one-half this value; the other parameters must be modified accordingly.

Note that in the situation described here, Equation (6.1) should be modified as follows to take account of the fact that the initial sampling rate of the analyzer is higher than the final sampling rate in the Fourier transformed record. Thus:

$$F = \frac{D \cdot N}{n_1} \quad (6.3)$$

where D is the decimation factor $\frac{25,6 \text{ kHz}}{f_{\max}}$, and f_{\max} is the selected frequency band on the analyzer. D is normally a power of 2 (in any case for the Analyzers Types 2032/2034).

From Eqn.(6.3) it can be seen that for Line No. values n_1 , other than powers of 2, a non-integer (though rational) multiplying factor must be used, requiring a multiplier/divider. The advantage of having an integer value for n_1 is that there will be an integer number of periods of all harmonics in the record length, and thus rectangular weighting can be used. This will concentrate these harmonics in a single line, but it should be kept in mind that other discrete components (e.g. from geared shafts or constant frequencies) in general will fall between analysis lines and therefore have poor filter characteristics (cf. Fig.4.12). Thus, it can be an advantage to use another window, such as Hanning, in any case. Note that it is possible to choose a multiplying factor such that n_1 is non-integer, in which case even the harmonic orders will fall between analysis lines.

The best solution to this problem would be to use "flat-top" weighting which would ensure that all discrete frequency components, including shaft harmonics, would be represented by a peak of the correct height, and it would not be necessary to make "picket fence" corrections.

For analyzers with non-destructive zoom, such as the B & K Types 2033 and 2515, a procedure similar to that illustrated in Fig.6.9 can be employed, where zoom is used to select a frequency range lower than that determined by the input sample rate (the baseband frequency range). In this case, an internal analog lowpass filter corresponding to the zoom band can be chosen.

It should be noted that with non-destructive zoom, the entire memory normally has to be replenished for each zoom analysis, and thus overlapping analyses are not possible. This would often limit the rate at which the original signal can change in order for the results (as a waterfall diagram) to be easily interpretable. Ref.6.3 gives further details of this technique.

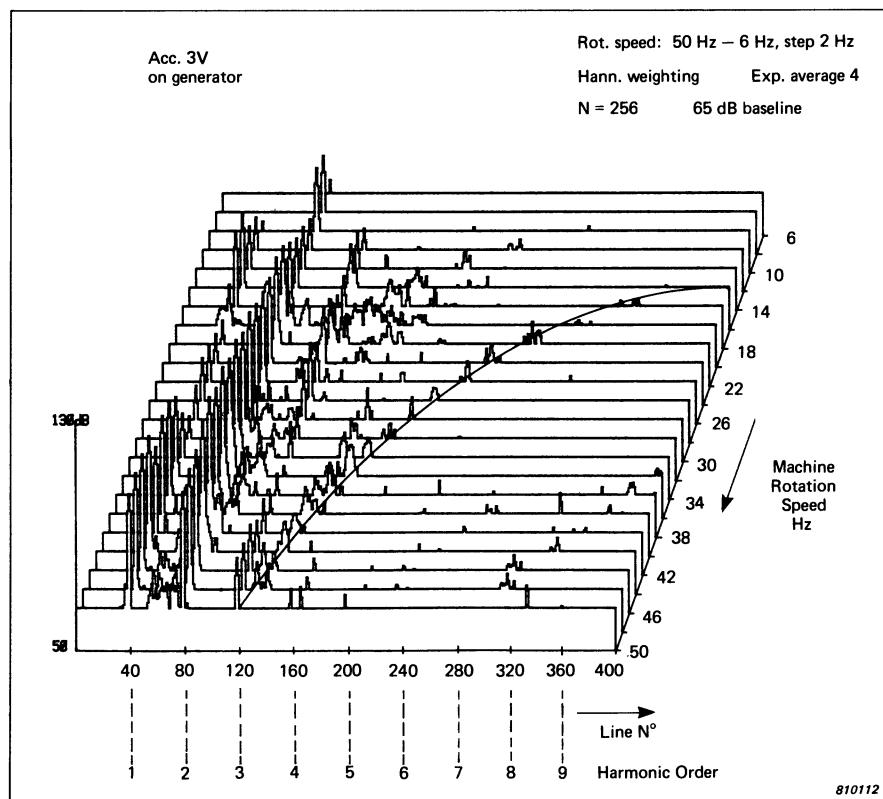


Fig. 6.10. Order analysis obtained during the rundown of a large turbo-generator, using zoom tracking on the 2033 Analyzer

Figure 6.10 (from Ref. 6.3) shows the application of this technique to a rundown of a large turbo-generator. With factor D corresponding to the zoom factor 10, $n_1 = 40$, and $N = 1024$ (Analyzer Type 2033), the multiplying factor used was 256. It is seen that the various harmonics remain in the same analysis line for the whole of the measurement (shaft speed range from 50 Hz down to 6 Hz), but that the amplitude of the components varies considerably with shaft speed as the latter runs through a number of "critical speeds". The curved line in the diagram traces the (hyperbolic) path of a constant (150 Hz) component, initially coinciding with the third harmonic of shaft speed. The fact that constant frequency lines follow a curved path means that the order tracking technique is slightly less suitable for separating them from shaft orders than the "waterfall" technique of Figure 6.5.

7. DUAL CHANNEL ANALYSIS

Simultaneous analysis of signals in at least two channels opens up new application areas, where it is no longer the signals themselves which are of primary interest, but rather the properties of the physical system responsible for the differences between them. The techniques can be extended to virtually any number of channels, but as they are basically analyzed two at a time, the topic is here referred to as dual channel analysis. The B & K Analyzer Type 2032 is an excellent example of a dual channel analyzer. Naturally, as is evident from the material of earlier chapters, a dual channel analyzer can be, and often is, used for single channel signal analysis.

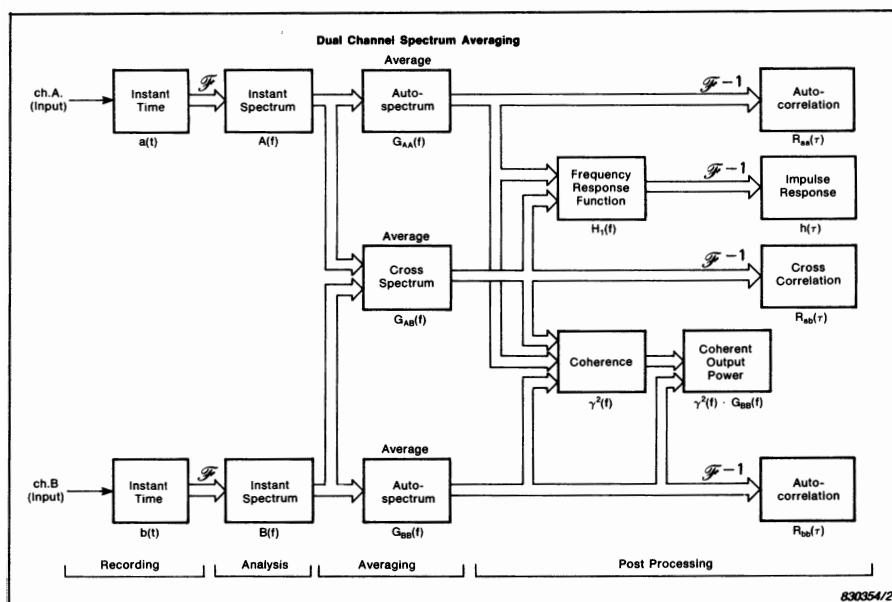


Fig. 7.1. Typical block diagram of a dual channel FFT analyzer in spectrum averaging mode

Figure 7.1 shows the basic internal structure of a typical dual channel analyzer, such as the 2032, in terms of the functions which are calculated (not all functions are shown). At the top and bottom of the diagram are what could be two parallel single channel analyzers, for Channels A and B, respectively. For each signal, the time function is first transformed (by a forward FFT) to a complex (instantaneous) spectrum. The squared amplitudes of a number of such instantaneous spectra are next averaged in an averaging buffer to give the "autospectrum" (alternatively known as power spectrum) for that particular signal. The autospectra can be further processed to give the "autocorrelation function" (see later) or the "power cepstrum" (see Chapter 8).

The basic new function, shown in the centre of the diagram, is the "cross spectrum", calculated from the instantaneous spectra of both channels. All other functions in the diagram are calculated (by post processing) on the basis of the two autospectra and the cross spectrum. In the following sections, the definitions, calculation procedures, properties, and major applications of the various functions are described.

7.1. CROSS SPECTRUM

7.1.1. Definitions and Calculation Procedures

In terms of the (complex) instantaneous spectra $A(f)$ and $B(f)$, respectively, the cross spectrum S_{AB} ("from A to B") is defined by the formula:

$$S_{AB}(f) = A^*(f) \cdot B(f) \quad (7.1)$$

Thus its amplitude is the product of the two amplitudes, and phase the difference of the two phases (from A to B). The cross spectrum S_{BA} ("from B to A") would thus have the same amplitude, but opposite phase. As for autospectra (see Fig.2.5) the cross spectrum may be expressed in a one- or two-sided form, and the commonly used one-sided form is often termed $G_{AB}(f)$, where $G_{AB}(f)$ is defined by:

$$\begin{aligned} G_{AB}(f) &= 0, & f < 0 \\ G_{AB}(f) &= S_{AB}(f), & f = 0 \\ G_{AB}(f) &= 2 S_{AB}(f), & f > 0 \end{aligned} \quad (7.2)$$

Similar relationships hold for the autospectra.

The real part of $G_{AB}(f)$ is known as the "coincident (or co-) spectrum", while the imaginary part is termed the "quadrature (or quad-) spectrum".

Strictly speaking, Eqn. (7.1) gives only a single estimate of the cross spectrum (such as would apply to a pair of transients). Where the two signals are stationary random, or a series of transients with some added noise, a better estimate of the cross spectrum can be obtained by averaging over a number of records (as for autospectra, see Fig.4.20). In general, when dealing with signals containing noise, the result can have both a systematic, or **bias** error, and a **random** error. The meanings of these two terms are illustrated in Fig.7.2 (from

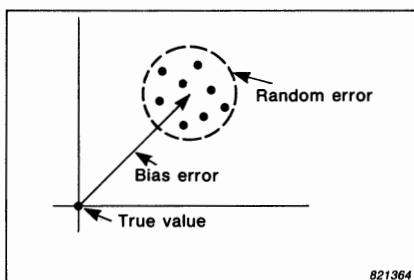


Fig. 7.2. Illustration of bias error and random error

Ref.7.1). Ref.7.1 discusses this question in some detail, making use of formulae developed by Bendat & Piersol (Refs.7.2, 2.2). For stationary, gaussian random signals, the random error in the cross spectrum is given by the formula:

$$\epsilon_r [| G_{AB}(f) |] = \frac{1}{\sqrt{\gamma^2(f)n_d}} \quad (7.3)$$

where $\epsilon_r [| G_{AB}(f) |]$ is the normalised error in the magnitude of the cross spectrum, and $\gamma^2(f)$ is the **Coherence function** (see later). n_d is the equivalent number of independent averages. Except for the addition of the coherence function, this is the same as the error in the **mean square** value of autospectral estimates (Eqn.(B.10) in Appendix B). The $\gamma^2(f)$ term allows for the fact that there will also be some random variation of the phase angle, when the two signals are not fully coherent (autospectral estimates are scalars, and therefore have no phase variation).

For linear systems, there is generally no bias error in the cross spectral estimates, provided the analysis is made with sufficient resolution (see the later discussion in connection with Fig.7.15).

7.1.2. Applications

The cross spectrum is not very widely used in its own right, it is mainly used as a basis for calculating other functions, but in general it can be said that the amplitude $|G_{AB}|$ gives a measure of how well the two functions correlate as a function of frequency, and the phase angle $\angle G_{AB}$ a measure of the phase shift between the two signals as a function of frequency. Since a time delay corresponds to a certain slope in a phase spectrum, the slope of the phase of the cross spectrum is sometimes used as a measure of time delay, in particular when this varies somewhat with frequency, such as in the case of dispersive systems (Ref.7.2). In cases where the signals in channels A and B represent an input and output, respectively, to a physical system, the **frequency response function** (see later) would generally be preferred to describe the relationship between the two signals in the frequency domain. The cross spectrum is a more symmetrical function which can be applied when there is no known cause/effect relationship between the two signals (e.g. response measurements at two points on a structure).

An important application area for the cross spectrum is its use in estimating **sound intensity** in the situation where $a(t)$ and $b(t)$ are sound pressure signals from closely spaced microphones (Ref.7.3). It can be shown that a finite difference approximation of the sound intensity is given by the (frequency weighted) value of the imaginary part of the cross spectrum. Ref.7.1 discusses this application in more detail and lists a number of other references.

7.2. COHERENCE

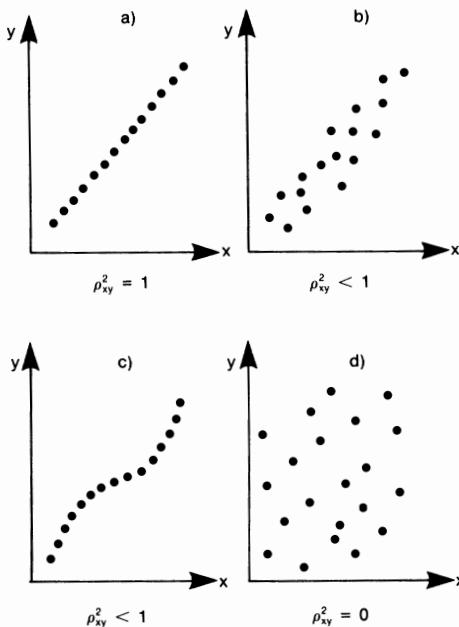
7.2.1. Definition and Properties

As implied in the previous section, the coherence gives a measure of the degree of linear dependence between the two signals, as a function of frequency. It is calculated from the two autospectra and the cross spectrum by the formula:

$$\gamma^2(f) = \frac{|G_{AB}(f)|^2}{G_{AA}(f) \cdot G_{BB}(f)} \quad (7.4)$$

As explained in Ref.7.1, the coherence (at each frequency) can be interpreted as a (squared) correlation coefficient (which expresses the degree of linear relationship between two variables) where the autospectral estimates correspond to the variances of the two variables, and the cross spectral estimate to the covariance. Fig.7.3 shows how the squared correlation coefficient ρ_{xy}^2 varies for different relationships between the variables x and y , and an equivalent relationship holds for the coherence function. In Fig.7.3(a) there is a perfectly linear relationship between x and y and $\rho_{xy}^2 = 1$. In (b) and (c), $\rho_{xy}^2 < 1$.

$$\text{Correlation Coefficient } \rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \cdot \sigma_y}$$



860769

Fig. 7.3. Dependence of the correlation coefficient ρ_{xy} on the linearity of the relationship between 2 variables x and y

but for different reasons. In (b) the relationship is reasonably linear, but there is some random spread due to added noise. In (c) there is no random spread, but the relationship is non-linear. In (d) there is no relationship whatsoever, and $\rho_{xy}^2 = 0$.

For a single estimate, Eqn. (7.1) indicates that:

$$\begin{aligned} |S_{AB}(f)|^2 &= |A(f)|^2 \cdot |B(f)|^2 \\ &= S_{AA} \cdot S_{BB} \end{aligned} \tag{7.5}$$

and therefore the coherence would always be unity. However, Fig. 7.4 shows that when the calculation of $\gamma^2(f)$ is based on averaged functions, where the individual estimates of G_{AB} are influenced by noise, the variation in phase angle

leads to the result that the modulus $|G_{AB}|$ would generally be less than that where no noise is present, the latter case corresponding to the situation where $\gamma^2 = 1$ (with all estimates adding up in phase). In Ref.7.1 it is shown that non-linearities give a similar result.

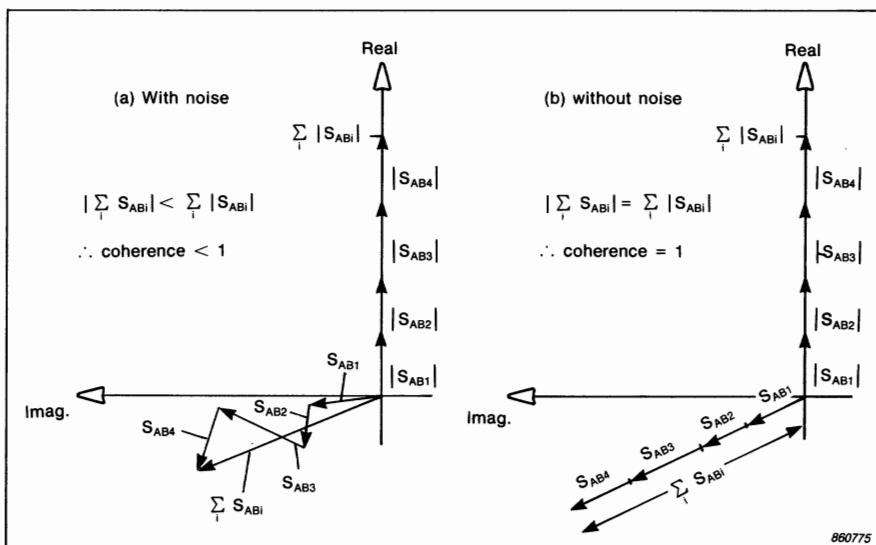


Fig. 7.4. Effect of noise in reducing coherence

Figure 7.5 (from Ref. 7.1) expresses the normalised error $\epsilon_r [\hat{\gamma}^2(f)]$ in the coherence estimate $\hat{\gamma}^2(f)$ as a function of the true coherence $\gamma^2(f)$ and the number of independent averages n_d . This applies to the effects of random noise in the two signals $a(t)$ and $b(t)$. The estimate is usually unbiased if the calculation is done with sufficient accuracy.

Reasons why the coherence $\gamma^2(f)$ may be less than unity include:

- 1) The presence of uncorrelated noise in $a(t)$ and/or $b(t)$.
- 2) A non-linear relationship between $a(t)$ and $b(t)$.
- 3) Leakage due to insufficient resolution, and/or wrong choice of window function.

This is a particular form of non-linearity introduced by the analysis. An example is given in Fig.7.17.

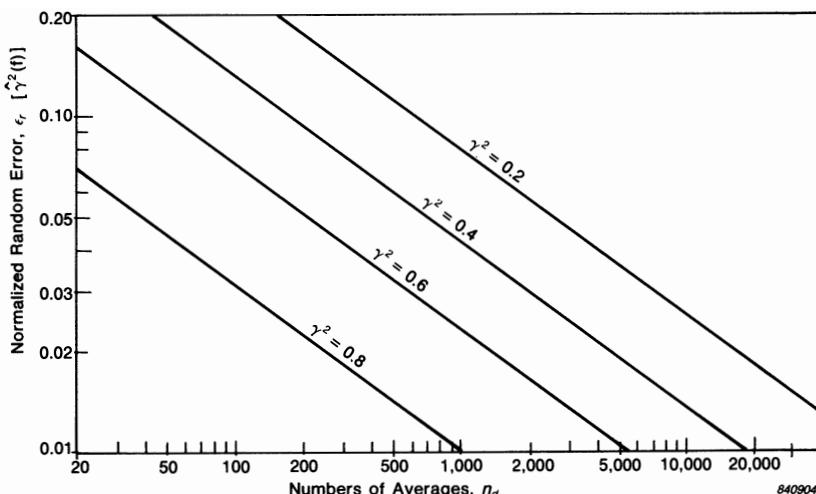


Fig. 7.5. Normalized random error $\epsilon_r [\hat{\gamma}^2(f)]$ of Coherence Function estimates as a function of number of averages n_d , for different values of (true) Coherence $\gamma^2(f)$

- 4) Time delay between the two signals, where this is of the same order as the length of record. Fig.7.6 illustrates that even when signal $b(t)$ results directly from $a(t)$, the apparent relationship will not be very strong if the delay τ between $b(t)$ and $a(t)$ is large with respect to the record length T . In fact, for white noise signals, Ref.7.2 shows that the estimated coherence will be lower in the proportion $(1 - \tau/T)^2$. This bias error can be removed by applying what is known as "pre-computational delay" to the signal $b(t)$ so that the sections of signal analyzed correspond to each other (Fig.7.6).

Pre-computational delay may be applied in the analyzers Types 2032/2034, with automatic adjustment of the time scales of such functions as cross correlation and impulse response (see later).

Where the output signal $b(t)$ contains reverberant components (i.e added coherent signals with different delay times) then it is not possible to fully compensate for them by a simple delay. The bias error will be small if the record length is made as long as the reverberation time (Ref.7.2), for example by zooming, but even where this is not done, the error can be minimised by using pre-comp. delay corresponding to the "mean" delay time. This can best be determined empirically by maximising the coherence.

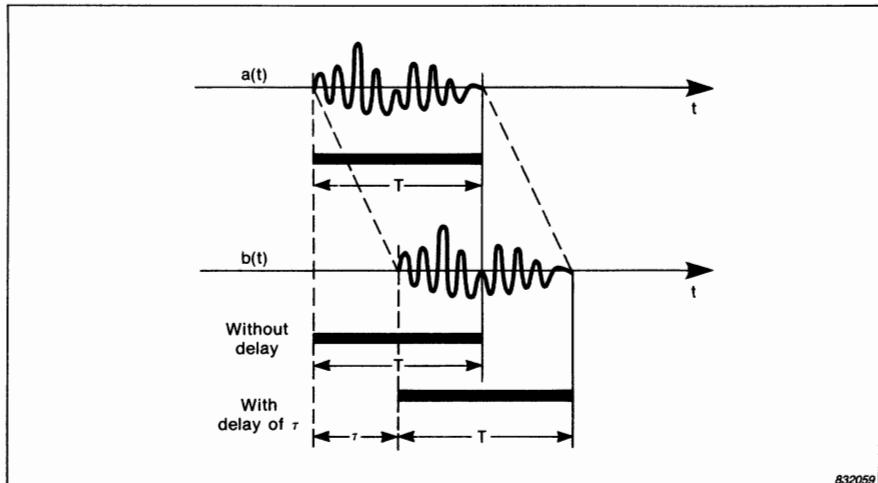


Fig. 7.6. Illustration of how system delays must be compensated for to avoid a reduction in coherence

7.2.2. Applications

The main application of the coherence function is in checking the validity of other functions, and in determining whether they are influenced by noise and/or non-linearities. Some examples of this are given in later sections. A low coherence does not necessarily mean that a measurement is invalid, but will sometimes indicate that a large number of averages are required to give a valid result (e.g. Eqn.(7.3)).

The coherence is also used to generate a number of derived functions which have various applications.

One of these is the so-called **Coherent Output Power** defined by:

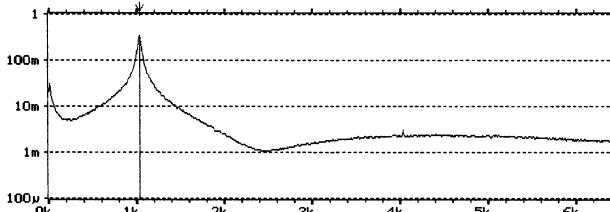
$$\text{COP} = \gamma^2 \cdot G_{BB}(f) \quad (7.6)$$

Thus, it gives a measure of what part of a measured (output) autospectrum, $G_{BB}(f)$, is fully coherent with a particular (input) signal represented by $a(t)$ and with autospectrum $G_{AA}(f)$.

Fig. 7.7. Example of the use of Coherent Output Power to remove the effects of noise added to the output signal

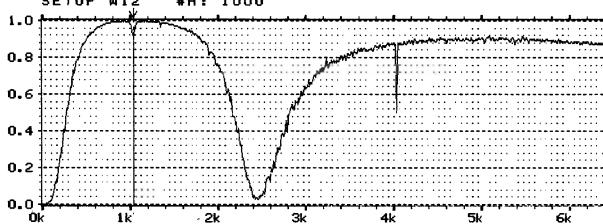
- (a) Measured output spectrum contaminated with added white noise
- (b) Coherence obtained in the measurement of (a)
- (c) Coherent output power
- (d) Output spectrum with no added noise

W12 AUTO SPEC CH.B INPUT MAIN Y: 342mV
 Y: 1.00V RMS 80dB X: 1032Hz
 X: 0Hz + 6.4kHz LIN
 SETUP W12 #A: 1000 ELEM #: 129



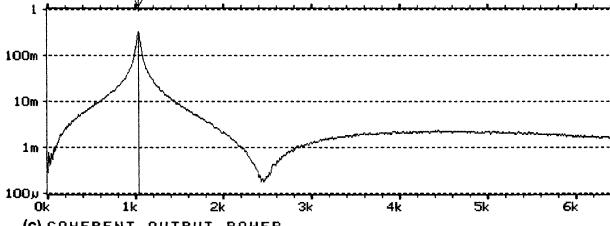
(a) OUTPUT SPECTRUM CONTAMINATED WITH WHITE NOISE

W2 COHERENCE MAIN Y: 916m
 Y: 1.00V RMS 80dB X: 1032Hz
 X: 0Hz + 6.4kHz LIN
 SETUP W12 #A: 1000



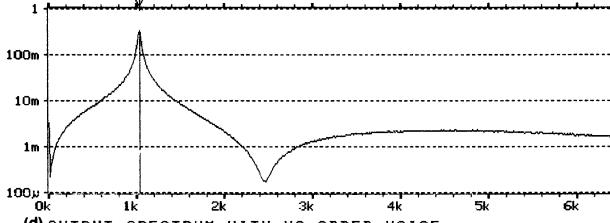
(b) Coherence

W12 COHERENT POWER STORED MAIN Y: 327mV
 Y: 1.00V RMS 80dB X: 1032Hz
 X: 0Hz + 6.4kHz LIN
 SETUP S12 #A: 1000 ELEM #: 129



(c) COHERENT OUTPUT POWER

W2 AUTO SPEC CH.B MAIN Y: 335mV
 Y: 1.00V RMS 80dB X: 1032Hz
 X: 0Hz + 6.4kHz LIN
 SETUP W12 #A: 1000 ELEM #: 129



(d) OUTPUT SPECTRUM WITH NO ADDED NOISE

860687

Figure 7.7 shows an example where the output autospectrum, obtained by applying a white noise signal to an electrical circuit, is contaminated by the addition of (white) noise from an uncorrelated source. Fig.7.7(a) shows the spectrum as measured, (b) shows the coherence, (c) the coherent output power, and for comparison purposes (d) shows the output auto-spectrum without the contaminating noise. This is seen to be very similar to (c), except at the resonance peak, where there is a lower coherence due to leakage. This emphasizes that coherent output power is only valid where the low coherence is entirely due to noise contamination of the output signal.

Fig.7.8 is an example of another application, the use of coherent output power to extract the harmonics of a particular shaft speed from a signal also containing noise and non-harmonically related frequency components. The signal used as input was a tachometer signal, consisting of a series of once-per-rev pulses, which had strong harmonics throughout the frequency range of interest. The result is in some ways similar to that achieved by synchronous averaging, but does not give the phase relationships of the various harmonics.

Ref.7.4 describes the application (and limitations) of the coherent output power technique to the separation of the contributions of a variety of inputs to a given measured spectrum, e.g. the sound measured at a particular point. It is pointed out that if the signals used to characterize the various inputs are in any way correlated (i.e. coherent), then much more complex "multiple coherence" and "partial coherence" techniques must be used. The latter are based on measurements of the coherence between all known inputs and outputs, and are fully described in Ref.7.2.

Another more direct application of the coherence function is to calculate the **Signal-to-Noise Ratio** defined as:

$$S/N = \frac{\gamma^2}{1-\gamma^2} \quad (7.7)$$

Where noise in the measured output is assumed to be the only factor influencing coherence, then the coherent output power (proportional to γ^2) gives a measure of the "signal" appearing in the output, while the **non-coherent power** (proportional to $1-\gamma^2$) gives a measure of the "noise" in the output. Figure 7.9 gives the S/N ratio for the case of Fig.7.7, and shows that where the S/N ratio is less than 1 (0 dB), the noise contaminated auto-spectrum departs significantly from the uncontaminated spectrum (cf.Fig.7.7(d)).

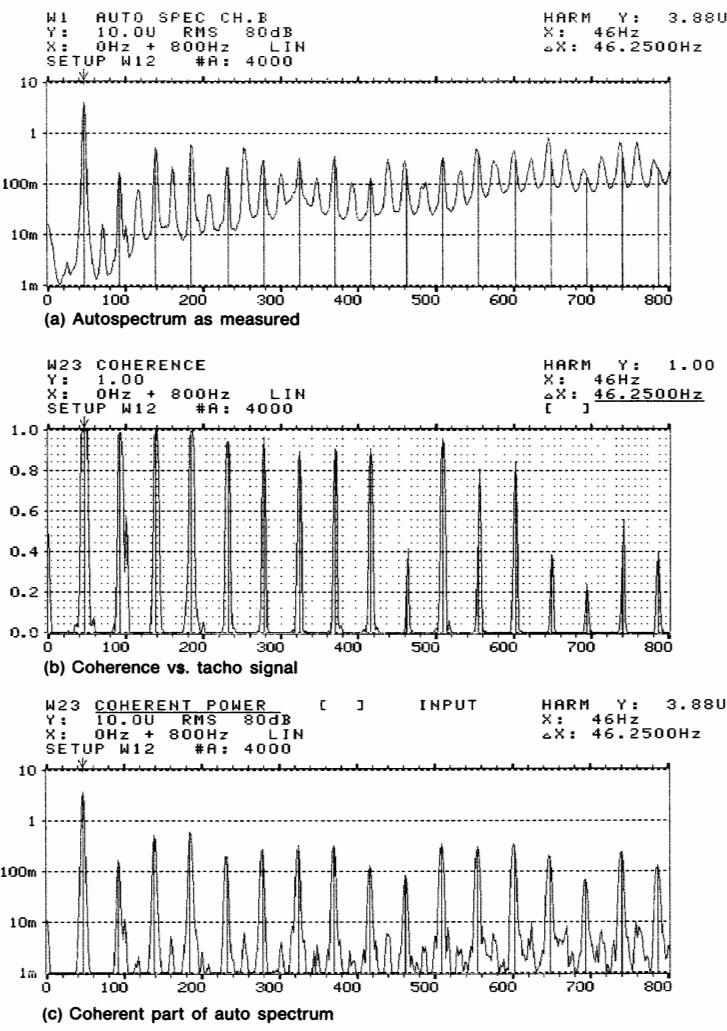


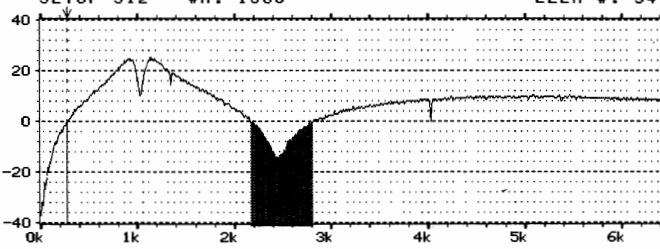
Fig. 7.8. Example of the use of Coherent Output Power to extract the harmonics of a particular fundamental frequency (defined by a tachometer signal)

(a) *Autospectrum as measured*

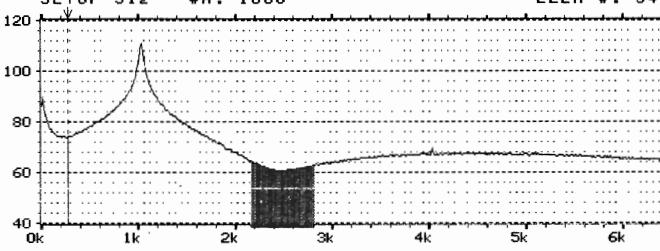
(b) *Coherence between the vibration signal and the tachometer signal*

(c) *The coherent part of the autospectrum, viz. the harmonics in the original signal corresponding to the harmonics of the tachometer signal*

W1 SIG/NOISE RATIO STORED DELT Y: 0.1dB
Y: 40.0dB 80dB X: 272Hz
X: 0Hz + 6.4kHz LIN ΔX: 640Hz
SETUP S12 #A: 1000 ELEM #: 34



W2 AUTO SPEC CH. B STORED DELT Y: 74.3dB
Y: 120.0dB /1.00μV RMS 80dB X: 272Hz
X: 0Hz + 6.4kHz LIN ΔX: 640Hz
SETUP S12 #A: 1000 ELEM #: 34



860666

7.3. FREQUENCY RESPONSE FUNCTIONS

Probably the most important use of dual channel analysis is in the measurement of **Frequency Response Functions**. These represent the ratio of output-to-input in the frequency domain, and thus fully characterize stable linear, time-invariant physical systems (e.g. structural, electrical, or acoustical systems). Frequency response functions are sometimes known as “transfer functions”, but strictly speaking this term applies to the complete description in the Laplace domain, of which the frequency response functions represent a special case (the values along the imaginary axis of the Laplace plane, Ref.7.5 and Appendix C).

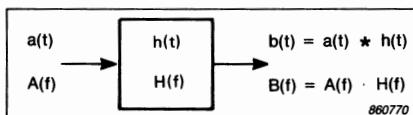


Fig. 7.10. Input – Output relations for a linear system

7.3.1. Definitions and Calculation Procedures

To assist in understanding the meaning of the frequency response function $H(f)$, Fig.7.10 shows the relationships between the input signal $a(t)$ and output signal $b(t)$ for a stable, linear, time-invariant system (henceforth termed an “ideal” system) in the absence of noise. The system is characterized by its **impulse response**, $h(t)$, and as described in connection with Eqn.(2.28), the output signal $b(t)$ is the convolution of $a(t)$ with $h(t)$, thus:

$$b(t) = a(t) * h(t) \quad (7.8)$$

By the convolution theorem, it follows that

$$B(f) = A(f) \cdot H(f) \quad (7.9)$$

where $H(f)$ is the Fourier transform of $h(t)$. Thus, in this situation, $H(f)$, the **frequency response function**, can be obtained from:

$$H(f) = \frac{B(f)}{A(f)} \quad (7.10)$$

In practice, there are found to be advantages in modifying Eqn.(7.10) in various ways. For example, if it is multiplied top and bottom by the complex conjugate of $A(f)$ a version known as H_1 is obtained:

$$\text{i.e. } H_1(f) = \frac{B(f)}{A(f)} \cdot \frac{A^*(f)}{A^*(f)}$$

$$= \frac{S_{AB}(f)}{S_{AA}(f)} = \frac{G_{AB}(f)}{G_{AA}(f)} \quad (7.11)$$

in other words, the cross spectrum normalised by the input autospectrum. In Ref.7.2 it is shown that in the presence of noise in the output signal, this definition minimises the error in the result.

If instead, the complex conjugate of $B(f)$ is used, another version known as H_2 is obtained (Ref.7.6):

$$\text{i.e. } H_2(f) = \frac{B(f)}{A(f)} \cdot \frac{B^*(f)}{B^*(f)}$$

$$= \frac{S_{BB}(f)}{S_{BA}(f)} = \frac{G_{BB}(f)}{G_{BA}(f)} \quad (7.12)$$

which involves the cross spectrum (from B to A) and the output auto-spectrum. It will be found that this version has advantages in other situations.

It is of interest to note that the ratio of H_1 and H_2 (always) equals the coherence, as follows:

$$\frac{H_1(f)}{H_2(f)} = \frac{\frac{G_{AB}(f)}{G_{AA}(f)} \cdot \frac{G_{BA}(f)}{G_{BB}(f)}}{\frac{G_{AB}(f) \cdot G_{AB}^*(f)}{G_{AA}(f) \cdot G_{BB}(f)}}$$

$$= \frac{|G_{AB}(f)|^2}{G_{AA}(f) \cdot G_{BB}(f)}$$

$$= \gamma^2(f) \quad \text{Q.E.D.} \quad (7.13)$$

It is perhaps worth emphasizing that although H_1 and H_2 in general have different properties, they "share" the same value of coherence. They also have the same phase spectrum, because

$$\angle(1/G_{BA}(f)) = \angle G_{BA}^*(f) = \angle G_{AB}(f) \quad (7.14)$$

and because the autospectra G_{AA} and G_{BB} are scalar quantities, the phase spectrum of $H(f)$ is the same as that of the cross spectrum, $G_{AB}(f)$, from Eqn.(7.11).

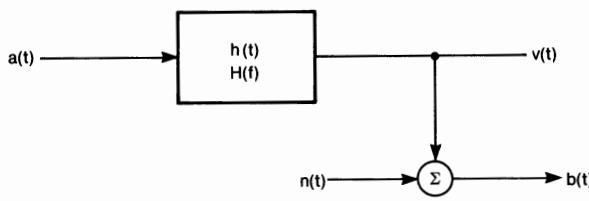
7.3.2. Effects of Noise on H_1 and H_2

Fig.7.11 illustrates the situation where the measured signal $b(t)$ includes not only signal $v(t)$, the output signal from the system in response to $a(t)$, but also some additive uncorrelated noise $n(t)$. $n(t)$ might also include some components actually transmitted by the system, but stemming from sources other than $a(t)$. As discussed in connection with Eqn.(7.3), $G_{AB}(f)$ gives an unbiased estimate of the true cross spectrum (across the system) $G_{AV}(f)$, and tends to it in the limit if sufficient averaging is performed. The output auto-spectrum $G_{BB}(f) = G_{VV}(f) + G_{NN}(f)$ (for uncorrelated signals the mean squares add directly) and so we have:

$$H_1(f) = \frac{G_{AB}(f)}{G_{AA}(f)} = \frac{G_{AV}(f)}{G_{AA}(f)} = H(f) \quad (7.15)$$

while,

$$\begin{aligned} H_2(f) &= \frac{G_{BB}(f)}{G_{BA}(f)} = \frac{G_{VV}(f) + G_{NN}(f)}{G_{VA}(f)} \\ &= H(f) + \frac{G_{NN}(f)}{H^*(f) \cdot G_{AA}(f)} \\ &= H(f) \left[1 + \frac{G_{NN}(f)}{|H(f)|^2 \cdot G_{AA}(f)} \right] \\ &= H(f) \left[1 + \frac{G_{NN}(f)}{G_{VV}(f)} \right] \end{aligned} \quad (7.16)$$



840152/1

Fig. 7.11. Ideal system with extraneous noise in measured output signal $b(t)$

Note that $G_{VV}(f)$ corresponds to the Coherent Output Power, and that the ratio $\frac{G_{NN}(f)}{G_{VV}(f)}$ is the reciprocal of the Signal-to-Noise ratio.

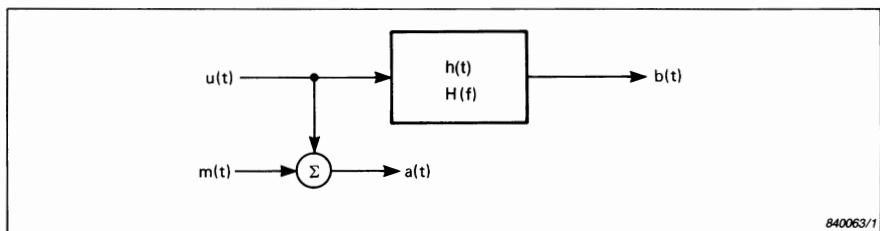


Fig. 7.12. Ideal system with extraneous noise in measured input signal a(t)

Figure 7.12 illustrates the opposite situation where the measured input signal $a(t)$ is contaminated by noise, $m(t)$, which does not contribute to the system response $b(t)$. By following a similar derivation to the previous case (Ref.7.1) it can be shown that in this case :

$$H_1(f) = H(f) \frac{1}{\left[1 + \frac{G_{MM}(f)}{G_{UU}(f)} \right]} \quad (7.17)$$

while, $H_2(f) = H(f) \quad (7.18)$

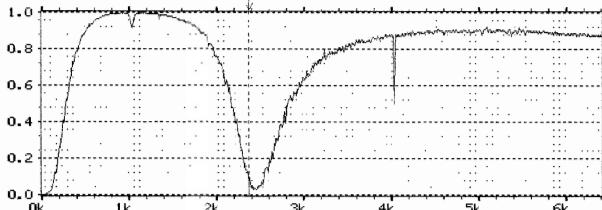
Similarly, for noise at both input and output, $H_1(f)$ is given by Eqn.(7.17) and $H_2(f)$ by Eqn.(7.16). In the absence of bias errors due to non-linearities (including leakage) $H_1(f)$ and $H_2(f)$ give a lower and upper bound, respectively, to the true value.

It should be noted that when the measured input signal is contaminated by noise, the concepts of Coherent Output Power, and Signal-to-Noise ratio, as defined in Equations (7.6) and (7.7) are no longer meaningful; (it can readily be shown that these definitions correspond to Equations (7.15) and (7.16)).

Fig. 7.13. Comparison of frequency response estimates H_1 and H_2 for the case of Fig.7.7 (noise added at output)

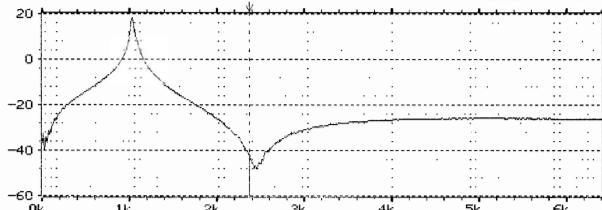
- (a) Coherence function with cursor placed in the vicinity of $\gamma^2 = 0,1$
- (b) Estimate H_1 which is good for $\gamma^2 > 0,1$
- (c) Estimate H_2 which is poor except at resonance peak
- (d) Estimate with no noise for comparison

W1 COHERENCE STORED MAIN Y: 99.4m
Y: 1.00 X: 2368Hz
X: 0Hz + 6.4kHz LIN ELEM #: 296
SETUP S12 #R: 1000



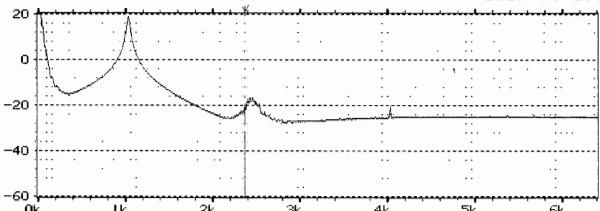
(a) COHERENCE WITH ADDED NOISE AT OUTPUT

W2 FREQ RESP HI MAG STORED MAIN Y: -42.0dB
Y: 20.0dB 80dB X: 2368Hz
X: 0Hz + 6.4kHz LIN ELEM #: 296
SETUP S12 #R: 1000



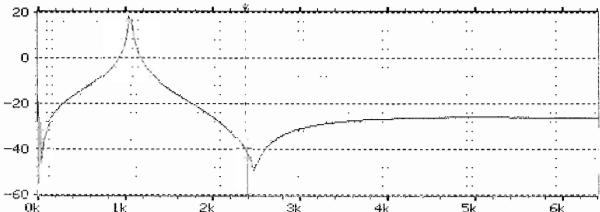
(b) H1 WITH ADDED NOISE AT OUTPUT

W1 FREQ RESP H2 MAG STORED MAIN Y: -21.9dB
Y: 20.0dB 80dB X: 2368Hz
X: 0Hz + 6.4kHz LIN ELEM #: 296
SETUP S12 #R: 1000



(c) H2 WITH ADDED NOISE AT OUTPUT

W2 FREQ RESP HI MAG MAIN Y: -42.3dB
Y: 20.0dB 80dB X: 2368Hz
X: 0Hz + 6.4kHz LIN ELEM #: 296
SETUP W12 #R: 100



(d) H1 WITH NO ADDED NOISE AT OUTPUT

Normally, when making measurements of frequency response, one has full control over the input signal, which is thus less likely to be influenced by noise, whereas the output signal is modified by the system response, and for example at anti-resonances is likely to be contaminated by noise. Thus, in these situations H_1 usually gives the best estimate over the whole frequency range, at least with respect to the effects of noise. Fig. 7.13 illustrates this for the same case as Figs. 7.7 and 7.9. After performing 1000 averages, the H_1 version is very close to the actual value of H , even in the vicinity of the anti-resonance, with the coherence as low as 0.1. H_2 is a poor estimate except in the vicinity of the resonance peak.

Figure 7.14 (From Ref. 7.1) gives the values of normalised (random) error $\epsilon_r [|\hat{H}(f)|]$ in the magnitude of either H_1 or H_2 as a function of the coherence $\gamma^2(f)$ and the number of independent averages n_d . Note that this is different from the **bias** errors (where applicable) expressed in Equations (7.16) and (7.17).

For $n_d = 1000$, and $\gamma^2(f) = 0.1$ (corresponding to the cursor position in Fig. 7.13) the random error from Fig. 7.14 is found to be 0.07 (0.6 dB), which corresponds well with the actual results for local random fluctuation.

There is at least one situation, however, where the measured input signal is likely to be contaminated by noise, and this occurs when lightly damped structures are excited with random noise by a shaker. At resonance peaks, the structure acts like a short circuit, and the input force spectrum is likely to be low (even when the generator signal feeding the power amplifier is white noise). The input signal can thus be contaminated by noise in the vicinity of resonance peaks, whereas the output signal is strong and usually free of noise. In these situations H_2 would generally be better than H_1 . Ref. 7.6 points out that the best solution is often a composite function where the peaks are taken from H_2 and the balance from H_1 .

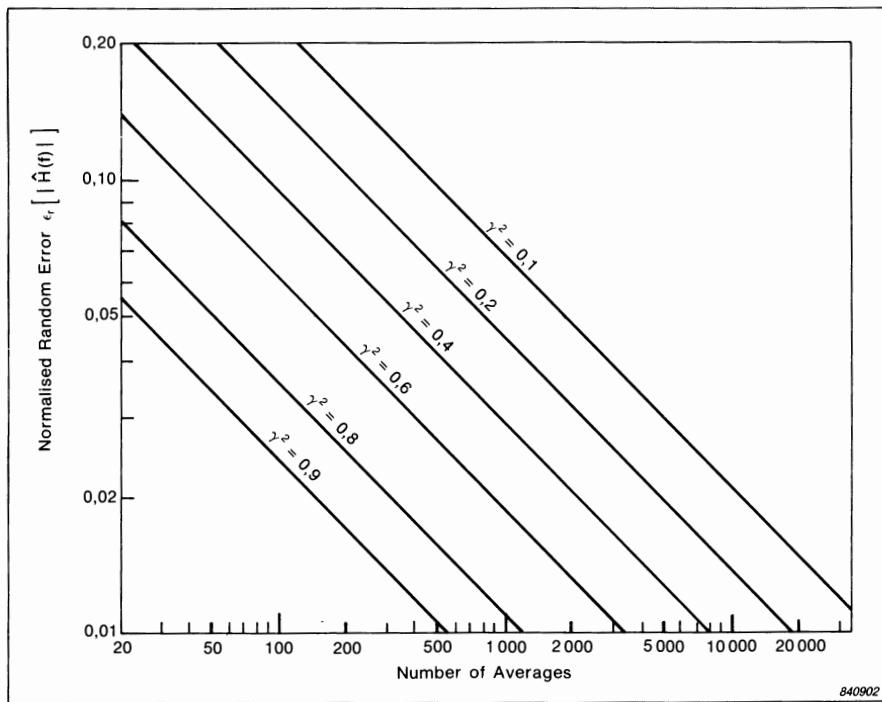


Fig. 7.14. Normalized random error $\epsilon_r [|\hat{H}(f)|]$ of magnitude of Frequency Response Function estimates ($H_1(f)$ or $\hat{H}_2(f)$) as a function of number of averages n_d for different values of Coherence $\gamma^2(f)$

7.3.3. Effects of Leakage on H_1 and H_2

So far, only the effects of noise have been considered. With respect to the effects of leakage, H_2 is almost always better than H_1 . Leakage problems tend to occur whenever the system is excited by a stationary random input signal, and where the resolution of the analysis (determined by a window function such as Hanning) is not sufficiently fine with respect to the resonance peaks being measured. Fig.7.15 illustrates this for the case of a flat input spectrum $G_{AA}(f)$. There is no leakage error in $G_{AA}(f)$, but the peaks in $G_{BB}(f)$ and $|G_{AB}(f)|$ are spread out as indicated by the dotted lines. In Ref.7.7, it is shown that in the calculation of H_2 , the errors in G_{BB} and $|G_{AB}|$ tend to cancel out and give the correct result, whereas in the calculation of H_1 , the error is in $|G_{AB}|$ only and therefore has a greater effect on the result. Ref.7.7 shows that this applies even where the input spectrum G_{AA} is not flat. Even though H_2 gives the correct peak value if it coincides with an analysis line, there can be a "picket fence" effect if the actual resonance falls between two lines, and Fig.7.16 (from Ref.7.7) compares the worst case errors for H_1 and H_2 against analysis resolution. It is found that in general a result with a given (maximum) error can be obtained from H_2 with about 3 times coarser resolution than for the same error in H_1 .

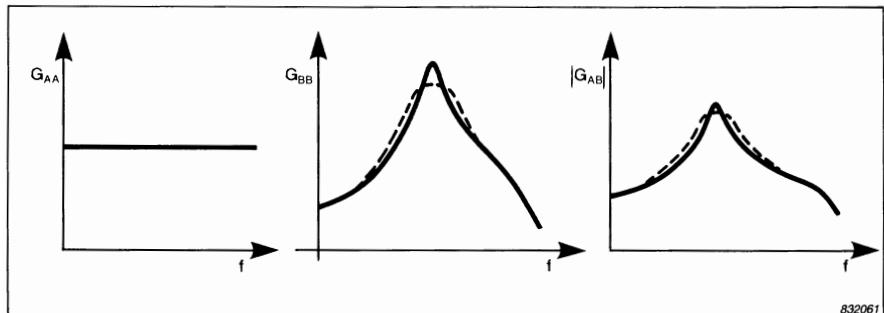


Fig. 7.15. Bias errors in $G_{BB}(f)$ and $|G_{AB}(f)|$ due to leakage. $G_{AA}(f)$ is assumed flat

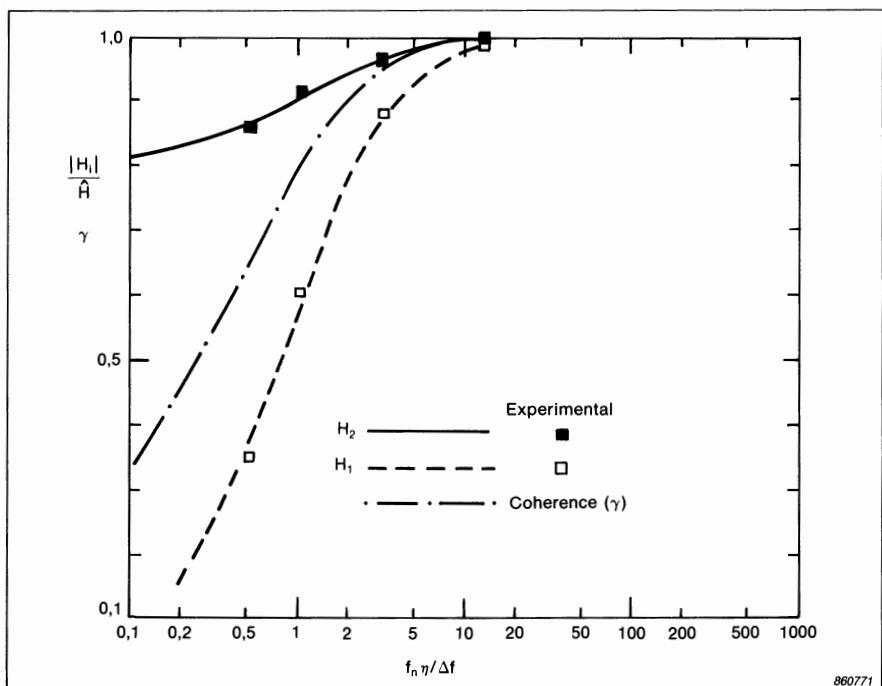


Fig. 7.16. (from Ref.7.7) Maxima of computed frequency response functions H_1 and H_2 (worst case value compared with the true value \hat{H}) vs. the resolution factor $f_n \eta / \Delta f$ for use of Hanning window

f_n = frequency of resonance peak

η = hysteretic loss factor defining width of resonant peak

Δf = analysis resolution (line spacing)

Note that the coherence is shown in terms of γ rather than γ^2

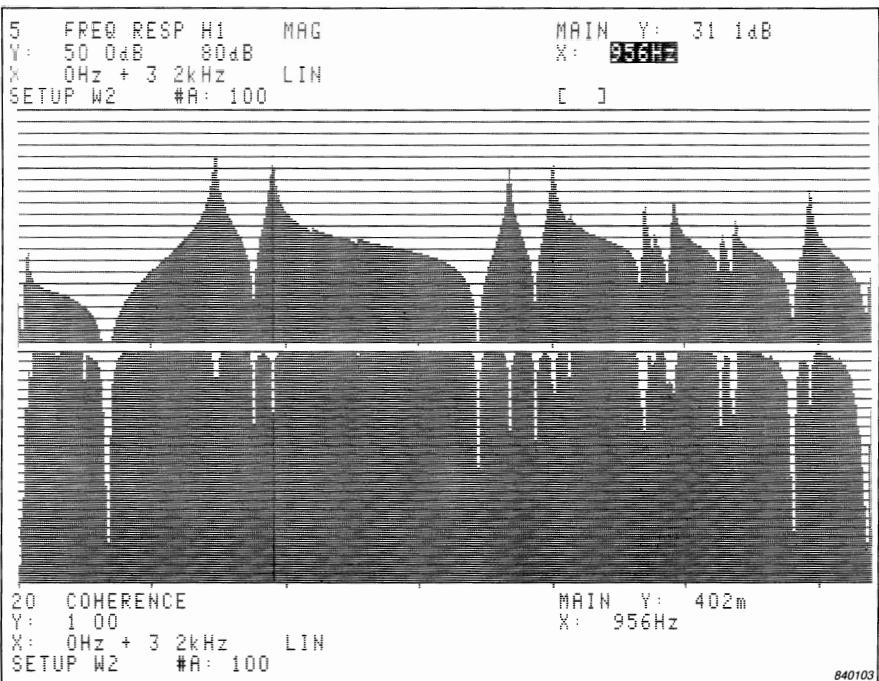


Fig. 7.17. Magnitude of Frequency Response Function estimate $H_1(f)$ and Coherence Function from a baseband measurement on a mechanical system. Random noise excitation

Figure 7.17 (from Ref.7.1) is an example showing the measurement of H_1 for a mechanical structure excited by a shaker with a random excitation. The coherence is seen to be low at all resonances (due to leakage) and anti-resonances (due to leakage and noise in the response signal). The resonant amplification of the peak selected by the cursor is 31,1 dB. Fig.7.18 shows H_2 for the same measurement (and thus having the same coherence), and for the same peak, the indicated amplification is 39,0 dB. Fig.7.19 shows a further measurement of H_1 obtained using zoom on the same peak, along with the coherence for this measurement. Since the coherence is 1 in this case, there is no leakage error and the value of H_1 (and H_2) would both be correct at 42,2 dB. Thus the value of H_2 in Fig.7.18 is only 3,2 dB in error, as compared with 11,1 dB error for H_1 in Fig.7.17.

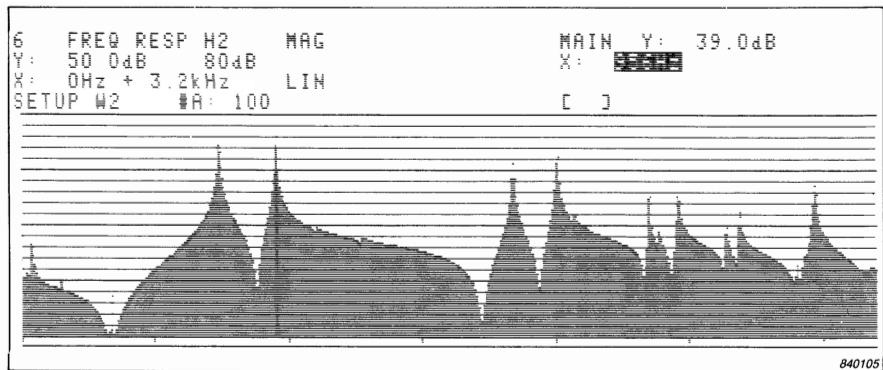


Fig. 7.18. Magnitude of Frequency Response Function estimate $H_2(f)$ for same measurement as Fig.7.17

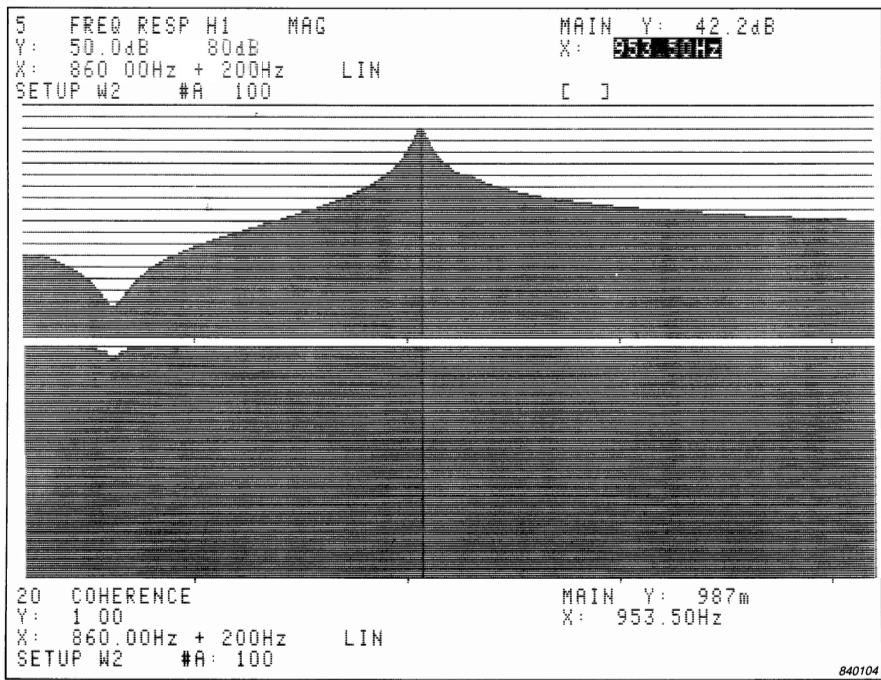
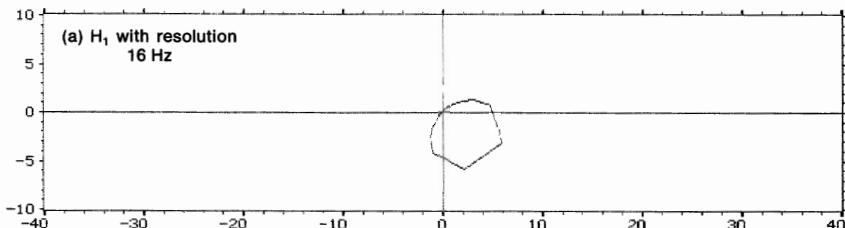


Fig. 7.19. Magnitude of Frequency Response Function estimate $H_1(f)$ and Coherence Function from a zoom measurement (Frequency span of 200 Hz) on the same system as in Figs.7.17 & 7.18. Random noise excitation

```

W12 FREQ RESP H1      NYQUIST     INPUT      MASK Y: 26.2m
Y: 10.0                F: 80Hz
X: 40.0                X: 1.58m
SETUP W12    #A: 100          ELEM #: 5

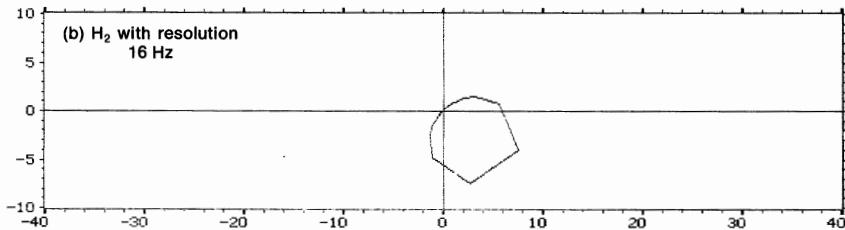
```



```

W2 FREQ RESP H2      NYQUIST      MASK Y: 27.4m
Y: 10.0                F: 80Hz
X: 40.0                X: 1.65m
SETUP W12    #A: 100          ELEM #: 5

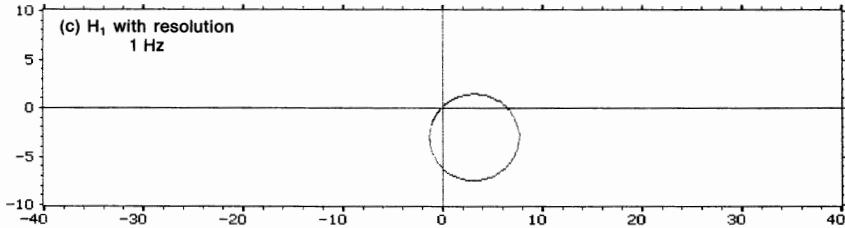
```



```

W12 FREQ RESP H1      NYQUIST     STORED      DELT Y: 251m
Y: 10.0                F: 624Hz
X: 40.0                X: 141m
SETUP S12    #A: 100          ELEM #: 0

```



860669

Fig. 7.20. Advantage of H_2 in a Nyquist plot

- (a) H_1 with poor resolution (16 Hz) showing distortion of Nyquist circle
- (b) H_2 with poor resolution (16 Hz) showing that the points lie on a circle
- (c) H_1 with good resolution (1 Hz) confirming that the H_2 circle in (b) is correct (though not the points on the circle)

One result of the analysis of Ref.7.7 is that when the values of H_2 are plotted in Nyquist format (real part vs imaginary part) they fall on the correct circle, even though not necessarily at the correct points. Thus by curve fitting a circle to the points, the correct value of resonant amplitude (residue) will be obtained (see later) even though it may correspond to an incorrect estimate of damping. Fig.7.20 shows examples of Nyquist plots of a resonance peak (of the system analyzed in Fig.7.7) with poor resolution such that H_1 differs from H_2 . Also shown is a correct result obtained using zoom. This has the same diameter as the other result for H_2 .

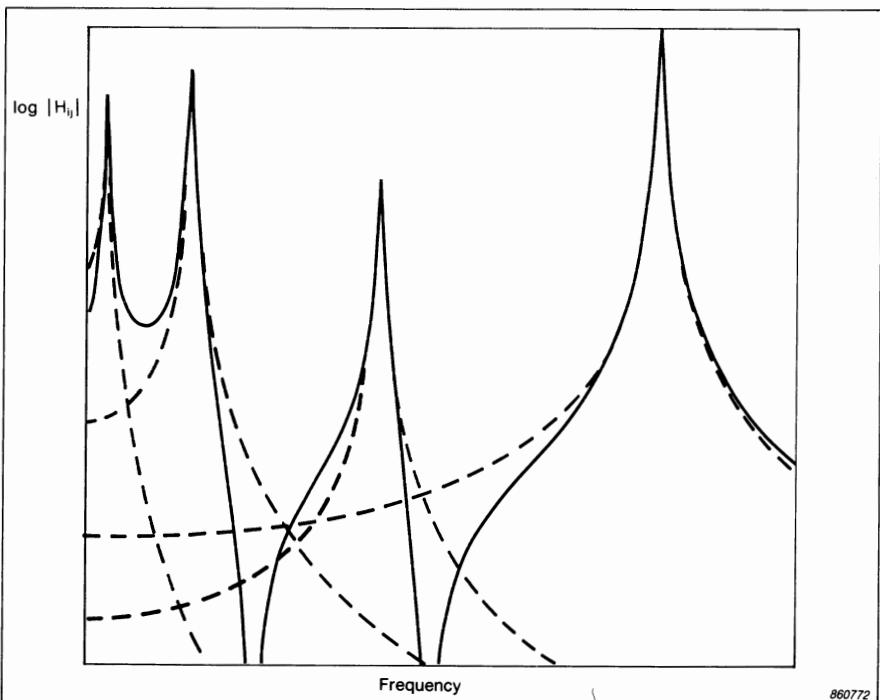
7.3.4. Applications of Frequency Response Functions

The main reason why frequency response functions are very widely used, is the simplicity with which the response of physical systems can be expressed in terms of them. One aspect of this, as evidenced by Equations (7.8) and (7.9) is that convolutions in the time domain reduce to multiplications in the frequency domain. More generally, as shown very briefly in Appendix C (and in more detail in Refs.7.5 and 7.8) for ideal physical systems whose properties can be described by systems of linear second order differential equations, application of the Laplace transform converts the differential equations to algebraic (polynomial) equations in the Laplace variable s . The solutions of these equations can be written in terms of “transfer functions”, $H_{ij}(s)$, which express the ratio of the response at point i to an input at point j . As shown in Appendix C, a typical transfer function for a system with n degrees of freedom can be represented as:

$$H_{ij}(s) = \sum_{k=1}^n \left[\frac{r_{ijk}}{s-p_k} + \frac{r_{ijk}^*}{s-p_k^*} \right] \quad (7.19)$$

where the p_k , or “poles”, are global properties common to all the transfer functions of the system, while the r_{ijk} , or “residues”, are specific to a particular transfer function. Each of the terms summed in Eqn.(7.19) represents the response of a single degree-of-freedom system with pole $p_k = -\sigma_k + j\omega_k$. The real part of this, $-\sigma_k$, represents the damping of the k^{th} mode, and the imaginary part, ω_k , its damped natural frequency (in radians/s). It can be shown that by substituting $j\omega$ for s in Eqn.(7.19), (thus evaluating it along the imaginary axis), and converting angular frequency ω to circular frequency f in Hz, the result is the equivalent frequency response function $H_{ij}(f)$, (see Appendix C).

As for the transfer function, the frequency response function can be interpreted as the sum of a number of components each equivalent to the response of a single degree of freedom system (Fig.7.21). In **Modal Analysis** applications, the global properties σ_k and ω_k can in principle be extracted from any of the $H_{ij}(f)$ by curvefitting, whereas the residues r_{ijk} define the mode shape ψ_k . It is not however necessary to measure all the H_{ij} , as it can be shown that the residues



860772

Fig. 7.21. The frequency response function H_{ij} represented as the sum of components corresponding to single degree-of-freedom systems

(and hence mode shapes) can be extracted from (almost) any row or column of the $[H(f)]$ matrix. Thus, it is possible to excite in one degree-of-freedom (DOF) only (for example with a shaker), and measure response (for example with an accelerometer) in all DOF's (each DOF corresponding to a point and direction). Alternatively, it is possible to measure response in one DOF only, and excite (for example with an instrumented hammer) at all DOF's. The restriction is that the particular row or column should not correspond to a node point in the mode shape (which nullifies all elements for that mode).

The system response in the frequency domain can thus be expressed (in terms of angular frequency ω) as:

$$\{X(\omega)\} = [H(\omega)] \{F(\omega)\} \quad (7.20)$$

where: $\{X(\omega)\}$ is the response vector

$\{F(\omega)\}$ is a vector representing the forcing functions

and $[H(\omega)]$ is the matrix of frequency response functions.

The elements of $[H(\omega)]$ can be synthesized from:

$$H_{ij}(\omega) = \sum_{k=1}^n \frac{r_{ijk}}{j\omega - (-\sigma_k + j\omega_k)} + \frac{r_{ijk}^*}{j\omega - (-\sigma_k - j\omega_k)} \quad (7.21)$$

even for frequency response functions which have not been measured (provided all relevant modes are included).

One major application of these techniques is so-called "Forced Response Simulation", which is effectively the application of Eqn. (7.20) to a prescribed set of forcing functions. Where the input forcing functions are uncorrelated random signals described only in terms of their autospectra $G_{jj}(f)$, the auto-spectrum of the response (at DOF i) can be obtained by summing the (mean square) effects of the various inputs, thus:

$$G_{ii}(f) = \sum_j G_{jj}(f) |H_{ij}|^2 \quad (7.22)$$

Modal models obtained from measured frequency response functions can also be used to verify and adjust analytical models, for example those based on Finite Element analysis.

7.4. CORRELATION FUNCTIONS

7.4.1. Definitions and Calculation Procedures

The **cross correlation** function $R_{ab}(\tau)$ gives a measure of the extent to which two signals correlate with each other as a function of the time displacement, τ , between them. For transient signals, the cross correlation function is defined by the formula:

$$R_{ab}(\tau) = \int_{-\infty}^{\infty} a(t) b(t + \tau) dt \quad (7.23)$$

For stationary signals, the formula is slightly modified (to convert from "energy" to "power") as follows:

$$R_{ab}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} a(t) b(t + \tau) dt \quad (7.24)$$

It can be shown, as follows, that the cross correlation function can alternatively be derived from the cross spectrum by inverse Fourier transformation, and this is the procedure used to calculate it in Fig.7.1. Eqn.(7.23) can be developed as follows:

$$R_{ab}(\tau) = \int_{-\infty}^{\infty} b(u) a(u-\tau) du, \quad \text{where } u = t + \tau \text{ and } du = dt$$

$$= \int_{-\infty}^{\infty} b(u) c(\tau-u) du, \quad \text{where } c(t) = a(-t)$$

By comparison with the convolution Equation (2.28), this is seen to be the equivalent of:

$$\begin{aligned} R_{ab}(\tau) &= b(\tau) * c(\tau) \\ &= b(\tau) * a(-\tau) \end{aligned} \tag{7.25}$$

Applying the Fourier transform and the Convolution theorem, this becomes:

$$\begin{aligned} \mathcal{F}\{R_{ab}(\tau)\} &= B(f) \cdot A(-f) \\ (\text{because when time runs backwards, phasors rotate backwards}) \\ &= B(f) \cdot A^*(f) \\ &= S_{AB}(f) \quad \text{Q.E.D.} \end{aligned} \tag{7.26}$$

Thus, the cross spectrum is the forward Fourier transform of the cross correlation function.

The **auto-correlation** function is a special case where $a(t) = b(t)$ and it is thus defined by

$$R_{aa}(\tau) = \int_{-\infty}^{\infty} a(t) a(t + \tau) dt \tag{7.27}$$

for **transients**, and

$$R_{aa}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} a(t) a(t + \tau) dt \tag{7.28}$$

for **stationary signals**.

The equivalent of Eqn. (7.26), viz.

$$\begin{aligned} \mathcal{F}\{R_{aa}(\tau)\} &= A^*(f) \cdot A(f) \\ &= S_{AA}(f) \end{aligned} \tag{7.29}$$

where $S_{AA}(f)$ is the autospectrum, is known as the "Wiener-Khinchin" relationship, at least for stationary signals (which require a somewhat more elaborate proof).

The correlation functions were previously calculated directly in the time domain, but since the advent of the FFT algorithm it has become much more efficient to go via the frequency domain, as indicated in Fig.7.1, making use of the relationships expressed in Equations (7.26) and (7.29). However, the circularity of the DFT (Section 4.3) leads to some special effects which must be taken into account when estimating the correlation functions by this means.

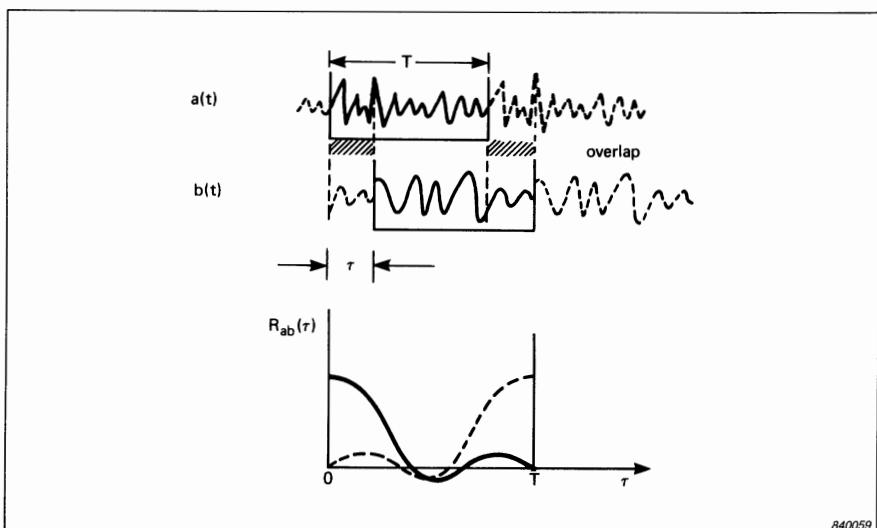
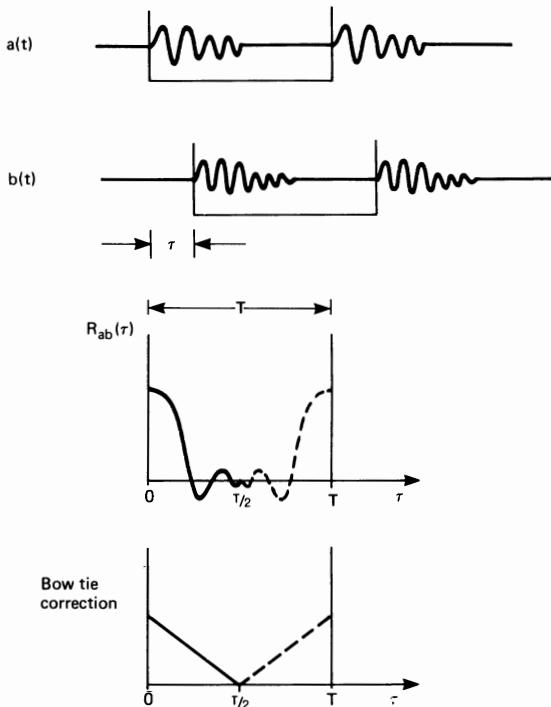


Fig. 7.22. Circular correlation effect. The Correlation Function is assumed to be even in the lower graph

Fig.7.22 illustrates the so-called “circular correlation” effect, which illustrates how the records are interpreted as being periodic, and thus when Channel B is displaced with respect to Channel A, it effectively overlaps the next period of the signal segment in Channel A, giving an apparent result not corresponding to the actual. Fig.7.23 illustrates how the use of so-called “zero padding”, whereby the second half of each record is set to zero, eliminates the overlap and thus gives a more correct result.

For transient signals, the circularity will be avoided if the frequency range is so chosen that the record length T is greater than the combined length of the two transients, and the result will be correct as measured. For stationary signals (to which zero padding is applied), the artificially introduced zero signal means that the total energy in the range where the two records overlap decreases in (inverse) proportion to the displacement τ , and this must be compensated for in order to obtain a correctly scaled result. The linearly decreasing function by



840060

Fig. 7.23. The use of zero padding to eliminate the circular correlation effect. The bow-tie correction is shown in the lower graph

which the measured correlation function should be divided is known as a “Bowtie correction” (because of its appearance in a 2-sided version) as illustrated in Fig.7.23. Note that zero padding doubles the bandwidth of the cross spectrum (and any other frequency functions) even though the line spacing remains the same.

When it is primarily the correlation functions which are of interest (as opposed to the frequency domain functions) it is usually advisable to use rectangular weighting in the initial analyses, as otherwise the correlations of the weighted time functions will be obtained.

7.4.2. Properties and Applications of Correlation Functions

7.4.2.1. Autocorrelation Function

Since the autocorrelation function is the inverse Fourier transform of the autospectrum (Eqn.7.29), the results are independent of the phase of the signal. Fig.7.24 illustrates this for three cases:

- (a) A sinusoidal signal
- (b) Band-limited white noise
- (c) Bandpass filtered white noise

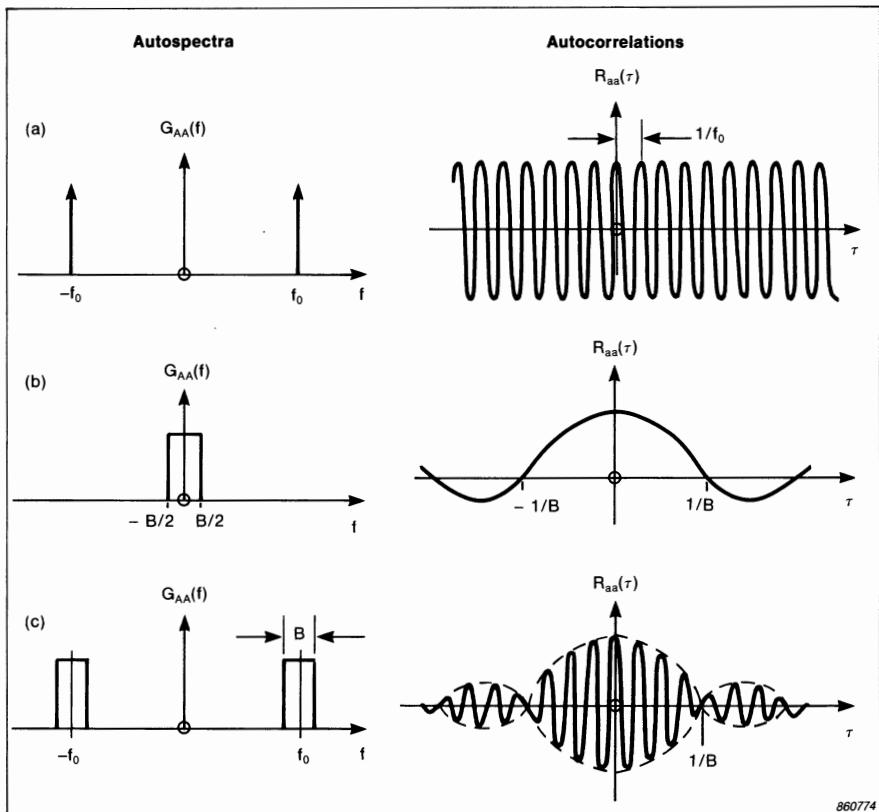


Fig. 7.24. Autospectra and Autocorrelation functions for various signals

- (a) Sinusoid, frequency f_0
- (b) Bandlimited white noise, cutoff frequency $B/2$
- (c) Bandpass filtered white noise, bandwidth B , centred on frequency f_0

The auto-spectrum for any sinusoid is the same as the spectrum of a cosine (apart from the scaling in squared units) and thus the autocorrelation function is always a cosine. Referring to Eqn.(7.28) it is seen that the value of autocorrelation for zero displacement, $R_{aa}(0)$, is equal to the mean square value of the signal $a(t)$. The function is often normalised to a maximum value of unity by dividing through by the mean square value, in which case it is strictly known as the "Autocorrelation Coefficient Function".

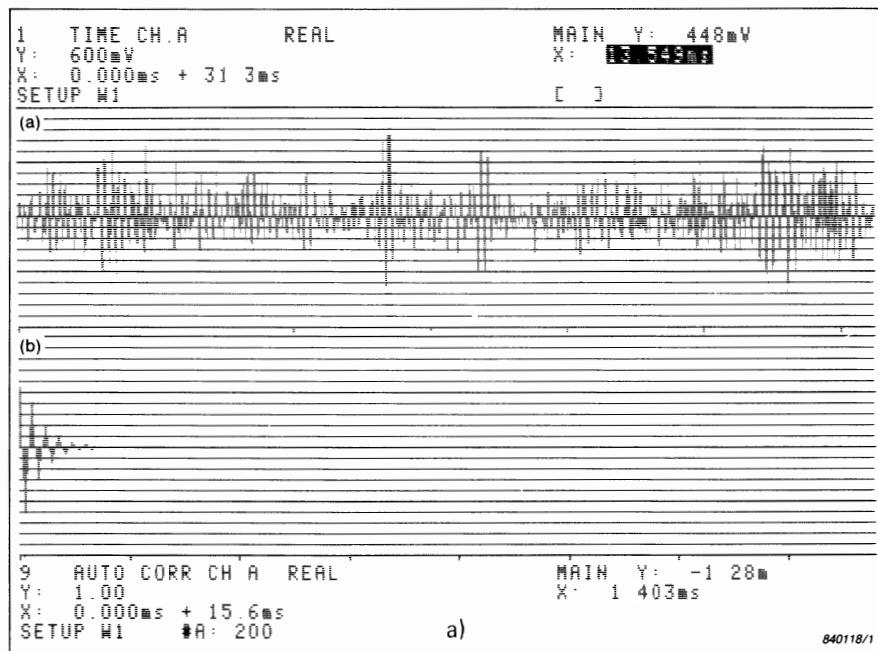


Fig. 7.25. (a) Time signal, and (b) Autocorrelation function for broadband random noise

The auto-correlation function for band-limited white noise, with (ideal) low-pass filter cutoff frequency $B/2$, is the well-known $\sin x/x$ function with zeroes at multiples of $1/B$. Thus for relatively broadband random signals, it is located close to the origin and dies away very rapidly. Fig.7.25 gives an example from Ref.7.1.

The auto-spectrum for Fig.7.24(c), band-pass filtered white noise, can be considered as the convolution of the spectra of cases (a) and (b). Thus, the auto-correlation function is given by the product of the autocorrelation functions (a) and (b). Consequently it is an amplitude modulated cosine, with envelope given by a $\sin x/x$ function.

The two main applications of the auto-correlation function are:

(1) Detection of a periodic signal buried in noise.

Fig.7.26 shows an example (from Ref.7.1) of a sine-wave buried in broadband random noise. It should be noted, however, that in order for the periodic structure to be apparent, the record length has to be so long (and thus bandwidth so narrow) that the discrete sinusoidal components would be readily observable above the noise level in the auto-spectrum as well. The use of the spectrum has the additional advantage that it can be used where more than one periodicity is present at the same time. This would give a "quasi-periodic" signal in the auto-correlation function, not easily recognizable by eye, whereas the use of a harmonic cursor in the auto-spectrum would allow rapid identification of the different sets of harmonics.

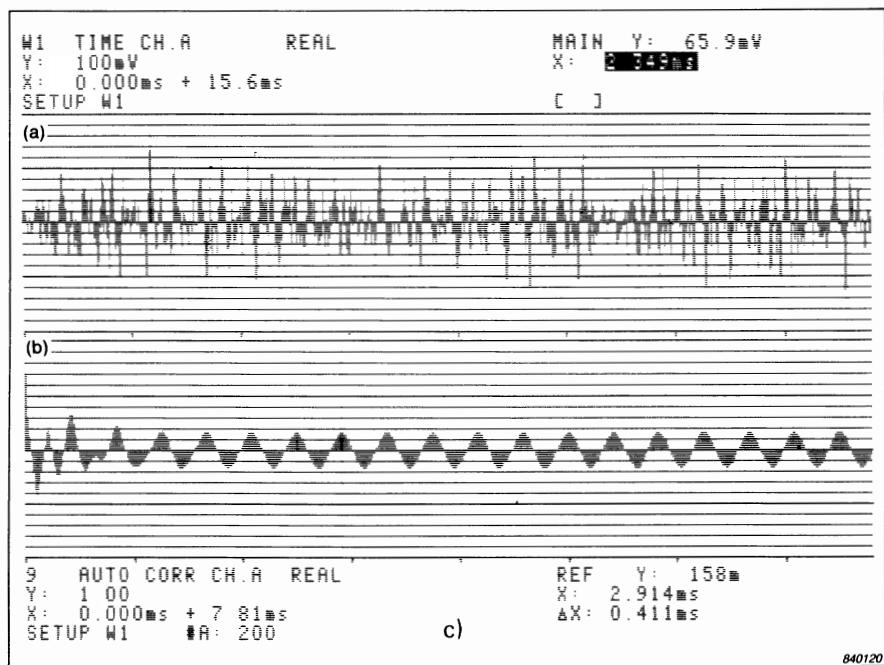


Fig. 7.26. (a) Time signal, and (b) Autocorrelation function for a sinusoid buried in broadband random noise

(2) Detection of echoes in a signal

When $a(t)$ contains echoes i.e. scaled down replicas of the main signal with different time delays, it can be seen from Eqn.(7.27) that when τ equals one of the echo delay times, the signal will correlate well with the delayed version of itself, and the auto-correlation will have a peak. The peak would however at best be a scaled down version of the autocorrelation function of the main signal, and would thus only be a narrow well-defined peak for broadband signals. It will be seen in Chapter 8 that the **cepstrum** in general has a much better defined peak corresponding to echoes, and is thus generally a better function to use to detect them.

7.4.2.2. Cross Correlation Function

Where signal $b(t)$ is simply a delayed version of $a(t)$, it will be appreciated that when τ equals the delay time τ_o the two signals will be identical. Thus the cross-correlation will simply be the autocorrelation of $a(t)$ displaced by τ_o . For a relatively broadband signal with a localised autocorrelation function, the location of the peak will indicate the delay time τ_o . However, Figures 7.24(b) and (c) show that where bandwidth B is narrow, the peak in the correlation function is so spread out that it is difficult to locate it exactly. Some benefit can be gained from taking the envelope of the analytic signal (Section 2.6), in order to avoid confusion from the local oscillations, but the width of the peak of the envelope function still limits the accuracy of determination of delay times.

Even where $b(t)$ is attenuated and/or contains additive noise, a local peak in the cross correlation function will indicate delay time and the degree of correlation between $a(t)$ and $b(t)$.

Thus, the major applications of the cross correlation function are to detect time delays between two signals, and to extract a common signal from noise.

(1) Determination of Time Delays

This is best illustrated by examples, which in this case are taken from Ref.7.1. Fig.7.27 gives an example where $a(t)$ is a sound signal measured close to a sound source, and $b(t)$ another at about 1,1 m distance from the source. The autospectrum G_{AA} (Fig.7.27(a)) shows the source signal to be very broadband, and thus in the cross correlation function $R_{ab}(\tau)$ (Fig.7.27(b)) the peaks are well-defined. The first and largest of these, selected by the cursor, is at 3,295 ms and thus corresponds to the direct path of 1,1 m. The other peaks with longer delay times represent reflected paths with more attenuation. Even in this fairly clear situation, the envelope (magnitude) of the analytic signal (Fig.7.27(c)) gives some improvement in detectability of the peaks.

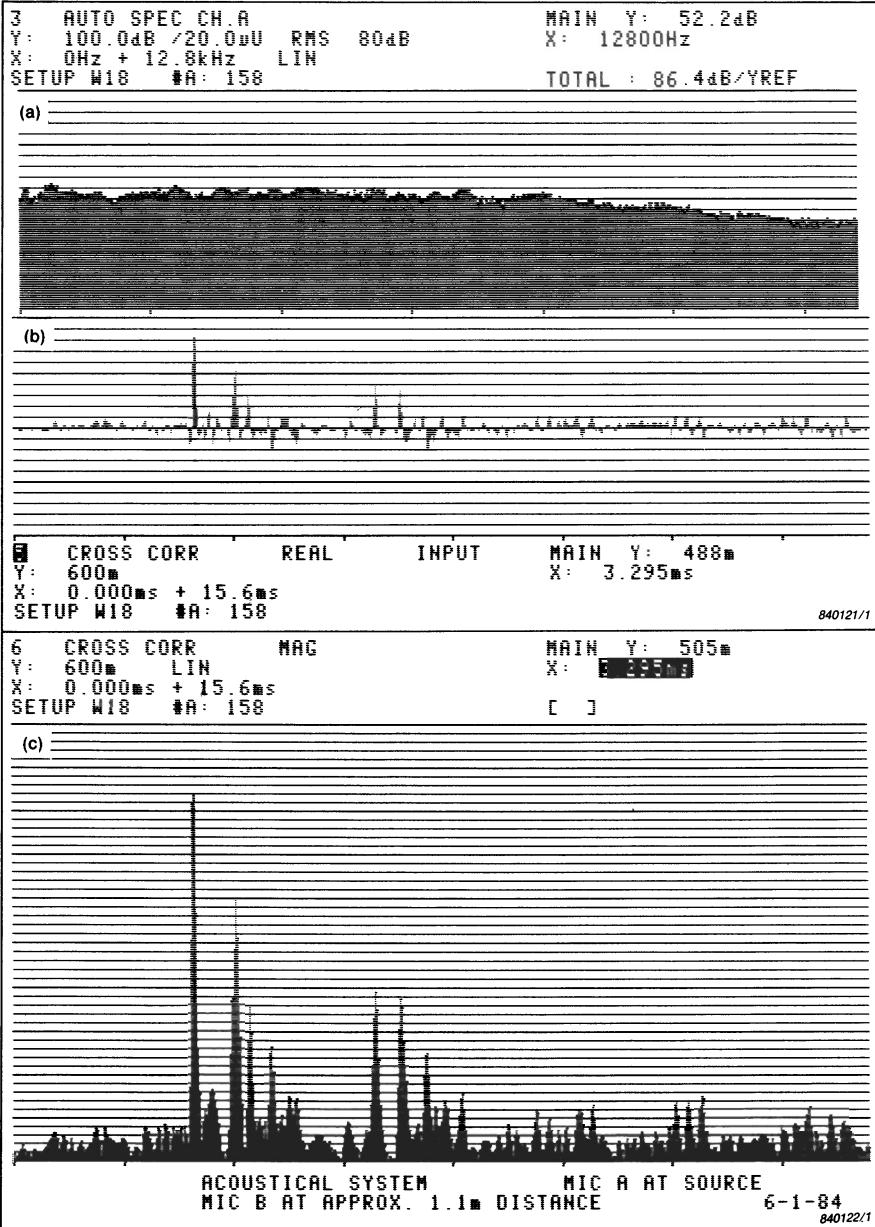


Fig. 7.27. Cross correlation with a broadband source

- (a) Autospectrum of input signal, $G_{AA}(f)$
- (b) Real part of the cross correlation function, $R_{ab}(\tau)$
- (c) Amplitude of the analytic signal, $|\hat{R}_{ab}(\tau)|$

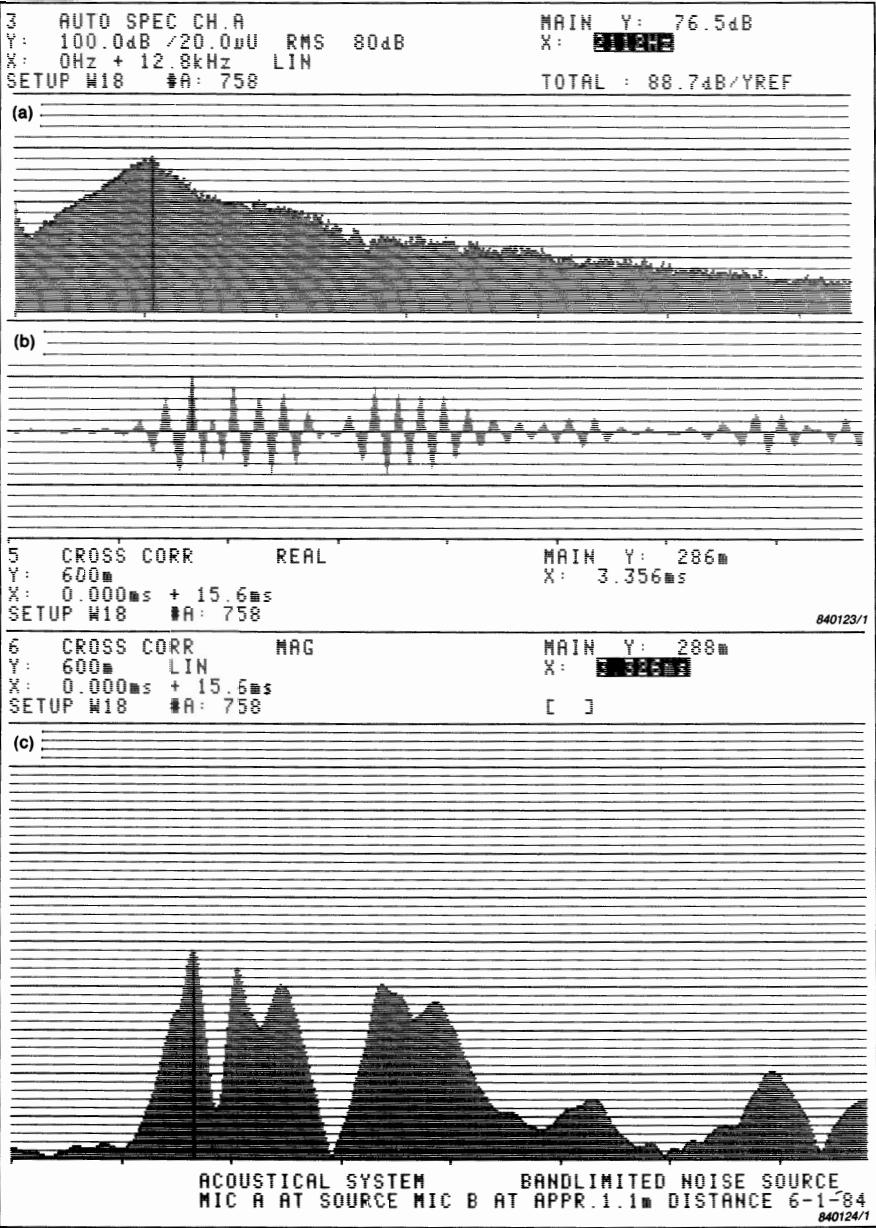


Fig. 7.28. Cross correlation with a band-limited source

- (a) *Autospectrum of input signal, $G_{AA}(f)$*
- (b) *Real part of the cross correlation function, $R_{ab}(\tau)$*
- (c) *Amplitude of the analytic signal, $|R_{ab}(\tau)|$*

With a constant wave speed, such as with sound signals, the delay times directly represent path length (and in the 2032/2034 analyzers the "FLEX" cursor can be calibrated in units of path length); thus, the cross correlation function can be used to determine the relative importance of the various paths by which a signal reaches a particular measurement point. However, as already mentioned, the results are very much influenced by the bandwidth of the signal. Fig.7.28 shows the results of a similar measurement to Fig.7.27, but where the source spectrum is band-limited (Fig.7.28(a)). It is now very difficult to determine the peaks corresponding to the various delay times in the real part of $R_{ab}(\tau)$ (Fig.7.28(b)) although the envelope of the analytic signal (Fig.7.28(c)) does improve the capability somewhat.

Even so, the information required here, the delays in the system response, is independent of the applied signal, and can better be determined using a measure of the system response itself, as will be seen in Section 7.5.

Another problem occurs in practice when the wave speed is not constant, for example in the case of bending waves in structures (and other dispersive systems). Here the wave velocity varies with frequency, as do delay times. One possible solution, already mentioned in Section 7.1.2, is to use the slope of the phase of the cross spectrum to indicate time delay. Another is to band-limit the cross-spectrum before inverse transformation, to limit the amount of dispersion. However, as we have seen in Fig.7.28, such band limitation gives problems in itself. The band limitation could be achieved by windowing in the cross spectrum, or possibly by using digital zoom (note that this will often be determined by the required time scale). Where real-time zoom is employed, even on a fairly broadband basis, it is virtually essential to use the magnitude of the analytic signal, as the real part is distorted by the zoom process. Fig.7.29 shows a result differing from Fig.7.27 only in that the 12.8 kHz band has been shifted upwards by 512 Hz by zooming. In this case the delay of the direct path (3,295 ms) is only correctly indicated in the magnitude of the analytic signal (Fig.7.29(b)).

(2) Extraction of a Common Signal from Noise

In the discussion of Section 7.1.1, it was pointed out that even in the presence of noise in the input and/or output signal, the correct cross spectrum can be measured if sufficient averages are made (Eqn.(7.3)). Correspondingly, the correct cross correlation function will be obtained from this cross spectrum. Hence, if the two signals $a(t)$ and $b(t)$ contain a common component, the cross correlation function will in principle show this independently of contaminating noise. If the common component were the signal $c(t)$ (with no attenuation or delay) the result would be the autocorrelation function of $c(t)$.

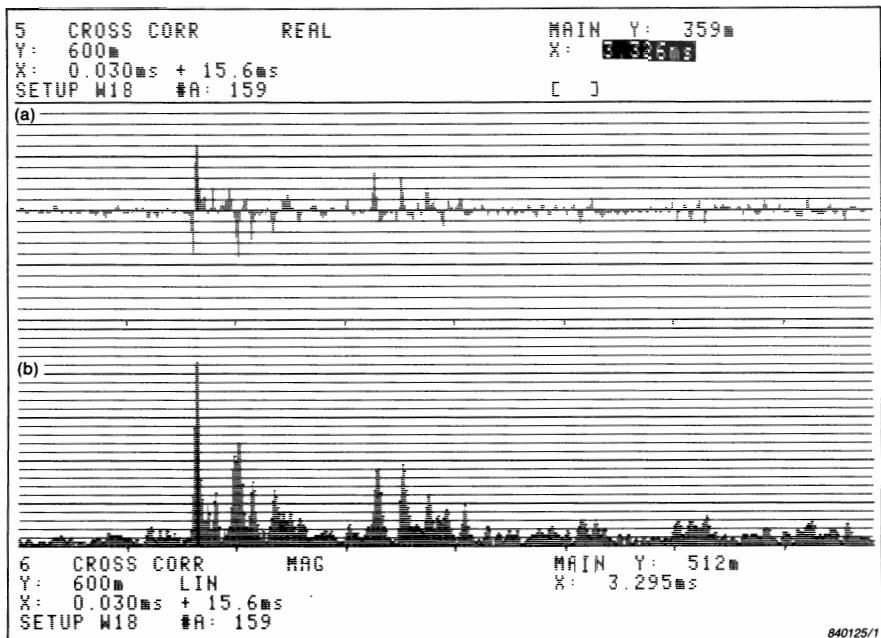


Fig. 7.29. Cross correlation using zoom. Same results as for Fig. 7.27, but using zoom in the frequency range 512 – 13312 Hz
 (a) Real part
 (b) Amplitude of the analytic signal

7.5. IMPULSE RESPONSE FUNCTIONS

7.5.1. Definitions and Calculation Procedures

As discussed in connection with Eqn.(2.28), the impulse response of a system is its output signal when a unit impulse (delta function) is applied at the input. It has also been pointed out that it is the inverse Fourier transform of the frequency response function, and this is the procedure used to calculate it in an FFT analyzer, thus:

$$h(t) = \mathcal{F}^{-1} \{ H(f) \} \quad (7.30)$$

Fig. 7.30 illustrates this for a small mechanical structure. Fig. 7.30(a) is a typical response to a hammer blow, which in this frequency range was very close to a true impulse (i.e. a flat spectrum). Fig. 7.30(b) shows the calculated impulse response obtained by inverse transforming the frequency response function, itself generated by averaging over a number of hammer blows (where the force

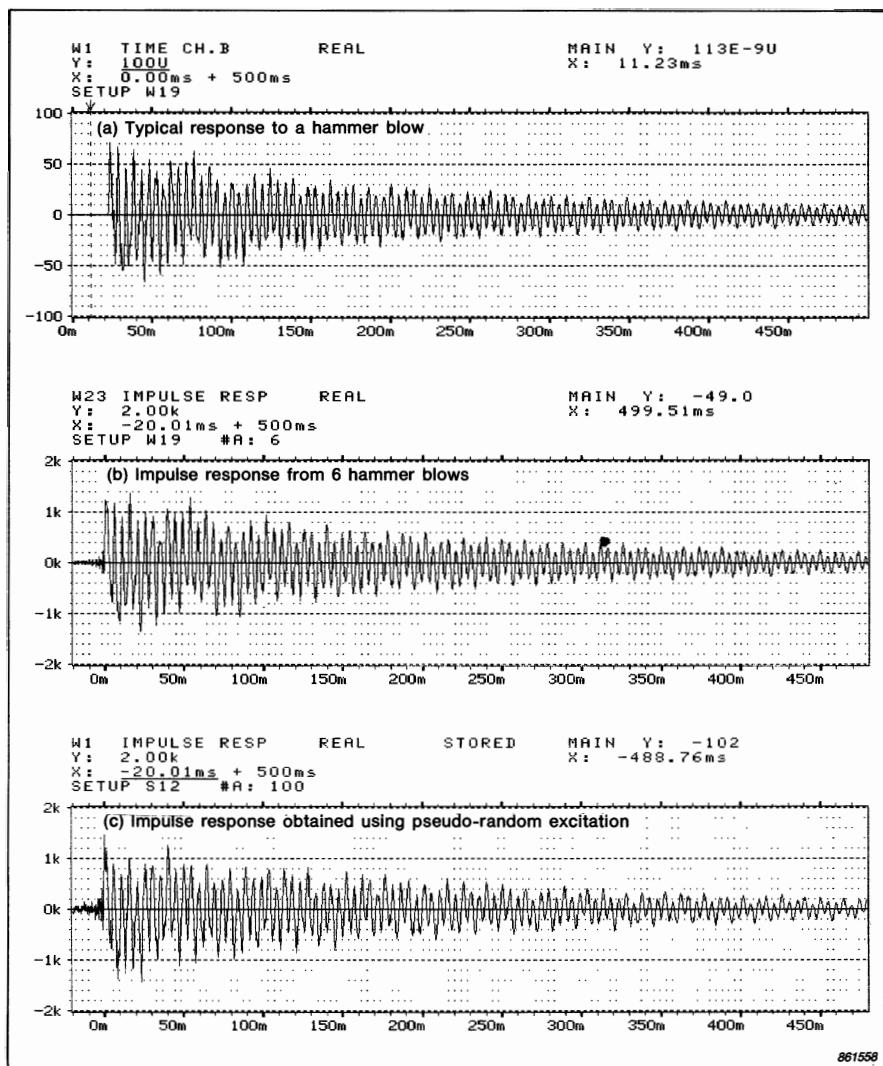


Fig. 7.30. Impulse response of a small structure

- (a) *Typical response signal to a hammer blow*
- (b) *Impulse response from the frequency response averaged over 6 hammer blows*
- (c) *Impulse response from a frequency response obtained using pseudo-random excitation*

applied by the hammer was measured by a force transducer). For comparison purposes, Fig.7.30(c) shows that the same result can be obtained from a frequency response function obtained in another way, in this case by applying a

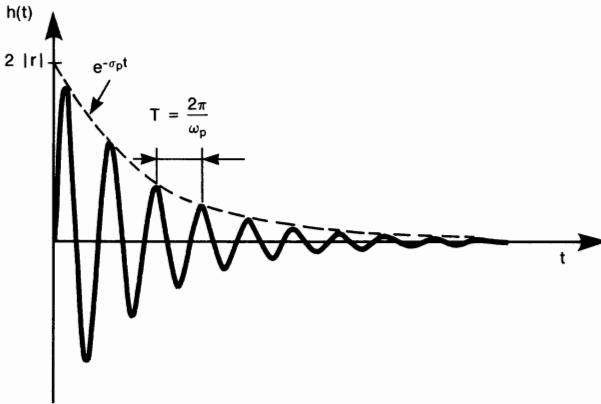
pseudo-random force signal via a shaker. It is seen that the results of (b) and (c) are virtually identical, and very similar to (a) which, however, is unscaled.

7.5.2. Properties and Applications

In the same way that the frequency response function $H(f)$ can be considered as the sum of the response of a number of single degree-of-freedom systems (Eqn.(7.21)) the linearity of the Fourier transform means that the impulse response is the sum of the individual single degree-of-freedom impulse responses. As shown in Appendix C, the impulse response of a single degree-of-freedom system has the general form of a "complex exponential" function, a damped one-sided sine-wave. As shown in Fig.7.31, this has the formula $2 | r | e^{-\sigma_p t} \sin(\omega_p t)$ in terms of the frequency response parameters of Eqns.(C.8) and (C.10). Summing over n modes (for an n degree-of-freedom system) the expression for a more general impulse response function is:

$$h_{ij}(t) = \sum_{k=1}^n 2 | r_{ijk} | e^{-\sigma_k t} \sin(\omega_k t) \quad (7.31)$$

One of the applications in modal analysis is effectively the application of Eqn.(7.31) to obtain the modal parameters of a multi degree-of-freedom system by curve fitting. The method is known as the "complex exponential" method (Ref.7.9). The impulse response function is obtained by inverse transforming a windowed section of the frequency response function containing the modes to be fit.



860773

Fig. 7.31. Impulse response of a single degree-of-freedom system in terms of the frequency response parameters r (residue), σ_p (damping factor) and ω_p (damped natural frequency)

The average decay rate of the impulse response can also be used to determine the average damping properties of a system. A typical parameter is the so-called "reverberation time", or time required for the response to decay by 60 dB. This is sometimes measured using the output of a short-term RMS detector applied to the decaying signal (initiated by removing a broad-band excitation, either random or impulsive). However, there is the danger that the averaging time used will itself limit the rate of decay, and there can be advantages in measuring directly the magnitude of the analytic signal corresponding to the impulse response. This has the advantage compared with the real part that it can be depicted on a logarithmic amplitude scale, and the mean decay rate determined by a straight line. Fig.7.32 illustrates this for the impulse response of a mechanical component.

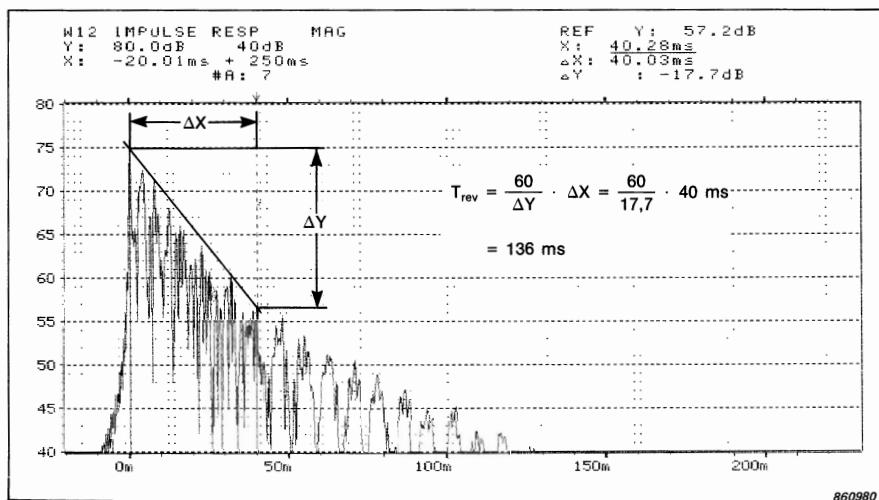


Fig. 7.32. Determination of "Reverberation Time" from the impulse response. The log magnitude of the impulse response is used

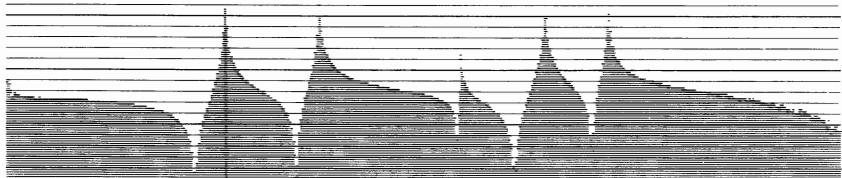
Fig. 7.33. Extraction of a single mode by windowing in the frequency response function

- (a) Total frequency response (magnitude)
- (b) Resonance peak separated out by windowing
- (c) Real part of the impulse response
- (d) Magnitude of the corresponding analytic signal on a log scale. The ΔX value gives the time constant τ for an 8,7 dB decay

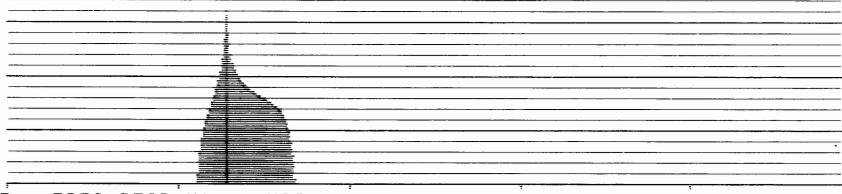
5 FREQ RESP H1 MAG
Y: 50.0dB 80dB
X: 0Hz + 3.2kHz LIN
SETUP W2 #R: 30

MAIN Y: 32 dB
X: 844Hz
ELEM #: 211

(a)



(b)



5 FREQ RESP H1 MAG
Y: 50.0dB 80dB
X: 0Hz + 3.2kHz LIN
SETUP W2 #R: 30

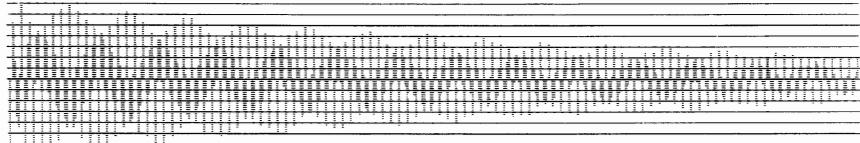
MAIN Y: 32 dB
X: 844Hz

840112/1

7 IMPULSE RESP REAL
Y: 1.50k
X: 0.00ms + 125ms
SETUP W2 #R: 30

MAIN Y: 313
X: 111.20ms

(c)



(d)

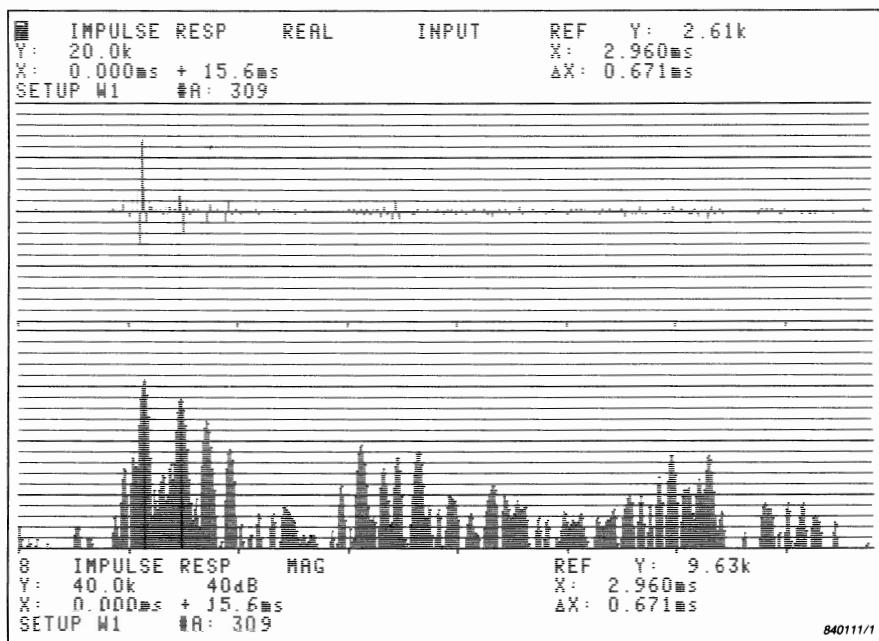


8 IMPULSE RESP MAG
Y: 80.0dB 40dB
X: 0.00ms + 125ms
SETUP W2 #R: 30

REF Y: 52.2dB
X: 111.20ms
ΔX: 103.63ms
ΔY: -8.74B

840114/1

For systems with well-separated modes, the individual resonances can be separated from each other by windowing in the frequency domain, after which an inverse transform gives the impulse response corresponding to that individual mode. Fig.7.33 shows an example from Ref.7.1, where in (a) the full frequency response is shown, in (b) the lowest resonance is windowed out using a frequency weighting function with a flat portion over most of the peak and a half-Hanning taper at either end. The impulse response from this mode alone is shown in (c), and (d) shows the corresponding envelope of the analytic signal on a logarithmic amplitude scale. This is a simple linear decay (cf. Fig.2.39), and shows at the same time how the modal damping information can be extracted. Considering the formula for the decay curve in Fig.7.31, $e^{-\sigma t}$, this can be expressed in terms of a "time constant" τ as $e^{-t/\tau}$, where $\tau = 1/\sigma$. τ is the time required for the amplitude to fall by a factor e (1 neper, or 8.7 dB). In Fig.7.33(d) the "REF" cursor is used to show that $\tau (= \Delta X) = 103.63$ ms and thus $\sigma = 1/\tau = 9.65$ rad/s. In Ref.7.1, the results are also given for the other resonances, and compared with results obtained directly in the frequency domain (from the half power bandwidth).



*Fig. 7.34. Impulse response function of an acoustical system.
 (a) Real Part
 (b) Log magnitude of the corresponding analytic signal*

As mentioned in Section 7.4.2.2, for systems with multiple delay transmission paths (and where there is a definite input and output signal) where one is interested in determining the delay times for each of the paths, the impulse response in general is better than the cross correlation function, as it is less dependent on the spectral shape of the input signal. Fig.7.34 (from Ref.7.1) shows a typical result both as the real part and log. amplitude of the analytic signal. In particular in the latter case, the peaks corresponding to the various delays are quite clear. By way of comparison, Fig.7.35 shows the corresponding cross correlation function (and cross spectrum). Because of the band-limited nature of the cross spectrum, the peaks in the cross correlation function are much less clear than in Fig.7.34.

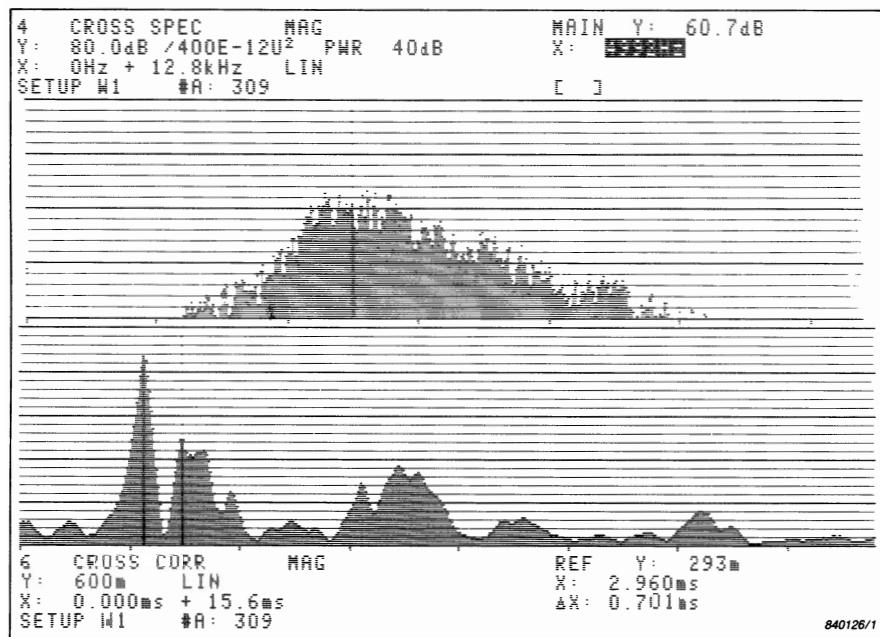


Fig. 7.35. Cross correlation for same case as Fig.7.34.

- (a) *Cross spectrum showing the band limitation*
- (b) *Linear magnitude of the cross correlation function*

8. CEPSTRUM ANALYSIS

Cepstrum analysis is the name given to a range of techniques all involving functions which can be considered as a “spectrum of a logarithmic spectrum”. In fact, the cepstrum was first defined (Ref.8.1) as far back as 1963 as the “power spectrum of the logarithmic power spectrum”. It was proposed at that time as a better alternative to the autocorrelation function for the detection of echoes in seismic signals. Presumably because it was a spectrum of a spectrum, the authors of Ref.8.1 coined the word **cepstrum** by paraphrasing **spectrum** and at the same time proposed a number of other terms derived in a similar manner. A list of the most common (sometimes encountered in the literature) is as follows:

Cepstrum from Spectrum

Quefrency from Frequency

Rahmonics from Harmonics

Lifter from Filter

Gamnitude from Magnitude

Saphe from Phase

and even such terms as “Short-pass lifter” from “Low-pass filter”. The first three of these terms are useful, and are used extensively in this chapter; even the word “lifter” is sometimes useful to indicate that filtration is carried out in the cepstrum domain. The analyzers Types 2032/2034 display the “Liftered Spectrum” obtained from applying a “Long-pass” or “Short-pass” lifter in the cepstrum.

However, the distinctive feature of the cepstrum is not that it is a spectrum of a spectrum, but rather the logarithmic conversion of the original spectrum. Note that the auto-correlation function is the inverse Fourier transform of the power spectrum (Eqn.7.29) and can thus also be considered a “spectrum of a

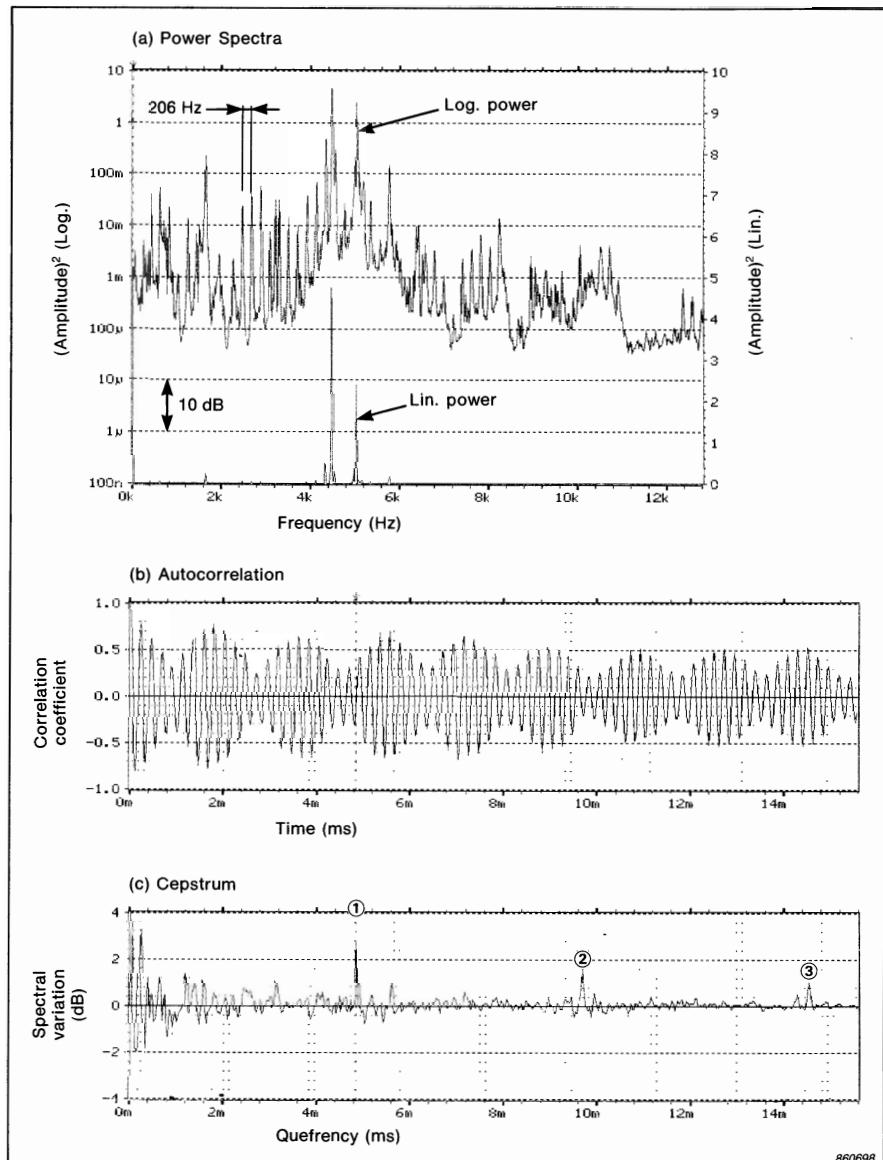


Fig. 8.1. Effect of linear vs. logarithmic amplitude scale in power spectrum
 (a) Power spectrum on linear scale (lower curve) and logarithmic scale (upper curve)
 (b) Autocorrelation function (obtained from linear representation)
 (c) Cepstrum (obtained from logarithmic representation) — ①, ② etc. are harmonics corresponding to harmonic series in spectrum ($4.85\text{ ms} = 1/206\text{ Hz}$). The harmonics result from a fault in a bearing

spectrum". In fact the most commonly used definition of the cepstrum nowadays is as the "inverse Fourier transform of the logarithmic power spectrum" which differs from the auto-correlation only by virtue of the logarithmic conversion of the spectrum.

Figure 8.1 illustrates one of the situations where the cepstrum gives advantages over the auto-correlation function. Fig.8.1(a) shows a particular power spectrum on both linear and logarithmic amplitude scales (the linear scale of course has units of "amplitude squared"). Fig.8.1(b) and (c) show the inverse Fourier transforms, the auto-correlation function and cepstrum, respectively. In the logarithmic version of Fig.8.1(a), a family of harmonics can be seen (which derive from a fault in a ball bearing). The effect of this harmonic family is seen clearly in the cepstrum (Fig.8.1(c)) as a series of rahmonics (denoted ①, ②, etc.) but is not at all evident in the autocorrelation function. The latter is dominated by the effects of the two largest peaks in the spectrum, which are all that show up on the linear amplitude squared scale.

Cepstra derived from power spectra are now known as "power cepstra". Another type of cepstrum which was defined later is the so-called "complex cepstrum" (Ref.8.2), derived from a complex spectrum, and thus using phase as well as log amplitude information at each frequency. For this reason the complex cepstrum operation is reversible back to a time signal; by contrast the new definition of the power cepstrum is reversible back to a power spectrum, while the old definition is not reversible at all.

Cepstrum analysis, in particular that involving editing in the complex cepstrum, is one example of so-called "homomorphic" signal processing. For further details of this wider topic, see Ref.8.2.

8.1. DEFINITIONS AND CALCULATION PROCEDURES

The original definition of the power cepstrum may be expressed as follows:

$$C_{AA}(\tau) = |\mathcal{F}\{\log S_{AA}(f)\}|^2 \quad (8.1)$$

in which the (2-sided) power spectrum, $S_{AA}(f)$, of a time signal $a(t)$ is given by:

$$S_{AA}(f) = \overline{|\mathcal{F}\{a(t)\}|^2} \quad (8.2)$$

where the bar means averaging over a number of records (where applicable).

The new definition of the power cepstrum is:

$$C_{AA}(\tau) = \mathcal{F}^{-1}\{\log S_{AA}(f)\} \quad (8.3)$$

As will be seen, there can be advantages in using the corresponding analytic signal (see Section 2.6.1) which can be obtained from the (one-sided) log power spectrum $\Gamma_{AA}(f)$ as follows:

$$\overset{\vee}{C}_{AA}(\tau) = \mathcal{F}^{-1}\{\Gamma_{AA}(f)\} \quad (8.4)$$

where $\begin{aligned} \Gamma_{AA}(f) &= 2 \log S_{AA}(f) & , f > 0 \\ \Gamma_{AA}(f) &= \log S_{AA}(f) & , f = 0 \\ \Gamma_{AA}(f) &= 0 & , f < 0 \end{aligned}$

This is the definition used in the analyzers Types 2032/2034, and its magnitude $|\overset{\vee}{C}_{AA}(\tau)|$ the version used in the analyzer Type 2515. Note that the real part of $\overset{\vee}{C}_{AA}(\tau)$ is the same as $C_{AA}(\tau)$ defined in Eqn.(8.3).

In comparing the definitions (8.1) and (8.3) the question of whether a forward or inverse transform is used is largely a formality. Because the power spectrum $S_{AA}(f)$ is a real even function, the cepstrum is also a real even function (Table 2.1), and the forward and inverse transforms (Eqn.(2.14) and (2.15)) give the same result (using the DFT there is a difference in scaling factor). The real difference is the squaring of the results in Eqn.(8.1) thus making the process irreversible and emphasizing the largest values. This is often not an advantage because the largest values usually occur at low quefrequencies, and are often less important than peaks at higher quefrequencies. The use of an inverse transform in the definition makes the relationship of the cepstrum to the auto-correlation function clearer, and is also more natural in going from a function of frequency to a function of time (Eqn.(2.36)). Thus, the parameter τ in the definitions is actually time, although it is referred to as "quefrency". As for the auto-correlation function, the time parameter τ can better be thought of as "delay time" or "periodic time" rather than absolute time.

The complex cepstrum may be defined as follows:

$$C_A(\tau) = \mathcal{F}^{-1}\{\log A(f)\} \quad (8.5)$$

where $A(f)$ is the complex spectrum of $a(t)$, i.e.:

$$A(f) = \mathcal{F}\{a(t)\} = A_R(f) + jA_I(f)$$

in terms of its real and imaginary parts (Eqn.(2.40)) or,

$$A(f) = |A(f)| e^{j\phi(f)} \quad (8.6)$$

in terms of its amplitude and phase at each frequency. Taking the (complex) logarithm of Eqn.(8.6) gives:

$$\log A(f) = \ln |A(f)| + j\phi(f) \quad (8.7)$$

and it is this complex function of frequency, with log amplitude as real part, and phase as imaginary part, which is inverse transformed in Eqn.(8.5) to give the complex cepstrum.

Note that when $a(t)$ is real-valued, as is normally the case, then $A(f)$ is conjugate even (Eqn.(2.37)) from which the following relationships follow:

$A_R(f)$ is even

$A_I(f)$ is odd

$|A(f)|$ is even

$\ln |A(f)|$ is even

$\phi(f)$ is odd

from which it further follows that $\log A(f)$ is conjugate even and the complex cepstrum $C_A(\tau)$ is a real-valued function, despite its name (Table 2.1).

8.2. PROPERTIES OF THE CEPSTRUM

As typified by the results of Fig.8.1, the cepstrum has the ability to detect periodic structures in the logarithmic spectrum, for example families of harmonics and/or sidebands with uniform spacing. Another effect which gives a periodic structure to the logarithmic spectrum is the presence of echoes, which can be understood as follows.

In Fig.8.2, a signal with an echo is modelled as the result of a convolution of the original signal with a function comprising two delta functions, a unit impulse at the origin, and an attenuated impulse at time τ (corresponding to the echo attenuation and delay time). By the convolution theorem, the spectrum of the total signal in (c) must be the product of the spectrum of the original function in (a) and the spectrum of (b). The latter can be derived intuitively by considering the analogous case of a (complex) time signal corresponding to a spectrum of the form of (b). As shown in (d) this would have a fixed (DC) component of unit length, and an additive rotating phasor rotating at a rate proportional to τ (and thus with period $1/\tau$ in the other domain). The resulting (complex) function has an amplitude varying periodically around a mean of unity, and a phase varying periodically around zero. Multiplying by this function would have the effect of adding a periodic function to the log amplitude spectrum, and another periodic function (with the same period) to the phase. Figure 8.3 illustrates this for the case of a structural response when the structure is excited by hammer blows.

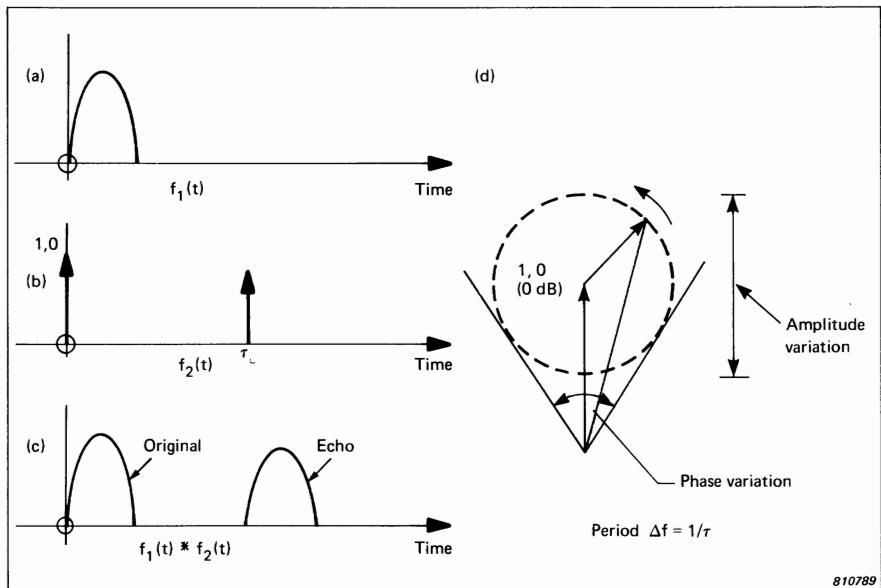


Fig. 8.2. Modelling a signal with echo as a convolution

810789

Where an echo is present (as a result of a "double hit"), the log amplitude and phase spectra both have an additive periodic structure with a period in frequency of 18.8 Hz, the reciprocal of the echo delay time, 53 ms. The power cepstra are shown in Fig.8.4, illustrating that the echo gives a peak at a quefrency corresponding to the echo delay time (plus some minor rahmonics). As a matter of interest, Fig.8.4 also shows the corresponding autocorrelation functions, which (as mentioned in Section 7.4.2.1) are not as efficient in revealing the echo.

Another property of the cepstrum, which gives rise to a large number of applications, is its ability to separate source and transmission path effects, i.e. to effect a "deconvolution". This can be understood as follows.

Referring to Fig.7.10, the relationships between the input and output signals of an ideal system can be expressed as:

$$b(t) = a(t) * h(t) \quad (8.8)$$

or in the frequency domain

$$B(f) = A(f) \cdot H(f) \quad (8.9)$$

in terms of the complex spectra, or

$$S_{BB}(f) = S_{AA}(f) \cdot |H(f)|^2 \quad (8.10)$$

in terms of the power spectra.

Taking the logarithm of Eqn.(8.9) gives

$$\log B(f) = \log A(f) + \log H(f) \quad (8.11)$$

and because of the linearity of the (inverse) Fourier transform, the additive relationship is maintained in the cepstrum, i.e..

$$\mathcal{F}^{-1} \{ \log B(f) \} = \mathcal{F}^{-1} \{ \log A(f) \} + \mathcal{F}^{-1} \{ \log H(f) \} \quad (8.12)$$

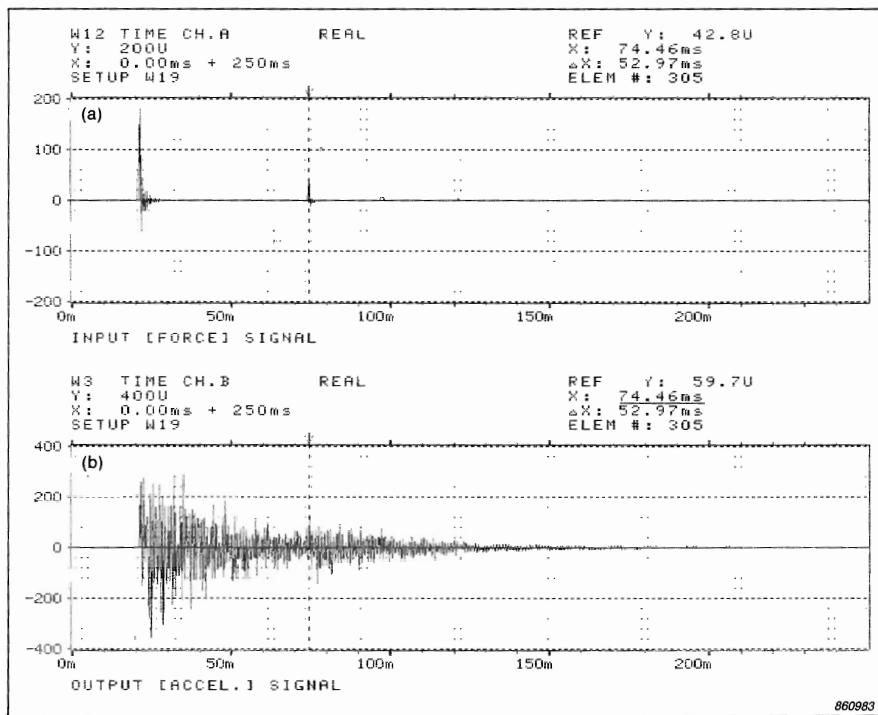


Fig. 8.3. Effect of echoes on the spectrum

- (a) Impulsive force signal with a dominant echo (double hit) at 53ms
- (b) Resulting structural response signal. The echo is not immediately obvious

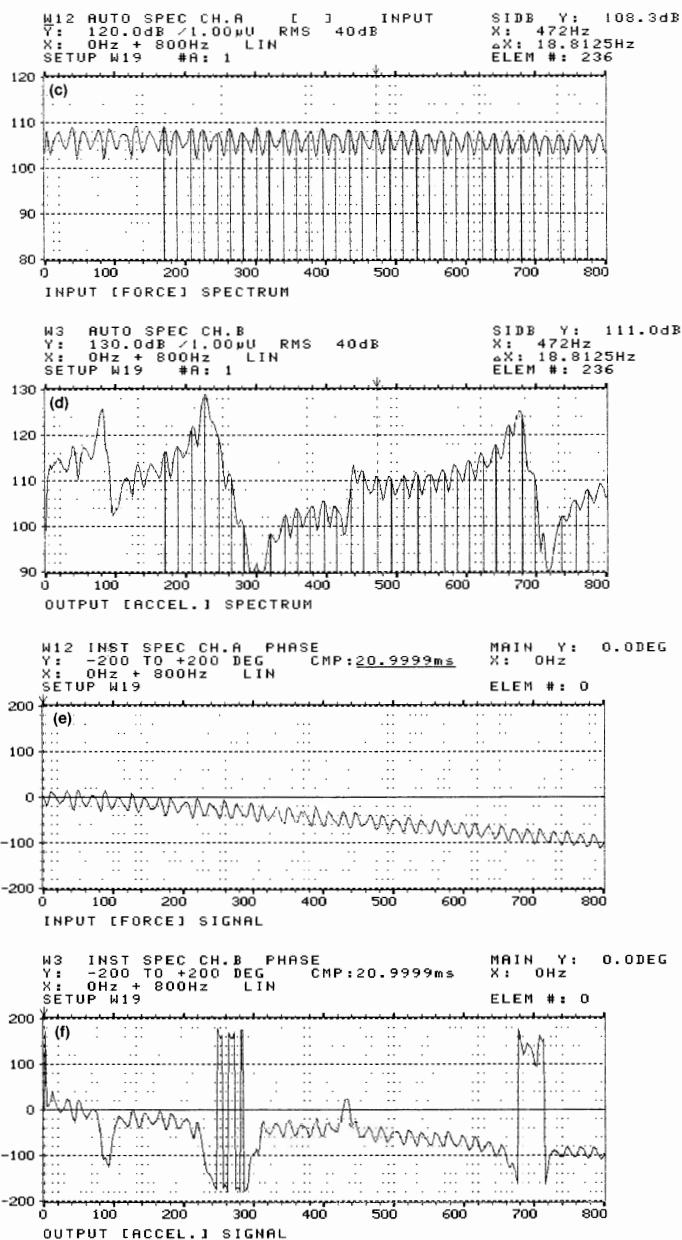


Fig. 8.3.cont.

860978

Exactly the same process can be applied to Eqn.(8.10), meaning that the source and transmission path effects are also additive in the power cepstrum. Note that by direct inverse transformation of Eqn.(8.10), the multiplication would result in a convolution of the corresponding effects in the autocorrelation functions (treating $|H(f)|^2$ as a power spectrum).

Applications of the cepstrum, based on these properties, range from simple diagnostic applications (e.g. determination of an echo delay time) to those involving editing in the cepstrum in order to remove an effect in another domain (e.g. echo removal, deconvolution).

- Fig. 8.3. (c) Log amplitude spectrum of force signal. The periodicity with cont. 18.8 Hz spacing corresponds to the 53ms echo delay time
(d) Log amplitude spectrum of the response signal. This has the same added periodicity
(e) Phase spectrum of the force signal. This has an added periodicity corresponding to the echo delay time
(f) Phase spectrum of the response signal. This has the same added periodicity

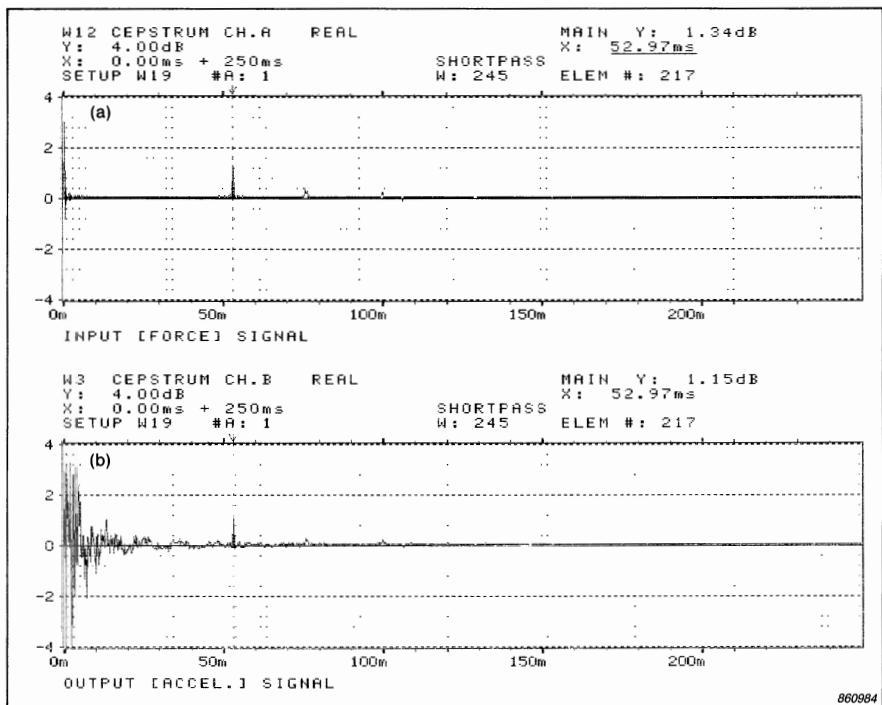


Fig. 8.4. Cepstra and autocorrelations for the same signals as Fig.8.3.

(a) & (b) Cepstra for the force and response signals, respectively. The echo delay time of 53ms is clearly evident in both

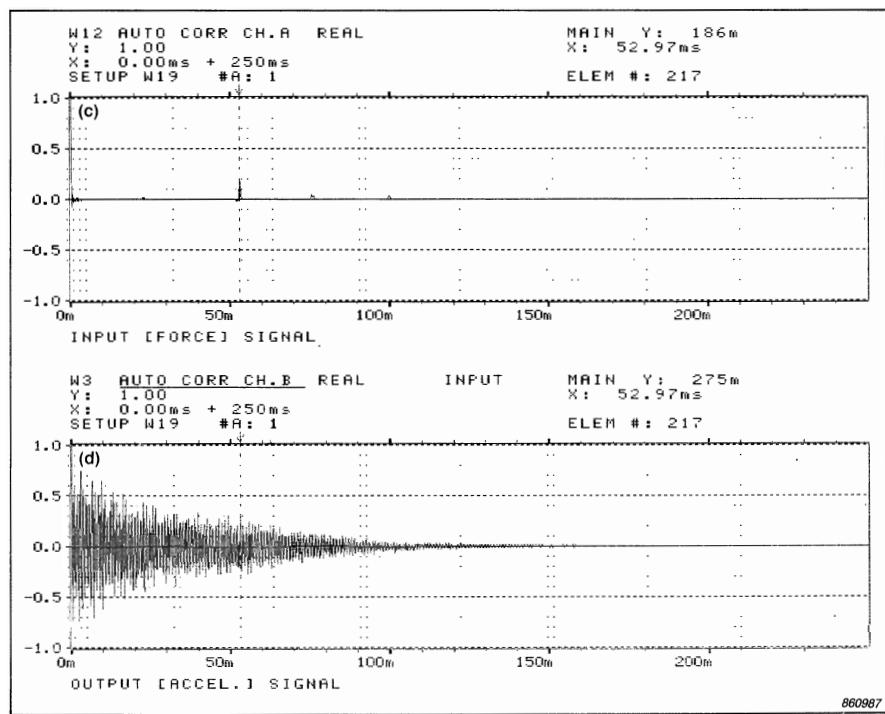


Fig. 8.4. (c) & (d) Autocorrelation functions for the force and response signals, cont. respectively. The echo is only apparent in the force signal, which has a flat spectrum

8.3. APPLICATIONS OF THE POWER CEPSTRUM

8.3.1. Echo Detection and Removal

Figure 8.5 (from Ref.8.3) shows the results of an experiment to remove the effects of an echo from the power spectrum, by editing in the power cepstrum.

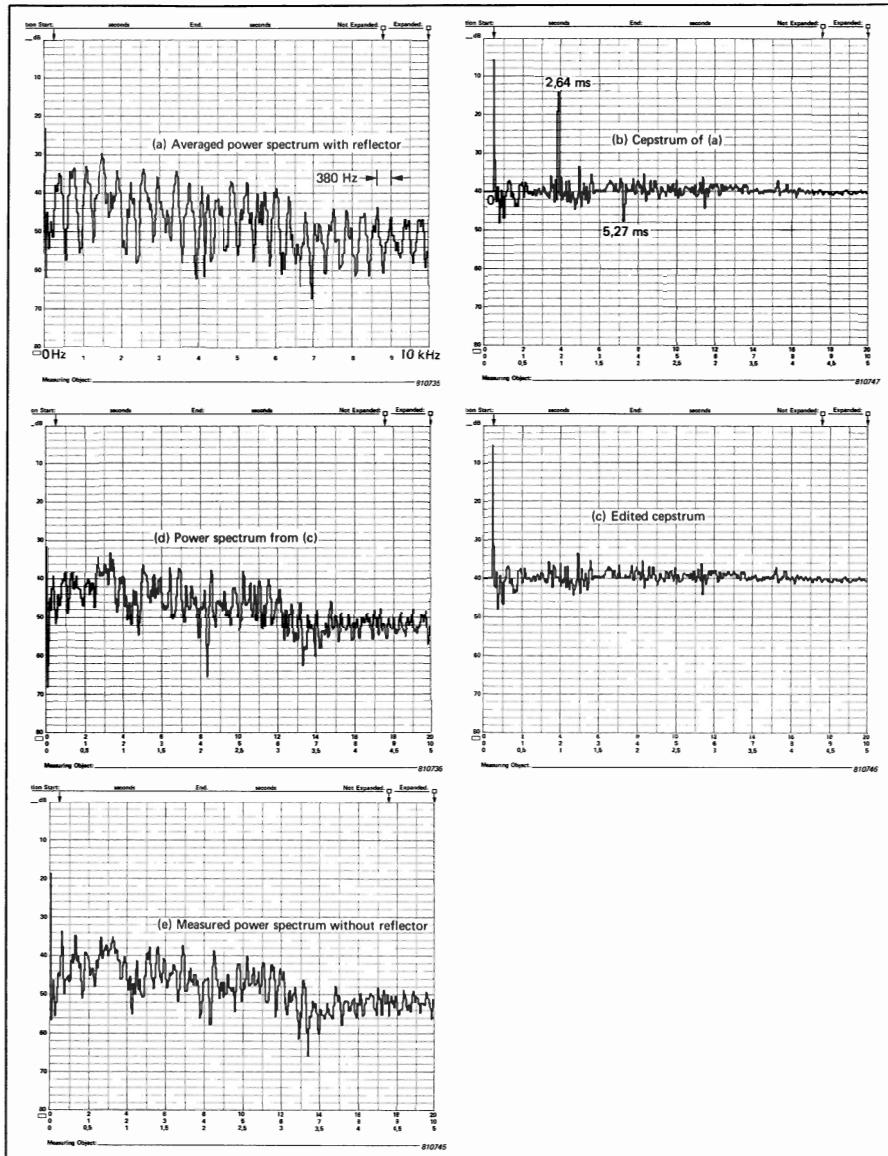


Fig. 8.5. Echo removal using power cepstrum

The spectrum of Fig.8.5(a) was obtained from a measurement with a reflecting board 43 cm behind a microphone receiving a broadband sound signal from a loudspeaker in a normal room. The peaks in (b) at multiples of the corresponding delay time 2.64 ms have been edited away in (c), and the spectrum of (d) obtained by a further forward transform of the edited cepstrum. This is seen to correspond very well with a direct measurement without the reflecting plane (Fig.8.5(e)).

A very similar technique has been used in Refs. 8.4 and 8.5 to remove the effects of ground reflections from measurements of aero engine noise on the ground. A result of this kind was given in Ref.8.3, which also lists a number of other references describing applications of the cepstrum in seismology and underwater acoustics. Some of these describe the use of the complex cepstrum for removing echoes from time signals, such as described here in Section 8.6.1.

8.3.2. Properties of a reflecting surface

In the analysis so far, the echoes and reflections have been considered to be perfect copies of the original signal, only attenuated in amplitude. Where the reflection is not perfect, the cepstrum can be used to determine the properties of the reflecting surface. The following derivation is largely taken from Ref.8.6.

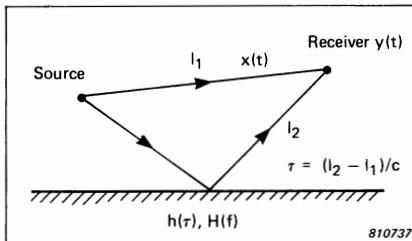


Fig. 8.6. Signal with a non-ideal reflection

Consider the case depicted in Fig.8.6, where the signal $y(t)$ received at a microphone is the sum of the direct signal $x(t)$, and a reflected signal modified by the reflecting surface and attenuated because of the longer path. Thus, in the time domain:

$$y(t) = x(t) + \frac{l_1}{l_2} x(t) * h(t-\tau) \quad (8.13)$$

Transforming this equation by the Fourier transform gives

$$Y(f) = X(f) \left[1 + \frac{l_1}{l_2} H(f) e^{-j2\pi f\tau} \right] \quad (8.14)$$

The power spectrum can be obtained as the square of the modulus:

$$|Y(f)|^2 = |X(f)|^2 \left[1 + \frac{l_1}{l_2} H e^{-j2\pi f\tau} \right] \left[1 + \frac{l_1}{l_2} H e^{j2\pi f\tau} \right]^* \quad (8.15)$$

from which the log power spectrum is:

$$\begin{aligned} \log |Y|^2 &= \log |X|^2 + \log \left[1 + \frac{l_1}{l_2} H e^{-j2\pi f\tau} \right] \\ &\quad + \log \left[1 + \frac{l_1}{l_2} H^* e^{+j2\pi f\tau} \right] \end{aligned} \quad (8.16)$$

Expanding the $\log(1+x)$ terms as

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} \dots$$

and inverse transforming to the power cepstrum gives

$$\begin{aligned} C_{YY}(t) &= C_{XX}(t) + \frac{l_1}{l_2} h(t-\tau) - \left(\frac{l_1}{l_2} \right)^2 \cdot \frac{1}{2} h(t-\tau) * h(t-\tau) + \dots \\ &\quad + \frac{l_1}{l_2} h(-t+\tau) - \left(\frac{l_1}{l_2} \right)^2 \cdot \frac{1}{2} h(-t+\tau) * h(-t+\tau) + \dots \end{aligned} \quad (8.17)$$

This means that the impulse response of the reflection, $h(t)$, will be found in the power cepstrum delayed by the echo time τ and scaled down by the attenuation factor $\frac{l_1}{l_2}$. At the higher harmonic quefrequencies of the delay time the impulse response is convolved with itself progressively more (and scaled down progressively more). At negative quefrequencies the mirror image is found, because the power cepstrum is an even function. Thus, provided both the cepstrum of the original signal, and the impulse response of the reflection are shorter than the delay time τ , it should be possible to extract the impulse response by simple windowing in the cepstrum. It is interesting that this impulse response may be Fourier transformed to give both the amplitude and phase characteristics of the reflection, even though the power cepstrum has been used. Ref.8.6 discusses a number of the practical points involved in making such measurements.

8.3.3. Speech analysis

One of the earliest applications of the cepstrum was to the detection of voiced speech and determination of voice pitch (Ref.8.7). This is because voiced speech distinguishes itself from unvoiced speech in that it has a harmonic structure, with a large number of harmonics of the fundamental voice pitch (see Fig.8.7).

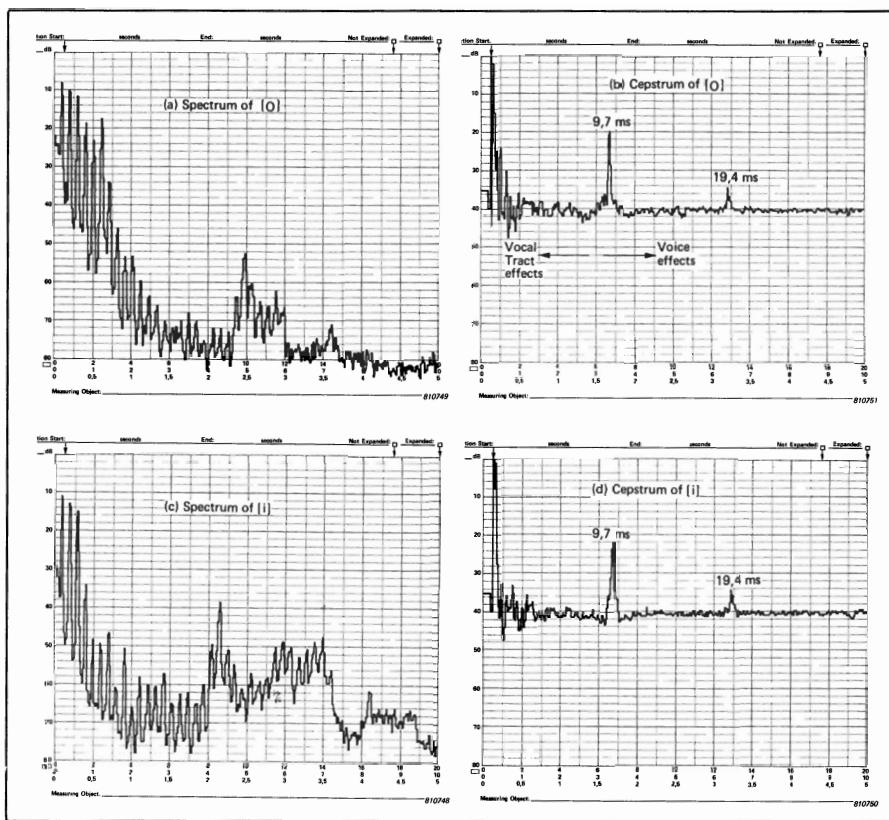


Fig. 8.7. Spectra and cepstra for two vowels

Other applications of the cepstrum are based on its ability to separate source and transmission path effects, provided they have different quefrency contents. This is usually the case with speech where the source spectrum (e.g. of the voice) is very flat, containing a large number of harmonics of the voice pitch, but is modified by the resonance characteristics of the vocal tract, the so-called "formants", which determine for example which vowel is being uttered. Fig.8.7 shows spectra and cepstra for the vowels "oh" [o] and "ee" [i] and illustrates how the differences mainly lie in the low quefrency part of the cepstrum, which is dominated by the formant characteristic. Non-voiced sounds, such as many consonants and whispered speech, do not give peaks in the cepstrum corresponding to the voice pitch, and this can be used to separate voiced from unvoiced speech (Refs.8.7, 8.8).

It is also possible by editing in the cepstrum to remove one effect completely, for example the voice, and thus simplify the tracking of the formants. Fig.8.9 (from Ref.8.9) shows a typical situation, a 3-dimensional representation of the section "ea" from the word "Montreal". The picture is quite confused, but by "shortpass filtering" each of the spectra to remove the voice components, as shown in Figs.8.9 and 8.10, only the formants are left and the picture becomes much clearer.

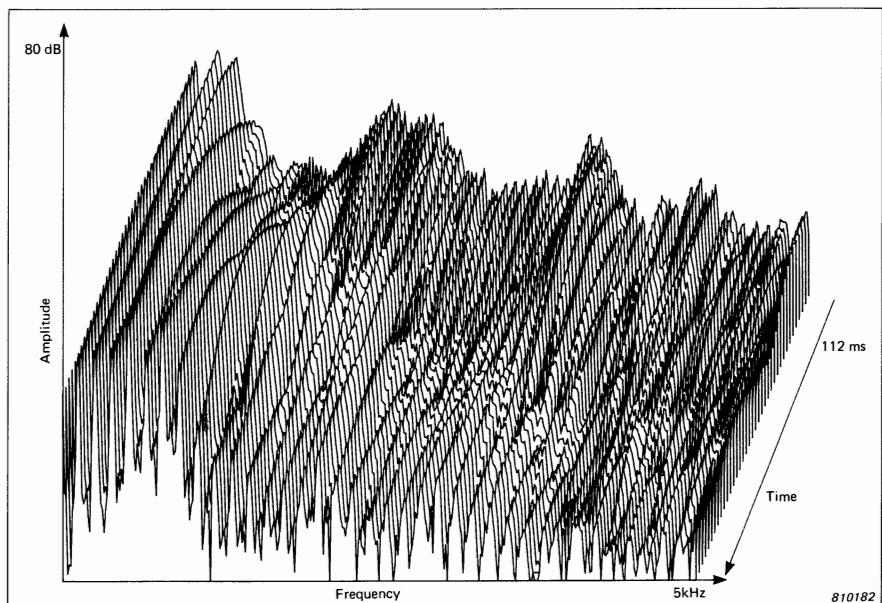


Fig. 8.8. Scan spectrum of "ea" in "Montreal"

As another application of the power cepstrum, Ref.8.10 shows how it can be useful to include it, along with spectral and other information, in pattern recognition algorithms for speaker identification. Inclusion of the cepstral information improved the ability of the technique to exclude impostors.

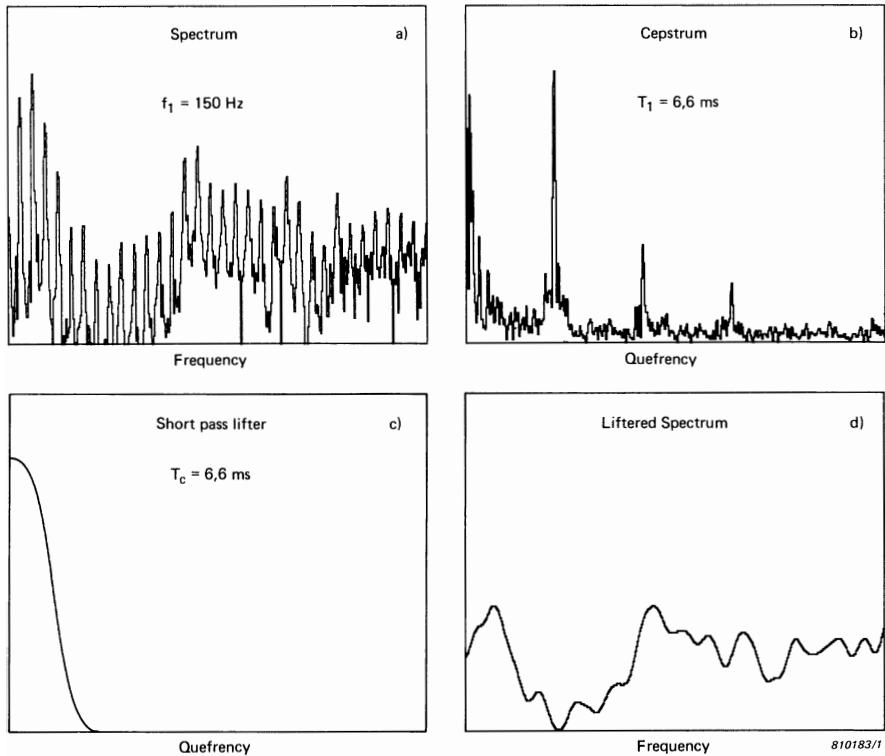


Fig. 8.9. Cepstrum filtering

- a) *log power spectrum of vowel*
- b) *magnitude of cepstrum*
- c) *short pass lifter characteristic*
- d) *short pass filtered log power spectrum*

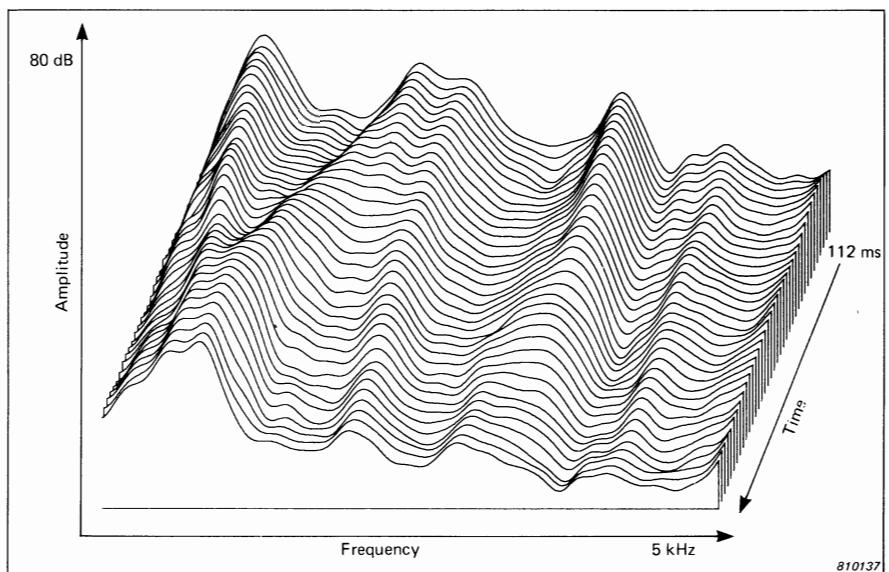


Fig. 8.10. Short pass filtered scan spectrum of "ea" in "Montreal"

8.3.4. Machine Diagnostics

The applications of the power cepstrum to machine diagnostics are based mainly on its ability to detect periodicity in the spectrum e.g. families of uniformly spaced harmonics and sidebands, while being insensitive to the transmission path of the signal from an internal source to an external measurement point.

Extensive families of harmonics are produced whenever short impulses are repeated periodically, for example from a localised fault in a ball-bearing. Fig.8.1 gives one such example. Note from the discussion of Section 4.6.1 (Fig.4.22) that clear patterns of harmonics are often only produced for the lower order harmonics; cepstrum analysis will not be applicable at very high frequencies where the harmonics merge.

Ref.8.11 discusses the case from which the signal of Fig.8.1 was taken. It is shown there that the use of the cepstrum gives three main benefits:

- (1) Detection of the harmonic pattern, indicating the fault, at an early stage, even though the fundamental component (the "ball-pass frequency") was not detectable initially.

- (2) An accurate indication of the harmonic spacing, showing it to be 4,1 times the shaft speed, which corresponded to the calculated frequency for an outer-race fault in a particular bearing.
- (3) The value of the main cepstrum peak was shown to be an excellent trend parameter; as it represents the average over a large number of individual harmonics, fluctuations in the latter (for example as a result of load variations) were largely averaged out in the cepstrum value, which gave a smooth trend curve with time.

Ref.8.12 describes a somewhat similar application, the detection of damaged blades in a turbine. The resulting flow anomaly gives an excitation pulse once per revolution, and a corresponding increase in the shaft harmonics in the mid frequency range, which in an actual case gave a dramatic increase in the corresponding cepstrum component.

Sidebands result from modulation of a carrier frequency by one or more lower frequencies, and can give information about machine faults in some cases. In the case of rolling element bearings, the presence of sidebands can give additional information as to the faulty component, as the fault signal tends to be unmodulated for outer race faults, modulated by the shaft speed for inner race faults, and by the cage speed for rolling element faults. In gear vibrations, modulation of the otherwise uniform toothmeshing component gives information about the **differences** between the teeth on a particular gear, with sidebands having a spacing corresponding to the rotational speed of that gear.

Fig.8.11 gives an example (from Ref.8.13) of the application of cepstrum analysis to gear diagnostics.

Fig.8.11(a) shows the spectrum and cepstrum for a speed-up gearbox between an electric motor running at 50 Hz and a centrifugal compressor running at 121 Hz. Components are seen in the cepstrum at 20ms (the time for one rotation of the 50 Hz gear) and at 8,2ms (corresponding to the 121 Hz gear). The second and third order "rahmonics" (harmonics in the cepstrum) of the latter are also apparent, but when falling off monotonically, as in this case, they can usually be disregarded; the first rahmonic contains the essential information. Fig.8.11(b) is a similar comparison where the spectrum has been edited to eliminate the low harmonics of both gear speeds (up to approx. half the toothmesh frequency). It is now evident from the cepstrum that only the 50 Hz gear gives significant sidebands around the toothmesh frequency; the cepstral components in Fig.8.11(a) from the 121 Hz shaft are apparently due entirely to low harmonics having nothing to do with the gear mesh. Fig.8.11(c) shows a similar comparison some months later, and it is now evident that even with the edited spectrum, some modulation by the 121 Hz gear is occurring. This was due in fact to a misalignment of this shaft which had developed in the meantime. Fig.8.11(d) shows the result of editing in the cepstrum to remove the

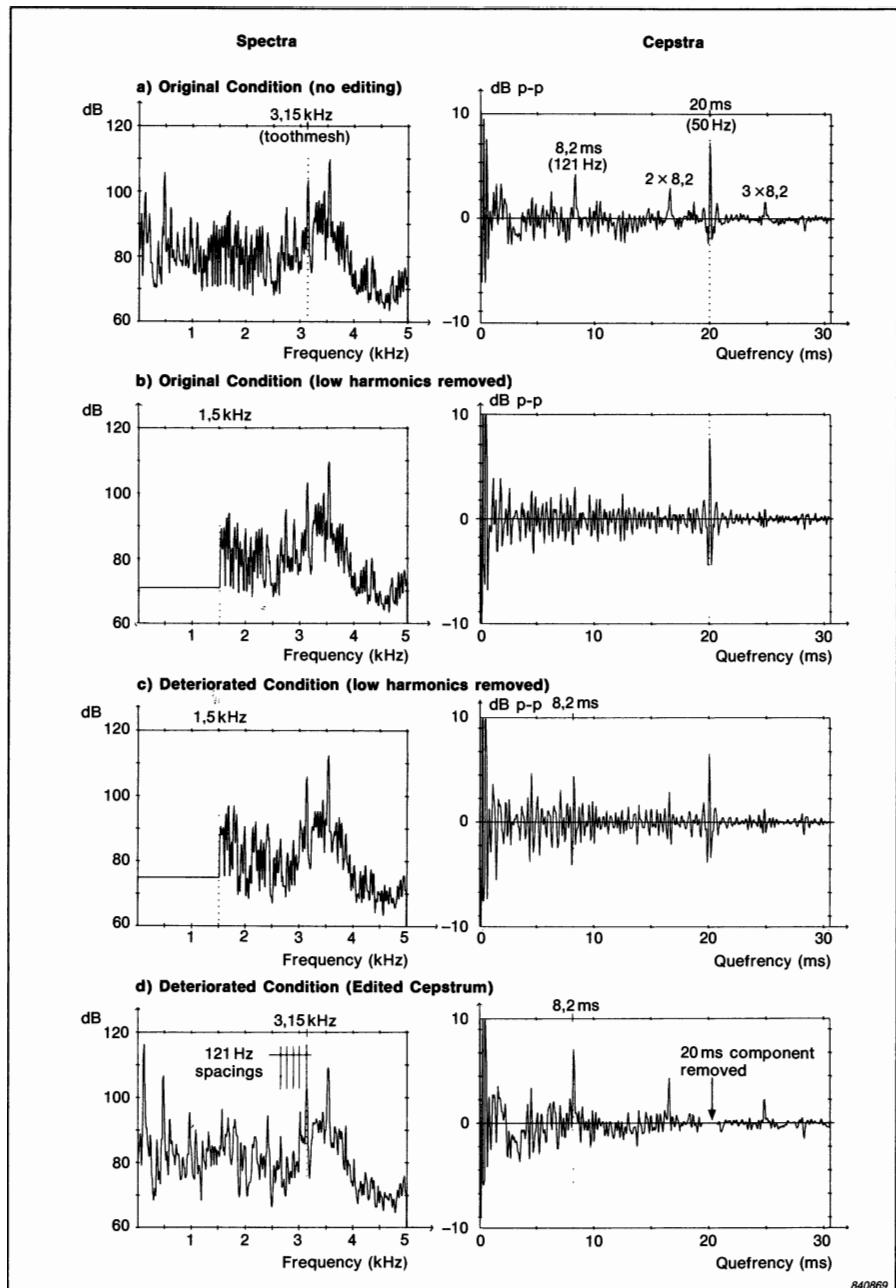


Fig. 8.11. Application of cepstrum analysis to gearbox diagnosis

effects of the 50 Hz gear. The original cepstrum was from the same signal as Fig.8.11(c), but comprising the whole frequency range. The spectrum shown was obtained by transformation back from the edited cepstrum, and shows the pattern of 121 Hz sidebands which were effectively masked in Fig.8.11(c).

Note that the procedure indicated in Fig.8.11(d) gives a similar result to that which could be achieved by synchronous averaging (synchronized with the rotation of the 121 Hz gear) without the need for a synchronizing signal.

Ref.8.14 contains a more detailed discussion of the application of the power cepstrum to gear diagnostics, while the application of the complex cepstrum is considered in Section 8.6.2.

8.4. PRACTICAL CONSIDERATIONS WITH THE POWER CEPSTRUM

The results of a cepstrum analysis are affected to a considerable extent (compared with normal spectrum analysis) by artefacts introduced by the analysis process, and by the signal itself, and so it is as well to discuss a number of practical points.

Firstly, wherever the spectrum value is zero, in theory it is not possible to take the logarithm. As an example, a truly periodic signal has finite values only at the harmonic frequencies, and cepstrum analysis is theoretically impossible. In practice, the presence of noise usually places a lower limit on (logarithmic) spectrum level, even if it is the quantization noise in the FFT process. This allows the cepstrum calculation to be carried out in a practical case. However, as illustrated in Fig.8.12, the noise level in the spectrum will obviously have an effect on the cepstrum results, and this should always be kept in mind when

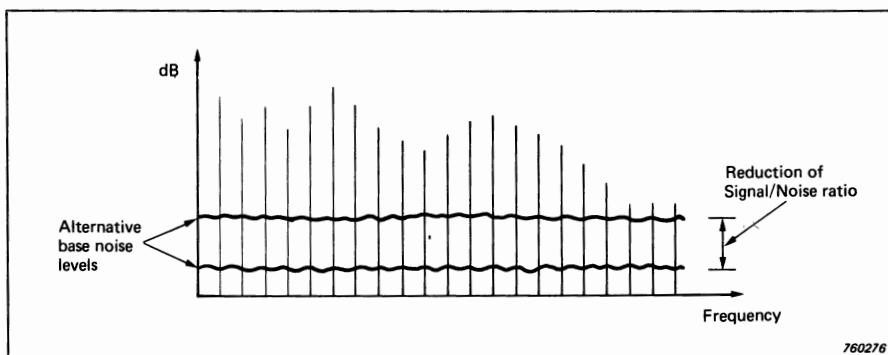


Fig. 8.12. Effect of noise level on size of cepstrum component

making comparisons. Often, comparative measurements are made on signals measured at the same point at different times, and it is valid to assume that the background noise level is roughly the same at all times. Note that the relative height of discrete frequency components above the noise is influenced by the analysis bandwidth, and thus comparisons should only be made between cepstra obtained under identical conditions.

The relative noise level is not the only factor influenced by the filter bandwidth (and characteristic). As illustrated in Fig.8.13, if discrete frequency components are not adequately separated (or if the filter characteristic is too poor) a bridging can occur between adjacent spectral components which reduces the corresponding cepstral component. As this bridging would tend to be a fixed number of dB below the peaks, growth in a group of harmonics/sidebands would tend not to register in the cepstrum component (as compared with growth above a fixed noise level). For this reason it is essential that equally spaced components are adequately resolved in the original spectrum. As a guide, the minimum spacing (i.e. the lowest periodic frequency) should represent a minimum of eight lines spacing in the spectrum (where Hanning weighting is used).

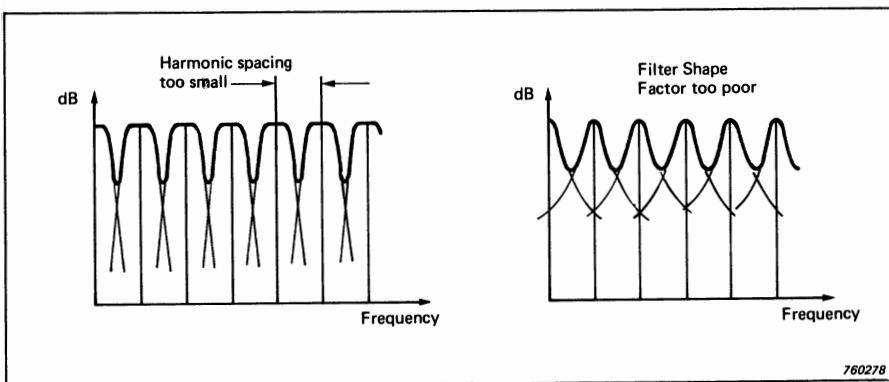


Fig. 8.13. Effect of filter bandwidth and shape factor

Note that the filter characteristic (and harmonic spacing) also affect the distribution in the cepstrum between the various rahmonics corresponding to a particular spectral spacing. This is another reason for only making comparisons under identical analysis conditions. In general the most important information is contained in the first rahmonic (i.e. how much the spectrum fluctuates up and down); the higher rahmonics only give information about the distortion of a particular periodicity, and this is influenced by artefacts such as the filter characteristic. If the higher rahmonics are decreasing monotonically they can

usually be ignored (except as confirmation that one is dealing with a periodic structure rather than a random peak). However, Fig.8.14 shows an example where every third rahmonic was predominant, and this indicates that both a fundamental rotational speed (corresponding to the third rahmonic) and its third harmonic (corresponding to the first rahmonic) each had a separate influence in the spectrum. As described in Ref.8.14, the third harmonic effect was due to "triangularity" of a gear.

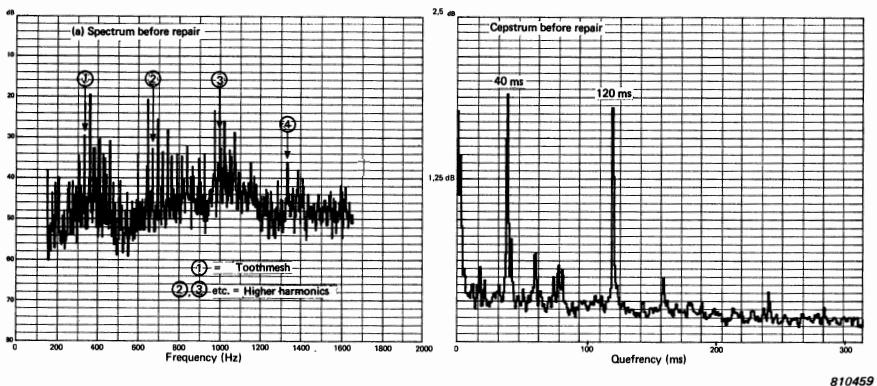


Fig. 8.14. Spectrum and cepstrum for a case where the third rahmonic (120 ms) represents a component in its own right

In order to achieve adequate resolution in the spectrum, it will sometimes be necessary to use zoom. A cepstrum analysis can be made on a zoom spectrum, but it is then virtually essential to use the magnitude of the analytic signal defined in Eqn.(8.4). Fig.8.15 shows an example similar to that of Fig.8.1. When the cepstrum is obtained from a baseband spectrum, the uniformly spaced components fall at exact harmonic frequencies and the peaks in the real part of the cepstrum are all positive. When a zoom analysis is made, with centre frequency shifted very slightly upwards, the peaks are no longer confined to the real part, but are correctly indicated in the magnitude $|C_{AA}(\tau)|$. A similar situation could arise in a baseband analysis where the uniformly spaced components (when projected) do not pass through zero frequency. This would for example be the result of modulation of a carrier frequency by a non sub-harmonic modulating frequency. Examples of this occur in the vibrations of planetary gears, and of rolling element bearings.

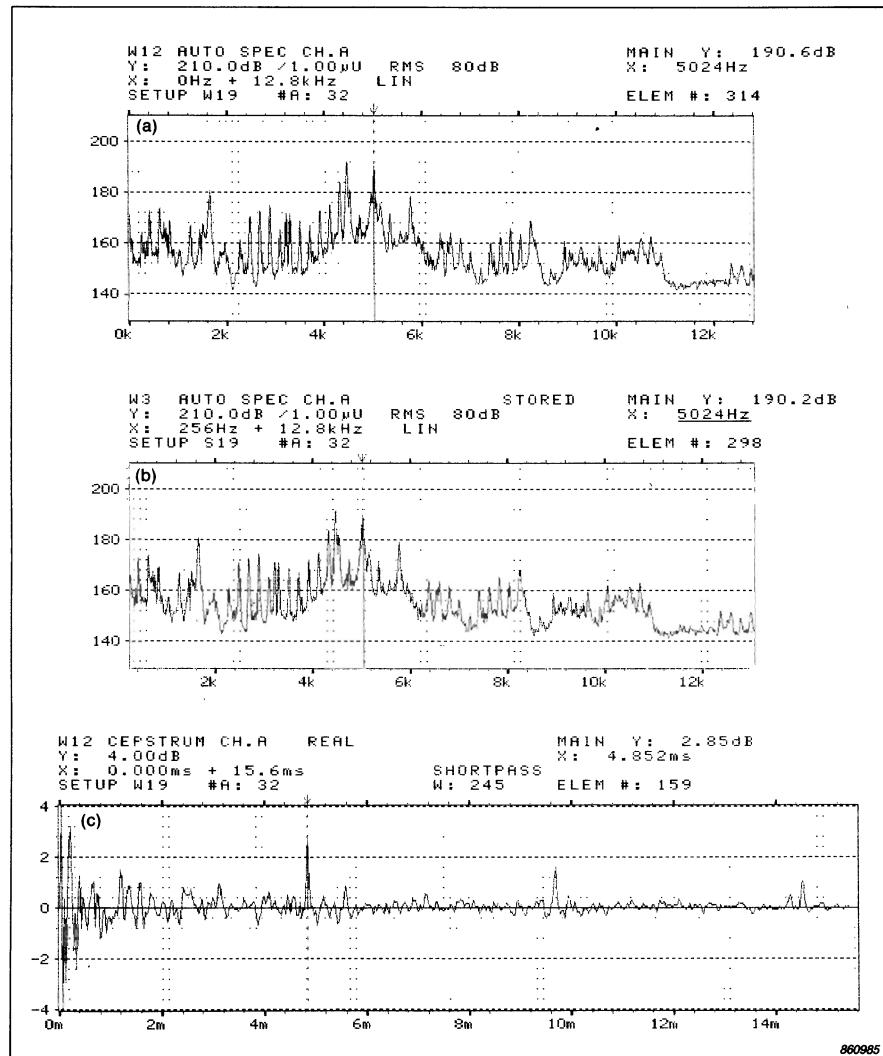
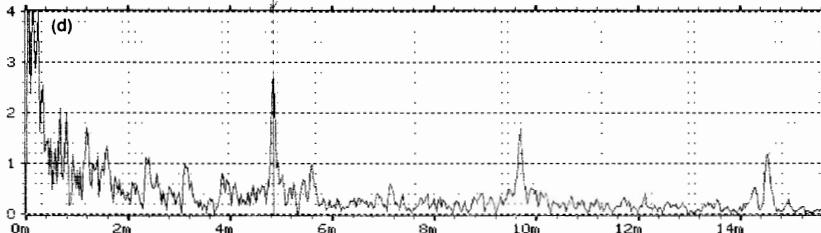


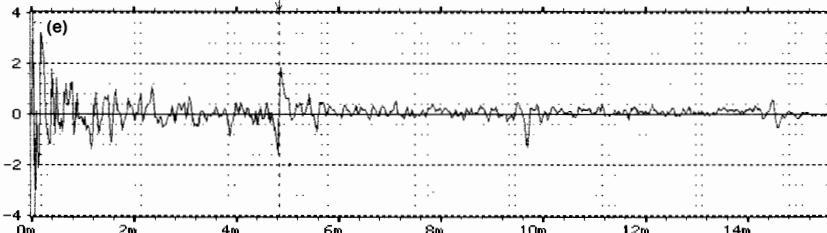
Fig. 8.15. Cepstra from spectra obtained by zoom

- (a) Baseband spectrum 0–12800 Hz
- (b) Zoom spectrum shifted upwards in frequency by 256 Hz
- (c) Real part of cepstrum (normal definition) from (a)

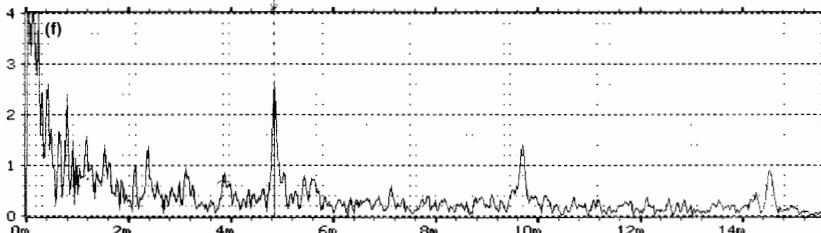
W3 CEPSTRUM CH.A MAG
 Y: 4.00dB X: 0.000ms + 15.6ms
 SETUP W19 #A: 32 SHORTPASS W: 245 ELEM #: 159
 MAIN Y: 2.86dB X: 4.852ms



W12 CEPSTRUM CH.A REAL INPUT MAIN Y: 0.34dB
 Y: 4.00dB X: 0.000ms + 15.6ms
 SETUP W19 #A: 32 SHORTPASS W: 245 ELEM #: 159
 X: 4.852ms



W3 CEPSTRUM CH.A MAG INPUT MAIN Y: 2.68dB
 Y: 4.00dB X: 0.000ms + 15.6ms
 SETUP W19 #A: 32 SHORTPASS W: 245 ELEM #: 159
 X: 4.852ms



860986

*Fig. 8.15. (d) Magnitude of analytic signal corresponding to (c)
 cont. (e) Real part of cepstrum from (b)
 (f) Magnitude of analytic signal corresponding to (e)*

There is no general agreement on the scaling of cepstral results. In the author's opinion, this can best be expressed in terms of "dB peak-to-peak". Since logarithmic power spectra are normally expressed on a dB scale, their Fourier transforms (the cepstra) can also be expressed in terms of dB. Fig.8.16 illustrates the meaning of scaling in dB peak-peak. If the spectrum had the appearance of a sinusoid fluctuating up and down over an 80 dB range, it would be scaled in the cepstrum as 80 dB p-p. In an actual case it tells how many dB the spectrum, on the average, fluctuates up and down with a certain frequency spacing. Note that only the zero quefrency ("DC") component in the cepstrum contains any information on actual scaling in terms of physical units. The zero quefrency component can be expressed in dB re a given reference level. All other quefrency components are in non-dimensional dB's, representing pure ratios (fluctuations about the "DC" level).

Note that scaling in this way implies that the spectrum is treated as a stationary signal. The results of Fig.8.11, for example, have been compensated for the shorter length of windowed spectrum in (b) and (c).

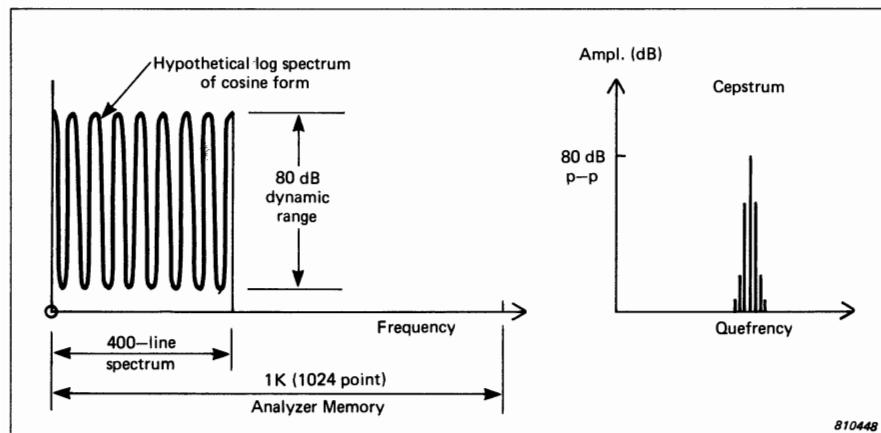


Fig. 8.16. Cepstrum amplitude calibration in dB peak-peak illustrated through the case of a 400-line spectrum transformed to the cepstrum using a 1 K (1024-point) transform

As already mentioned, the results are in any case dependent on analysis method (and the signal itself) and the most important thing is to use consistent scaling techniques and only compare cepstra obtained under similar conditions. Note that the analyzers Types 2032/2034 do not scale in terms of dB peak-peak, but in consistent units.

8.5. THE COMPLEX CEPSTRUM

In principle the complex cepstrum is considerably more powerful than the power cepstrum, but it is much more difficult to deal with. The main reason for this is that the phase function $\phi(f)$ in Eqn. (8.7) must be a continuous function of frequency rather than the principal value modulo 2π as is normally measured. Fig.8.17 illustrates what is meant by "phase unwrapping", to obtain a continuous phase function. Where the phase curve is fairly smooth, it is not very difficult to unwrap, using a simple criterion such as that the phase jump between adjacent samples should be less than π . This is not necessarily always correct, however, and in the general case it is necessary to use a complex algorithm such as that of Trigolet (Ref.8.15).

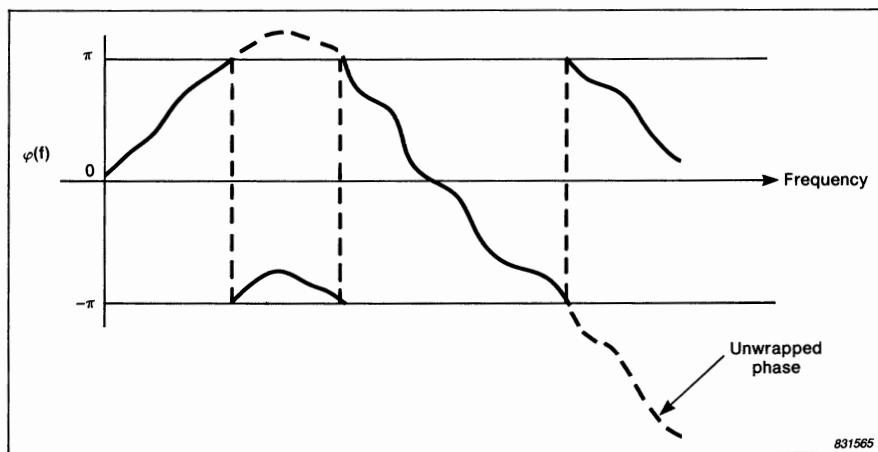


Fig. 8.17. Illustration of the meaning of "Phase-unwrapping", whereby the phase spectrum is made a continuous function of frequency

There is one situation, however, where the phase does not have to be measured, and that is in the case of so-called "minimum phase" functions. As shown in Ref.8.16, these have the property that the phase function $\phi(f)$ is the (inverse) Hilbert transform of the log magnitude function $\ln |A(f)|$. It follows immediately that the complex cepstrum of a minimum phase function is causal (i.e. exists for positive quefrequencies only), because the real and imaginary parts of its Fourier transform are related by a Hilbert transform (Section 2.6). The even part is the power cepstrum (though obtained from a log amplitude spectrum scaled in nepers) and the odd part is the "phase cepstrum", i.e. the cepstrum of the phase function alone. By analogy with Fig.2.31, the power cepstrum and phase cepstrum of a minimum phase function must be identical at positive quefrequencies.

Thus if one is dealing with minimum phase functions, the complex cepstrum can be obtained from the power cepstrum by simply doubling the positive quefrency components, and setting negative quefrency components to zero.

A minimum phase function is one which has minimum phase lag, or phase delay, for a given spectrum amplitude function. As explained in Ref.8.16, and briefly in Appendix C, the transfer function of a stable, causal function has no poles in the right half of the Laplace plane (where damping exponent $-\sigma$ is positive). A minimum phase function must neither have poles nor zeroes in the right half plane, because on taking logarithms the zeros become poles, which would not allow the cepstrum to be stable and causal. Thus, a stable, causal, non minimum phase function differs from a minimum phase function only by the presence of zeroes in the right half plane. As shown in Figure C7 of Appendix C, such a transfer function can always be considered to be the product of a minimum phase function with an all-pass function having pole-zero pairs of the general form $\frac{s-a}{s+a^*}$. The all-pass function has unit amplitude, but adds a phase lag of 2π for the passage of each pole-zero pair. This explains the term "minimum phase function". Note that because of the finite impulse response time of the all-pass function, the minimum phase function also has the shortest impulse response for a given frequency response amplitude function.

Because the effect of a frequency response function is additive in the cepstrum domain (Eqn. (8.12)) it is of interest to investigate what form this will have. It is simplest to do this using the z-transform, rather than the Laplace transform (Ref.8.17). As shown in Ref.8.17, and briefly in Appendix C, the z-transform version of a typical transfer function may be expressed as

$$H(z) = \frac{|A| \prod_{k=1}^{m_i} (1 - a_k z^{-1}) \prod_{k=1}^{m_o} (1 - b_k z)}{\prod_{k=1}^{p_i} (1 - c_k z^{-1}) \prod_{k=1}^{p_o} (1 - d_k z)} \quad (8.18)$$

where the a_k and c_k are zeroes and poles, respectively, inside the unit circle (of Fig.3.34), which corresponds to the left half of the s-plane, and the b_k and d_k are zeroes and poles outside the unit circle (corresponding to the right half of the s-plane). The moduli $|a_k|$, $|b_k|$, $|c_k|$ and $|d_k|$ are all < 1 .

Taking logs and expanding the $\log(1 - \alpha z^{\pm 1})$ terms as power series, gives the (sampled) cepstrum in terms of the poles and zeroes, for incremental values of the time sample number n .

$$\text{Thus, } C_a(n) = - \sum_{k=1}^{m_i} \frac{a_k^n}{n} + \sum_{k=1}^{p_i} \frac{c_k^n}{n} \quad , \quad n > 0 \quad (8.19)$$

$$\text{and, } = \sum_{k=1}^{m_o} \frac{b_k^{-n}}{n} - \sum_{k=1}^{p_o} \frac{d_k^{-n}}{n} \quad , \quad n < 0$$

As expected, the minimum phase part appears only at positive quefrequencies, and the “maximum phase” part at negative quefrequencies. For stable causal functions, the d_k will be zero, so the maximum phase part will come entirely from the zeroes, b_k . Each of the terms is a complex exponential (in conjugate pairs a damped sinusoid) further damped by multiplication by the hyperbolic function $1/n$.

8.6. APPLICATIONS OF THE COMPLEX CEPSTRUM

The complex cepstrum must be used whenever it is desired to return to the time domain after editing in the cepstrum.

8.6.1. Echo Removal

Fig.8.18 shows a numerically generated example of echo removal from a time signal using the complex cepstrum. The original signal is a damped one-sided sinewave (the impulse response of a single degree-of-freedom system) to which has been added two equi-spaced echoes. Even though these overlap the original signal, the delta functions in the complex cepstrum are reasonably well removed from the cepstrum of the basic signal (which dies out more rapidly than the signal itself). It is thus relatively simple to remove the effects of the echo from the cepstrum, and transforming all the way back to the time signal shows that the removal has been quite efficient.

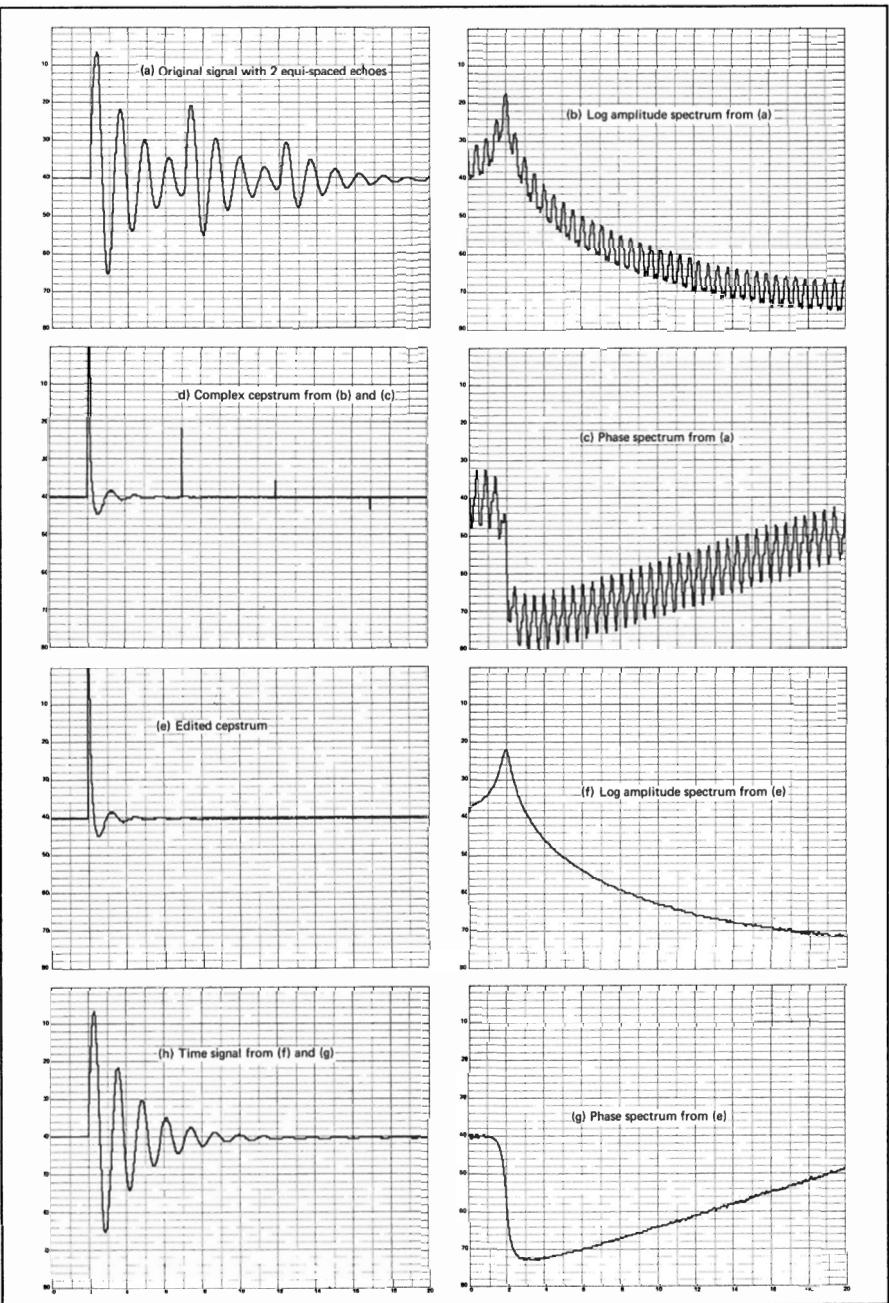


Fig. 8.18. Echo removal using complex cepstrum

Fig.8.19 shows the result of applying the same technique on an actual loudspeaker excited by a $50\ \mu\text{s}$ square pulse in a normal room. The (first part of the) time signal in Fig.8.19(a) shows a number of reflections starting at about 6.2 ms delay time. In contrast to the numerically generated example of Fig.8.18, the phase spectrum had to be “unwrapped” and scaled down appropriately before calculation of the complex cepstrum, which is shown in Fig.8.19(b). The effect of the reflections in the cepstrum is seen to be rather spread out, indicating non-ideal reflections and so a fairly drastic “short-pass lifter” was applied to remove all reflections (a half Hanning window falling to zero at 5 ms). The resulting log amplitude and phase characteristics (to 10 kHz) are shown in Fig.8.19(c). For comparison purposes, a measurement made on the same loudspeaker using the TDS (Time Delay Spectrometry) technique is shown in Fig.8.20. Even though this was made in a different room, the similarities are quite striking.

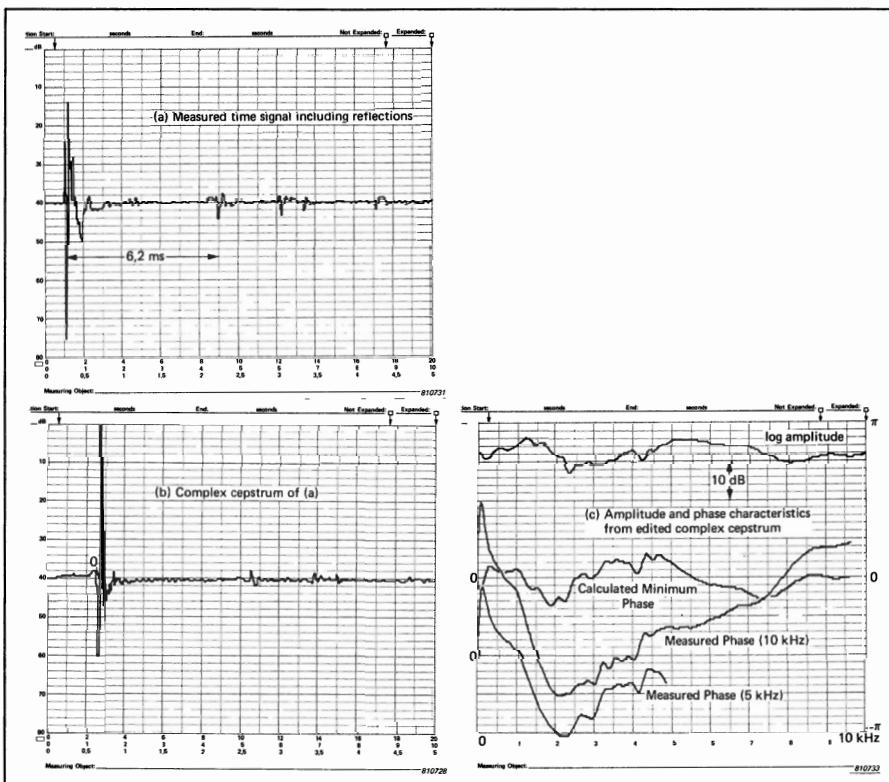


Fig. 8.19. Measurement of loudspeaker characteristics in a hard room using complex cepstrum

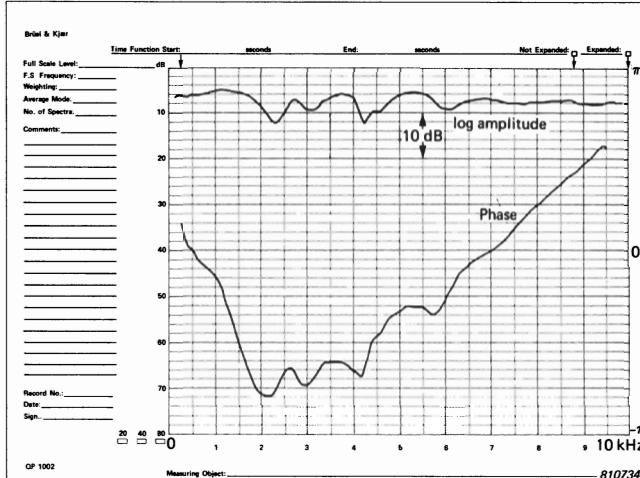


Fig. 8.20. Loudspeaker characteristics measured using TDS system

At the same time as the 10 kHz results of Fig.8.19 another measurement was made using the cepstrum technique but with an upper limiting frequency of 5 kHz. Because of the double length of record, the added noise (and reflections) caused problems with the phase unwrapping algorithm. There were three places in the spectrum where the algorithm "jumped" the wrong way, introducing a discontinuity. In this case it was possible to solve the problem by applying a decaying exponential window to the original time signal, thus considerably reducing the effect of noise (and reflections) at the end of the record. After removal of the reflections in the cepstrum, the signal was transformed all the way back to the impulse response which was compensated for the exponential window before forward transforming again to obtain the amplitude and phase characteristics. The phase result is drawn in Fig.8.19(c) on the same scale (but with twice the resolution).

The application of the complex cepstrum to loudspeakers is one case where the phase unwrapping is not too difficult because the phase spectrum should be smooth and the amplitude spectrum should not contain zeroes. For more general signals, the phase unwrapping can present considerable problems.

As a matter of interest, Fig.8.19(c) also includes the minimum phase characteristic corresponding to the measured log amplitude characteristic (to 10 kHz), calculated as outlined in Section 8.5.

8.6.2. Deconvolution

One of the applications in speech analysis and resynthesis (Refs.8.18, 8.19) involves deconvolution of the formant information from the voice/sound source. As discussed in Section 8.3.3., most of the intelligence in the speech is contained in the low quefrency part of the cepstrum (only a fraction of the total), and so this can be efficiently transmitted along with information as to whether the speech is voiced, and if so, the voice pitch. At the receiver end, the speech is reconstituted using the low quefrency information to generate a filter characteristic (or impulse response) for a source which would either be a variable frequency pulse generator, for the voiced sections, or a noise generator for the unvoiced sections. Despite the synthetic voice, the speech was reported as sounding natural.

Another area where some progress has been made, is in the field of machinery diagnostics. Vibration signals measured externally on a machine are always a compound of source and transmission path effects. With the machine in operation it is normally too difficult to measure the forcing functions at the same time as the response in order to determine the frequency response function of the transmission path. However, as we have seen in Equations (8.11) and (8.12), for a single dominant source and transmission path, the two effects are additive in both the logarithmic spectrum and cepstrum. Moreover, as for speech analysis, they are often largely separated into different regions in the cepstrum.

One example is in gearbox vibrations, where the forcing function at the tooth mesh has well-defined characteristics (Ref.8.20). The three dominant effects are:

- (1) Mean effects showing up at the toothmesh frequency and its harmonics.
- (2) Non-uniformly distributed effects (e.g. local faults) showing up at the harmonics of the individual gear speeds.
- (3) Mesh transfer functions, expressing the lowpass filter effects of loadsharing between teeth and dominated by the "contact ratios" (average number of teeth in contact in either the peripheral or axial directions).

The first two of these effects are confined to very localised regions in the cepstrum, while the third has a calculable effect. Ref.8.21 shows that a simple windowing in the cepstrum (of a response signal) below the lowest harmonic of effects (1) and (2) above, and allowing in a simple way for the contact ratio effect, can give a frequency response function (from the source to the external measurement point) which compares favourably with a direct measurement (applying a force at the mesh with the machine stopped). The major difference between the two results appeared to be a lack of high quefrency information in

the result obtained from the cepstrum, and Ref.8.22 suggests that high frequency information could be obtained selectively from between the harmonics (from the forcing function) and used in curve fitting algorithms to derive the frequency response functions based on Eqn.8.19. This application requires further development, but it would be very valuable to be able to determine whether changes in an externally measured vibration signal result from a change in the forcing function, or in the structural response.

A somewhat similar technique has been used in Ref.8.23 to regenerate internal cylinder pressure signals in a diesel engine, from externally measured vibration signals, making use of windowing in the cepstrum domain.

APPENDIX A

FOURIER ANALYSIS

(a) Integration of a vector rotated through π

Without loss of generality, the case illustrated in Fig.2.16 can be taken, where the resultant is directed along the real axis. Fig.A1 illustrates a typical component vector of length A at angle θ .

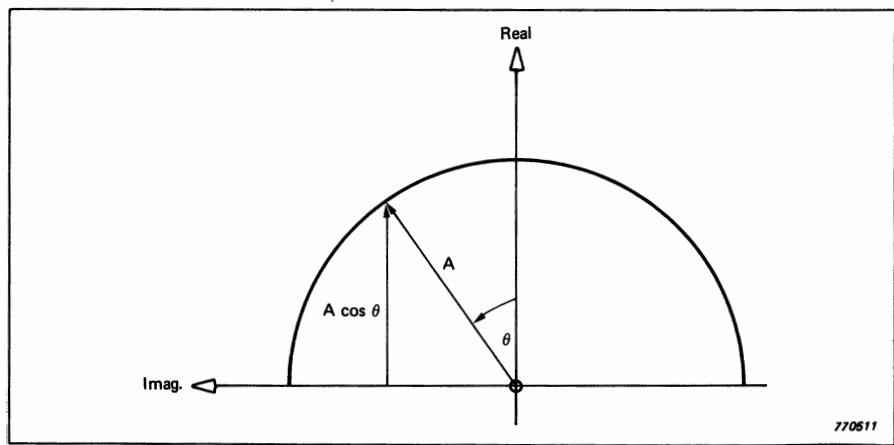


Fig. A1. Integration of a rotating vector from $-\pi/2$ to $\pi/2$

Its contribution to the resultant is $A \cos \theta$ as illustrated, and thus the average over the total angle between $-\pi/2$ and $\pi/2$ is given by

$$A_{\text{result}} = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} A \cos \theta \, d\theta \quad (\text{A.1})$$

$$\begin{aligned}
 &= \frac{A}{\pi} \left[\sin \theta \right]_{-\pi/2}^{\pi/2} \\
 &= \frac{2A}{\pi} \quad \text{as stated in Eqn.(2.21)}
 \end{aligned}$$

(b) Fourier Transform of a rectangular function length T

Fig.A2 depicts a rectangular function of length T and height A , evenly divided about zero time.

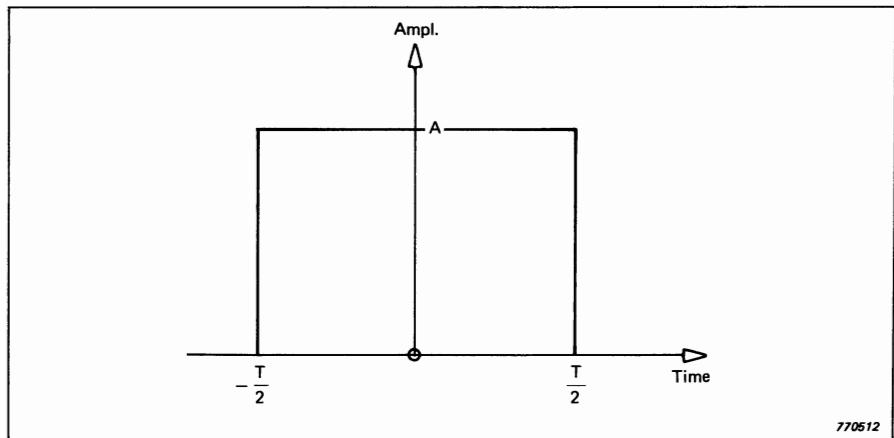


Fig. A2. Rectangular function of length T and height A

Mathematically, the time function may be defined as:

$$\begin{aligned}
 g(t) &= A \quad , \quad -\frac{T}{2} < t < \frac{T}{2} \\
 &= 0 \quad , \quad \text{otherwise}
 \end{aligned} \tag{A.2}$$

Thus, from Eqn.(2.14) the Fourier Transform is given by:

$$\begin{aligned}
 G(f) &= \int_{-\infty}^{\infty} g(t) e^{-j2\pi ft} dt \\
 &= A \int_{-T/2}^{T/2} e^{-j2\pi ft} dt
 \end{aligned}$$

$$\begin{aligned}
 &= -\frac{A}{j2\pi f} \left[e^{-j2\pi ft} \right]_{-T/2}^{T/2} \\
 &= \frac{jA}{2\pi f} \left[\cos(2\pi ft) - j\sin(2\pi ft) \right]_{-T/2}^{T/2} \\
 &= AT \frac{\sin(\pi f T)}{(\pi f T)}
 \end{aligned} \tag{A.3}$$

(c) Bandwidth of a $\sin x/x$ function

Fig.A3(a) shows the amplitude characteristic corresponding to Eqn.(A.3), normalised to a peak value of unity (this can be done without loss of generality when it is the bandwidth which is required). The amplitude characteristic is thus a $|\sin x/x|$ function where $x = \pi f T$. The amplitude squared, or power transmission characteristic is shown in Fig.A3(b) and it is this which must be integrated to obtain the total power transmitted from a unit white noise source.

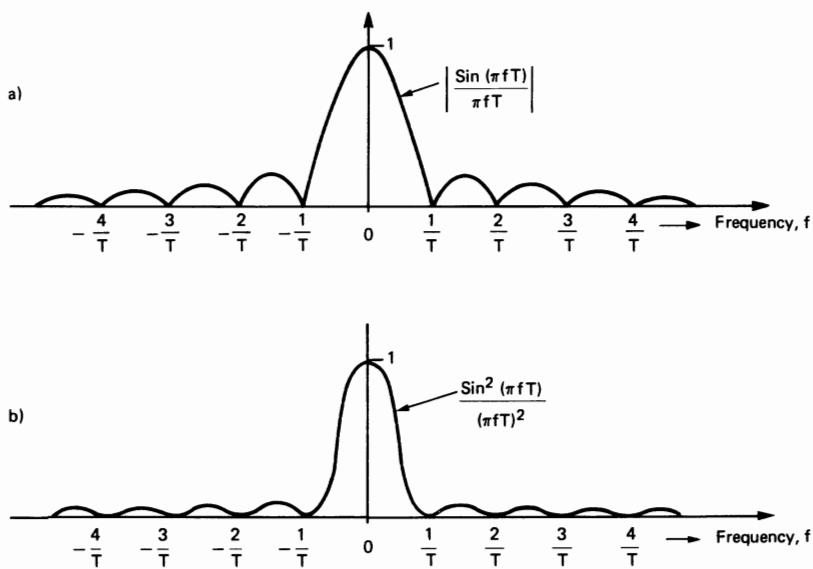


Fig. A3. (a) $|\sin x/x|$ (b) $\sin^2 x/x^2$ where $x = \pi f T$

770519

Thus, the total area under the power transmission curve:

$$\begin{aligned} &= \int_{-\infty}^{\infty} \frac{\sin^2(\pi f T)}{(\pi f T)^2} df \\ &= \int_{-\infty}^{\infty} \frac{\sin^2 x}{x^2} \cdot \frac{df}{dx} \cdot dx , \quad \text{where } x = \pi f T \\ &= \frac{1}{\pi T} \int_{-\infty}^{\infty} \frac{\sin^2 x}{x^2} dx \end{aligned} \tag{A.4}$$

From tables of standard integrals, the value of the integral is found to be π , resulting in a total area of $1/T$. Since the peak amplitude was normalised to unity, dividing by this gives:

$$B_{\text{eff}} = \frac{1}{T} \quad \text{as stated in Eqn.(2.23)}$$

(d) Fourier Transform of an exponential function

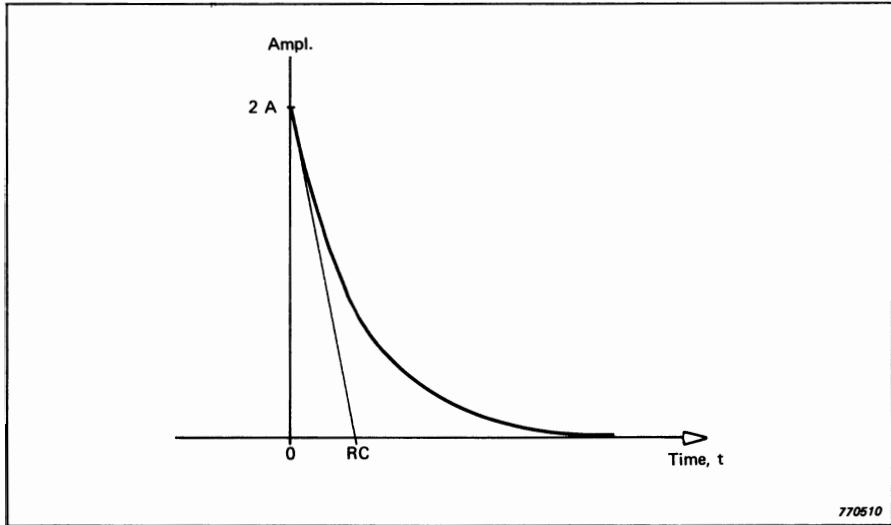


Fig. A4. Decaying exponential $2Ae^{-t/RC}$

770510

Fig.A4 depicts the exponential function defined by:

$$g(t) = \begin{cases} 2Ae^{-t/RC} & , \quad t \geq 0 \\ 0 & , \quad \text{otherwise} \end{cases} \quad (\text{A.5})$$

Thus, from Eqn.(2.14) its Fourier Transform is given by:

$$\begin{aligned} G(f) &= 2A \int_0^\infty e^{-t/RC} e^{-j2\pi ft} dt \\ &= 2A \int_0^\infty e^{-(1/RC + j2\pi f)t} dt \\ &= \frac{-2A}{\left(\frac{1}{RC} + j2\pi f\right)} \left[e^{-(1/RC + j2\pi f)t} \right]_0^\infty \\ &= \frac{2ARC}{1 + j2\pi fRC} \end{aligned} \quad (\text{A.6})$$

The peak amplitude of this function, occurring where $f = 0$, is thus given by $A(2RC)$ which will be seen to be the same as for the spectrum of the rectangular function (Eqn.A.3) if $T = 2RC$.

(e) Bandwidth of the spectrum of a decaying exponential

Normalising the function of Eqn.(A.6) to a peak value of unity gives $1/(1 + j\pi f 2RC)$ which can be expressed as

$$\frac{1 - j\pi f 2RC}{1 + (\pi f 2RC)^2}$$

The modulus squared of this function is:

$$\frac{1}{1 + (\pi f 2RC)^2}$$

and the total integral under this power transmission curve is

$$\int_{-\infty}^{\infty} \frac{df}{1 + (\pi f 2RC)^2} = \frac{1}{\pi 2RC} \int_{-\infty}^{\infty} \frac{dx}{1 + x^2} \quad , \quad \text{where } x = \pi f 2RC$$

$$\left(\because \frac{df}{dx} = \frac{1}{\pi 2RC} \right)$$

From tables of standard integrals the value of the infinite integral is π and thus the total area under the curve is $1/2 RC$.

The effective bandwidth is obtained by dividing this by the peak value (unity) and thus:

$$B_{\text{eff}} = \frac{1}{2RC} \quad (\text{A.7})$$

as illustrated in Fig.3.14.

(f) The Convolution Theorem

Referring to Section 2.5.3, it is given that:

$$G(f) = \int_{-\infty}^{\infty} g(t) e^{-j2\pi ft} dt \quad (\text{A.8})$$

$$F(f) = \int_{-\infty}^{\infty} f(t) e^{-j2\pi ft} dt \quad (\text{A.9})$$

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{-j2\pi ft} dt \quad (\text{A.10})$$

$$\text{and } g(t) = f(t) * h(t) = \int_{-\infty}^{\infty} f(\tau) h(t - \tau) d\tau \quad (\text{A.11})$$

Substituting Eqn.(A.11) in (A.8) gives

$$G(f) = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(\tau) h(t - \tau) d\tau \right] e^{-j2\pi ft} dt$$

which by reversing the order of integration gives:

$$\begin{aligned} G(f) &= \int_{-\infty}^{\infty} f(\tau) \left[\int_{-\infty}^{\infty} h(t - \tau) e^{-j2\pi ft} dt \right] d\tau \\ &= \int_{-\infty}^{\infty} f(\tau) \left[\int_{-\infty}^{\infty} h(u) e^{-j2\pi f(u + \tau)} du \right] d\tau \end{aligned}$$

where $u = t - \tau$ (and thus $du = dt$)

This reduces to:

$$G(f) = \left[\int_{-\infty}^{\infty} f(\tau) e^{-j2\pi f\tau} d\tau \right] \left[\int_{-\infty}^{\infty} h(u) e^{-j2\pi fu} du \right]$$

which by substituting Equations (A.9) and (A.10) gives:

$$G(f) = F(f) \cdot H(f) \quad \text{as stated in Eqn.(2.31)}$$

APPENDIX B

Mean Square Error for Narrow Band Random Noise

The development given here basically follows that of Ref.3.1 but makes use of concepts developed in Chapter 2 to clarify the argument.

A narrow band stationary random signal of bandwidth B is considered, such as would be passed by an ideal filter of bandwidth B from a white noise source. Fig.B1(a) illustrates the power spectrum of this signal which has a constant power spectral density W_B within the passband. (Note that the 2-sided spectrum representation is assumed.)

The complex spectrum corresponding to Fig.B1(a) (i.e. the direct Fourier transform of the time signal $g(t)$) may be termed $G(f)$ and has the properties:

$$\left. \begin{array}{l} |G(f)|^2 = W_B \\ \angle G(f) = \text{random} \end{array} \right\} \text{inside the passbands}$$
$$\left. \begin{array}{l} |G(f)|^2 = 0 \\ \angle G(f) = \text{undefined} \end{array} \right\} \text{outside the passbands} \quad (\text{B.1})$$

When the signal $g(t)$ is passed through a squaring circuit to obtain its power, the resulting multiplication in the time domain ($g(t) \times g(t)$) transforms by the Fourier transform to a convolution in the frequency domain (see Section 2.5.3).

Thus, from Eqn.2.30 the frequency spectrum of $g^2(t)$ ($= y(t)$) will be given by:

$$Y(f) = \int_{-\infty}^{\infty} G(\phi)G(f-\phi) d\phi \quad (\text{B.2})$$

The evaluation of this convolution integral must be treated separately for the two cases:

- (1) Displacement frequency $f = 0$
- (2) Displacement frequency $f \neq 0$

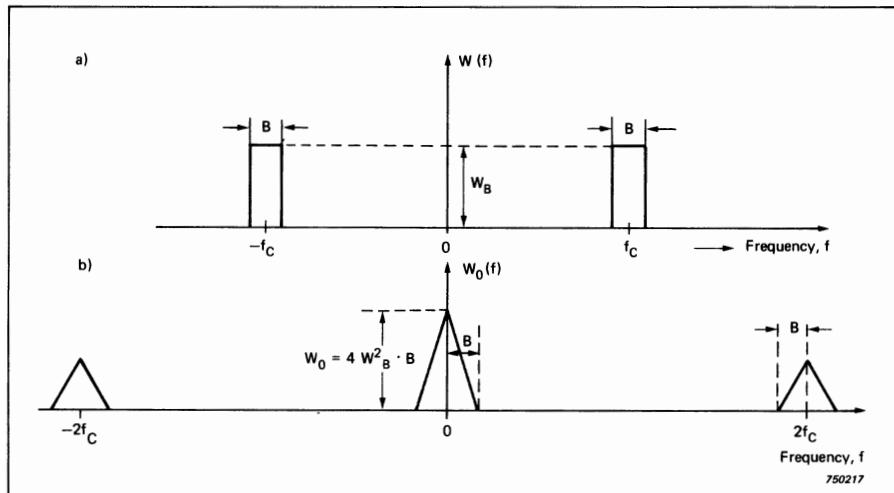


Fig. B1. Sketches illustrating the effect of squaring a narrow frequency band of random noise.

- (a) *The noise frequency spectrum before squaring*
- (b) *The output spectrum from the squaring device*

For $f = 0$, Equation (B.2) reduces to

$$Y_{DC}(0) = \int_{-\infty}^{\infty} G(\phi)G(-\phi) d\phi \quad (\text{B.3})$$

Since $g(t)$ is real, its spectrum $G(\phi)$ is conjugate even (Eqn.(2.37)) and thus:

$$G(-\phi) = G^*(\phi) \quad (\text{B.4})$$

Substituting in Eqn.(B.3) gives:

$$\begin{aligned}
 Y_{DC}(0) &= \int_{-\infty}^{\infty} G(\phi) \cdot G^*(\phi) d\phi \\
 &= \int_{-\infty}^{\infty} |G(\phi)|^2 d\phi \\
 &= 2 \int_{f_c-B/2}^{f_c+B/2} W_B d\phi, \text{ making use of Eqns.(B.1)} \\
 &= 2BW_B
 \end{aligned} \tag{B.5}$$

$Y_{DC}(0)$ is of course the DC component of the squared signal, and considering its dimensions (PSD \times bandwidth) it obviously has finite power. Thus, to express it on a power spectral *density* scale it must be a delta function weighted with the value indicated by Eqn.(B.5).

As soon as the origins of the two spectra $G(\phi)$ and $G(-\phi)$ are displaced even slightly from each other, i.e. $f \neq 0$ in Eqn.(B.2), then the remarkable effect due to Eqn.(B.4) no longer applies, and juxtaposed frequency components will now have completely random phase relationships (instead of the phase always cancelling out to zero at each frequency so that the amplitudes W_B add directly). Thus the integration in frequency must be a vector addition of components with random phase (even though their amplitudes will still be equal to W_B) and a different approach is necessary.

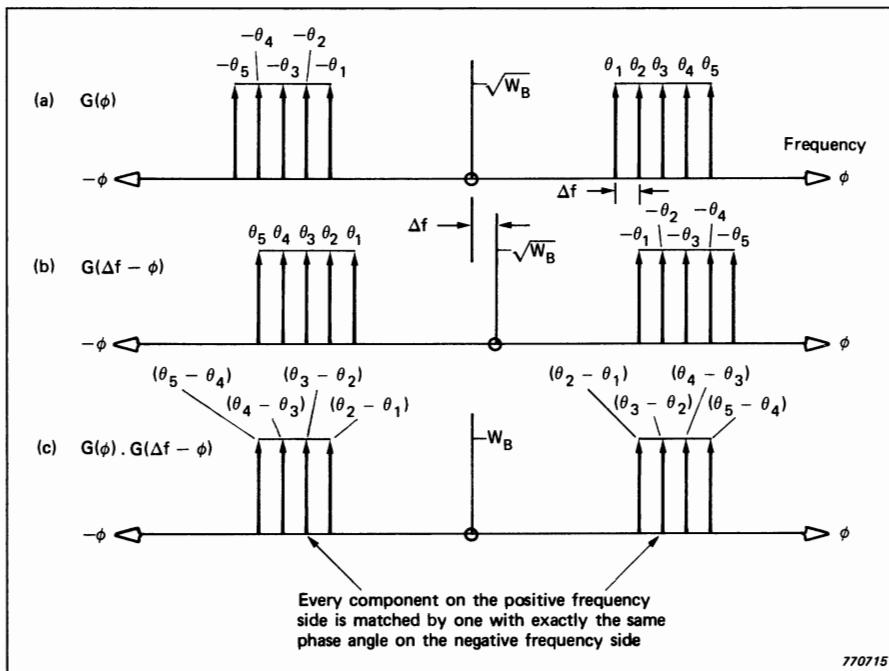
As indicated at the end of Section 2.5.3, for signals with random phase (which are multiplied in the time domain) it is valid to obtain the power spectral density of the result by convolving the individual PSD spectra. Fig.B1(b) indicates the results of doing this for the spectrum of Fig.B1(a). The central triangular portion results from the fact that the greater the displacement of the two identical power spectra, the less the overlap of the rectangular passbands, the result decreasing linearly down to zero at a displacement equal to the bandwidth B (zero overlap). The two smaller triangles occur similarly when the positive frequency passband of one spectrum coincides with the negative frequency passband of the other. In fact it is not strictly true that the power spectra are convolved in this case because of the symmetry which exists about zero frequency, even when the two complex spectra $G(\phi)$ and $G(-\phi)$ are displaced. (This would not be the case if $g(t)$ were multiplied by a different signal with the same PSD.) The scaling of the result can be determined by considering the limiting case as displacement frequency $f \rightarrow 0$.

For the maximum value of the distributed portion of the spectrum (as opposed to the DC component already obtained), Eqn.(B.2) gives:

$$Y_{AC}(0) = \lim_{f \rightarrow 0} \int_{-\infty}^{\infty} G(\phi)G(f-\phi) d\phi \tag{B.6}$$

The difference with respect to Eqn.(B3) is that although the amplitude of the vector to be integrated is still $|G(f)|^2$ ($= W_B$) its phase angle is random (however small f is). Thus, considering the positive frequencies only, the amplitude of the vector resulting from the integral of Eqn.(B6) will add according to the *square root* of the range of the integration and thus equals $W_B \sqrt{B}$. Because of symmetry, it is found that the resultant vector from the negative frequency side has *exactly the same phase angle* (Fig.B2) and so the amplitude of the total integral is equal to $2W_B \sqrt{B}$. The peak value of the power spectrum of the result (as shown in Fig.B1(b)) is thus the square of this or

$$|Y_{AC}(0)|^2 = 4W_B^2 B \quad (\text{B.7})$$



*Fig. B2. (a) Spectrum $G(\phi)$ of narrow band random noise, showing amplitude $\sqrt{W_B}$ and indicating (random) phase angles θ_i ,
 (b) Reversed and displaced spectrum $G(\Delta f - \phi)$
 (c) Product $G(\phi) \cdot G(\Delta f - \phi)$ which is to be integrated to obtain convolution for $f = \Delta f$*

The continuous spectrum of Fig.B1(b) can be said to represent the AC fluctuations of the squared signal around the DC component which represents the long-term average. The fluctuations can be reduced by low-pass filtering, and Fig.B3 shows the effect of doing this with an averaging network whose lowpass filter bandwidth is $\ll B$. Over a small frequency range in the vicinity of zero frequency the spectrum level can be considered constant, and equal to the value at zero frequency ($4 W_B^2 B$). Thus, the AC power transmitted by an averaging network of bandwidth $1/T_A$ (and thus with effective averaging time T_A , see Appendix A) is obtained by integrating the PSD of the fluctuations over the bandwidth, and is given by

$$\text{AC power} = 4 W_B^2 \frac{B}{T_A} \quad (\text{B.8})$$

The standard deviation of these fluctuations is given by the square root of this, or

$$\sigma_1 = 2 W_B \sqrt{\frac{B}{T_A}} \quad (\text{B.9})$$

where σ_1 is the standard deviation of the power fluctuations (remembering that the input *voltage* to the averaging circuit was in fact proportional to the *power* of the original unsquared signal).

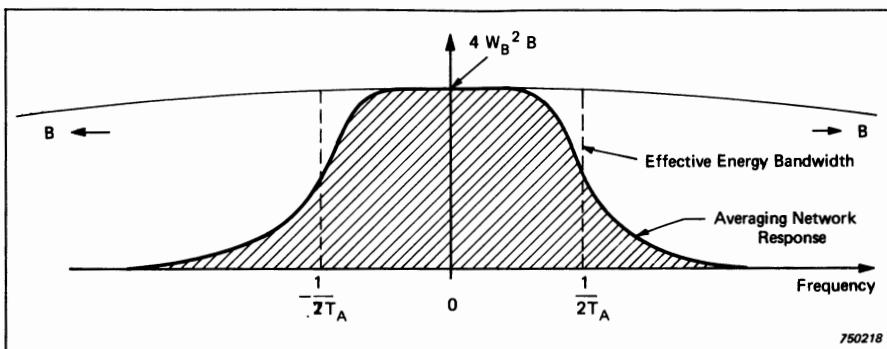


Fig. B3. The low-frequency portion of the squared signal spectrum shown with expanded frequency scale, together with the effect of passing the spectrum through an averaging network

The relative standard deviation of these power fluctuations can be obtained by normalisation with respect to the DC power representing the true result (Eqn.B.5) and thus:

$$\frac{\sigma_1}{Y_{DC}(0)} = \frac{2 W_B \sqrt{\frac{B}{T_A}}}{2 W_B B} = \frac{1}{\sqrt{BT_A}} \quad (\text{B.10})$$

The relative standard deviation of an RMS estimate will be 1/2 of that given by Eqn.(B.10) for a mean square or power estimate (because of taking the square root) and that is the result given in Eqn.(3.9).

APPENDIX C

Response of Physical Systems

One of the most important measurements in dual channel FFT analysis (Chapter 7) is that of frequency response functions. This is because of the simplification which results from formulating the response of linear physical systems in terms of them. Appendix C gives a brief introduction to this formulation. The particular case of mechanical structures will be taken, although a parallel formulation can be made for electrical or acoustical systems.

The structure is assumed to be modelled as an assembly of masses, springs and (viscous) dampers, acted on by a system of forces represented by the vector $\{f(t)\}$, with responses in the same degrees-of-freedom (DOF's) represented by the displacement vector $\{x(t)\}$. In terms of the mass matrix $[M]$, the stiffness matrix $[K]$, and the damping matrix $[C]$, the force balance can be expressed as:

$$[M] \{\ddot{x}(t)\} + [C] \{\dot{x}(t)\} + [K] \{x(t)\} = \{f(t)\} \quad (\text{C.1})$$

Applying the Laplace transform to this system of equations gives:

$$([M] s^2 + [C] s + [K]) \{X(s)\} = \{F(s)\} \quad (\text{C.2})$$

where $X(s)$ is the Laplace transform of $x(t)$ and $F(s)$ the Laplace transform of $f(t)$ in terms of the Laplace variable s . Eqn. (C.2) can be expressed more concisely as:

$$[B] \{X\} = \{F\} \quad (\text{C.3})$$

where $[B]$ is termed the "Impedance Matrix"[‡], and is given by:

$$[B] = [M] s^2 + [C] s + [K] \quad (\text{C.4})$$

[‡] strictly speaking "impedance" refers to a formulation in terms of velocity, while B here is actually "dynamic stiffness", related to displacement.

The elements of $[B]$ can be formulated analytically, but are difficult to measure in practice, because this would require the measurement of forces at all points resulting from the displacement of one point only (while restraining the motion of all other points). The relationship of Eqn. (C.3) can alternatively be expressed in the form:

$$\{X\} = [H] \{F\} \quad (C.5)$$

where $[H]$ ($= [B]^{-1}$) is known as the "Compliance Matrix". The elements of $[H]$ can be measured in practice because they require only the measurement of displacement at all points resulting from the application of a force at one point.

For no externally applied forces ($\{F\} = \{0\}$ in Eqn. (C.3)) the equation for free vibrations results, i.e.:

$$[B] \{X\} = \{0\} \quad (C.6)$$

This constitutes an eigenvalue problem, with the only non-trivial solutions occurring for values of s (eigenvalues) for which $\text{Det } [B] = 0$, and with corresponding arrangements (eigenvectors, or mode shapes) of the displacement vector $\{X\}$.

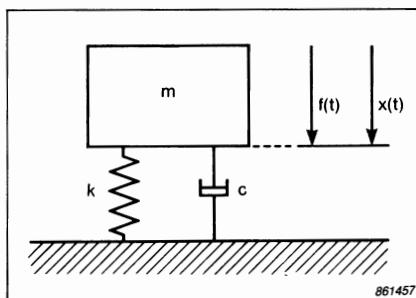


Fig.C1. Single degree-of-freedom system with mass m , spring constant k , and (viscous) damping coefficient c . Applied force $f(t)$ and response $x(t)$ are also shown

For the simple case of a single degree-of-freedom system (Fig.C1) with mass m , damping c , and spring stiffness k , Eqn.C.6 reduces to:

$$(ms^2 + cs + k) X(s) = 0 \quad (C.7)$$

with the two solutions:

$$s_{1,2} = -\frac{c}{2m} \pm j\sqrt{\frac{k}{m} - \left(\frac{c}{2m}\right)^2}$$

$$= -\sigma_p \pm j\omega_p \quad (C.8)$$

where $\sigma_p = c/2m$, ω_p is the damped natural frequency (in rad/s) corresponding to the undamped natural frequency $\omega_0 = \sqrt{\frac{k}{m}}$, and where less than critical damping has been assumed (i.e. $\sigma_p^2 < \omega_0^2$)

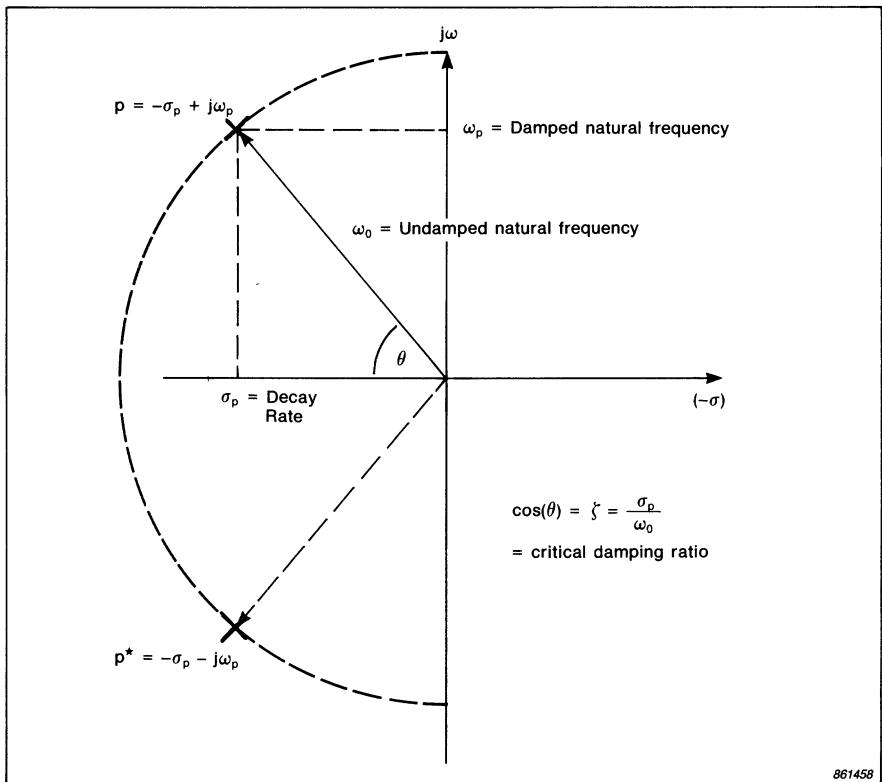


Fig.C2. Figure illustrating the frequency and damping parameters associated with a complex conjugate pair of poles for a single degree-of-freedom system

In this case $H = 1/B = \frac{1}{ms^2 + cs + k}$

$$= \frac{1/m}{(s-s_1)(s-s_2)} \quad (\text{C.9})$$

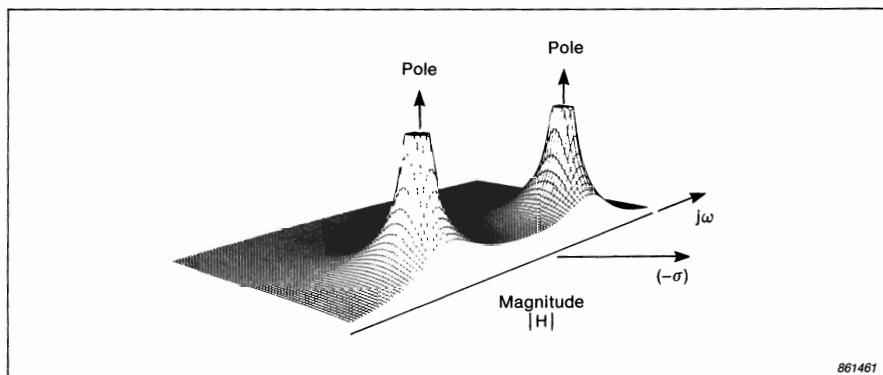
in terms of the "poles" s_1 and s_2 defined by Eqn. (C.8). It will be seen that at the poles ($s = s_1$ or s_2) the value of H goes to infinity.

Eqn. (C.9) can also be expanded in a partial fractions expansion to give the result:

$$H(s) = \frac{r}{s-s_1} + \frac{r^*}{s-s_1^*} \quad (\text{C.10})$$

where r is called the "residue" and in this case equals $\frac{1}{2jm\omega_p}$

Figure C2 shows the location of the poles in the complex s plane, and illustrates the relationship of σ_p , ω_o and ω_p . Fig.C3 (from Ref.7.5) shows a 3-dimensional representation of $|H|$ in the s -plane, highlighting the intersection with the imaginary ($j\omega$) axis. This is recognized as being the amplitude of the **frequency response** function $H(j\omega)$ or $H(f)$ (where $\omega = 2\pi f$) for a single degree-of-freedom system. Figure C4 displays such a frequency response function in terms of its amplitude, phase, real part, imaginary part, and real part vs. imaginary part (Nyquist diagram). Thus, as stated earlier, the **frequency response function** is a restricted part of the **transfer function**, $H(s)$, evaluated along the $j\omega$ axis. Even so, the whole function is defined by the pole location and the residue (see Eqn. (C.10)), and can thus be obtained as follows from the frequency response function: ω_p is the (angular) frequency corresponding to the resonance peak, σ_p is a direct measure of the damping, which can be deduced from the width of the resonance peak (e.g. from the 3 dB bandwidth), while the residue can be determined from the height of the peak.



861461

Fig.C3. 3-dimensional plot of $|H|$ in the s -plane, highlighting the intersection with the $j\omega$ axis

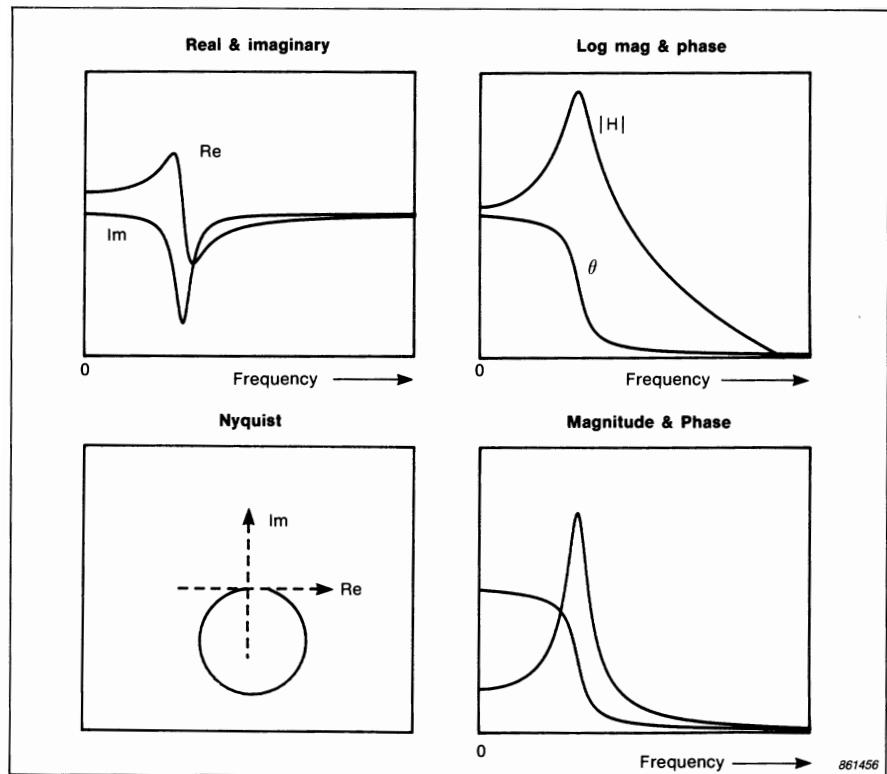


Fig.C4. Various representations of the frequency response function of a single degree-of-freedom system

More generally, for a multi degree-of-freedom system, with n degrees of freedom, the determinant $\text{Det}[B]$ will be a polynomial in s of order $2n$. An element $H_{ij}(s)$ of the compliance matrix $[H(s)]$, indicating the response in DOF i to a force applied in DOF j , can be written in the form:

$$H_{ij}(s) = \frac{a_0 + a_1 s + a_2 s^2 + \dots + a_m s^m}{b_0 + b_1 s + b_2 s^2 + \dots + b_{2n} s^{2n}} \quad (\text{C.11})$$

This can be expanded by a partial fractions expansion to give a sum of terms representing n complex conjugate pairs of poles and residues, each pair of which can be considered as the response of a single degree-of-freedom system (cf. Eqn.(C.10)). Thus:

$$H_{ij}(s) = \sum_{k=1}^n \left[\frac{r_{ijk}}{s-p_k} + \frac{r_{ijk}^*}{s-p_k^*} \right] \quad (\text{C.12})$$

Similarly, $H_{ij}(f)$, the frequency response function can be expressed as the sum of the responses of n single degree-of-freedom systems.

Because of the linearity of the Laplace transform, Eqn.(C.12) can be inverse transformed to give the impulse response corresponding to $H_{ij}(s)$ as a sum of complex exponentials (the impulse response function for the single degree-of-freedom components). Thus:

$$h_{ij}(t) = \sum_{k=1}^n 2 |r_{ijk}| e^{-\sigma_k t} \sin(\omega_k t) \quad , \quad t \geq 0 \quad (\text{C.13})$$

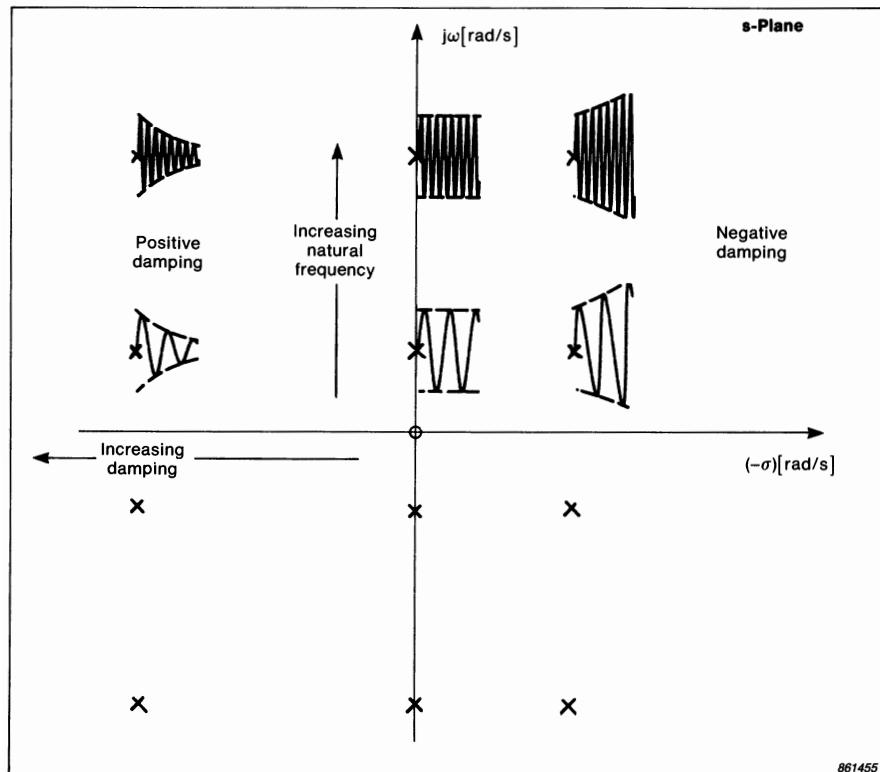


Fig.C5. How the position of a pole (pair) affects the impulse response

861455

The interpretation of the modal parameters σ_k , ω_k and r_{ijk} is the same as for the single DOF parameters σ_p , ω_p and r in the impulse response function of Fig.7.31. Figure C5 illustrates how the position of a pole in the s -plane influences the corresponding impulse response function. This makes it clear that passive linear systems cannot have poles in the right-hand plane, as this corresponds to negative [†] damping and would require a continuous input of energy.

Equation (C.11) can alternatively be factorised in terms of the roots of the polynomials in both numerator and denominator. Thus:

$$H_{ij}(s) = \frac{\prod_{l=1}^m (s - z_l)}{\prod_{k=1}^{2n} (s - p_k)} \quad (\text{C.14})$$

Here, the poles, p_k , are the same as in Eqn.(C.12) but the roots of the numerator, z_l , are known as "zeroes" because the value of the transfer function falls to zero at these points.

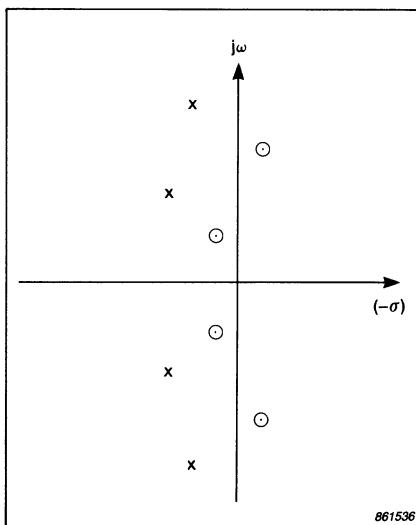


Fig.C6. Representation of a transfer function by the position of poles (x) and zeroes (o) in the Laplace plane

[†] Note that by the definition of damping factor σ (Eqn.(C.8)), positive damping corresponds to a negative exponent of e , and negative damping to a positive exponent.

Fig.C6 illustrates how a transfer function can be represented in terms of the positions in the s -plane of its poles and zeroes. The case illustrated could represent the frequency response of a stable causal system, because even though it has zeroes in the right half plane, it has no poles there. In cepstrum analysis (Chapter 8), one deals with the logarithm of the frequency response function (and hence transfer function) and on taking logs, zeroes become poles. The cepstrum will thus only be causal when there are no poles **or zeroes** in the right half plane for the original transfer function. Functions which have this property are known as “minimum phase” functions, because it can be shown that for a given amplitude characteristic they exhibit minimum phase lag, and the shortest possible impulse response. This can be briefly explained as follows:

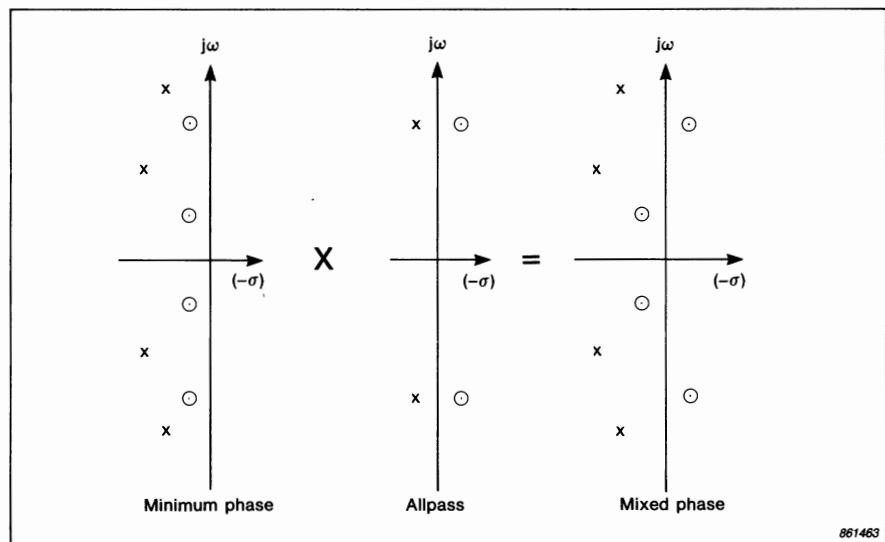


Fig.C7. Representation of a mixed phase function as the product of a minimum phase and an allpass function

A transfer function such as that depicted in Fig.C6 can be considered to be made up as a product of two transfer functions as illustrated in Fig.C7. The first is a minimum phase function (with no zeroes in the right half plane) and the second is a so-called “all-pass” function with pole-zero pairs arranged so that any zero in the right half plane of the final product comes from the all-pass function, while the pole from the latter cancels a zero in the left half plane of the minimum phase function. The equation for one pole-zero pair of an all-pass function is:

$$H_{ap}(s) = \frac{s-a}{s+a^*}$$

and the frequency response function, evaluated along the $j\omega$ axis, is

$$H_{ap}(j\omega) = \frac{j\omega - a}{j\omega + a^*} \quad (C.15)$$

From Fig.C8 it will be seen that the amplitude $|H_{ap}(j\omega)|$ is unity, and this is the reason for the term "all-pass", since (as a filter) all frequency components would be transmitted with unchanged amplitude. In going from small to large values of ω , the phase angle $\angle(j\omega - a)$ changes from near zero to near -180° ($-\pi$), while $\angle(j\omega + a^*)$ changes through $+\pi$, and $\angle\left[\frac{1}{j\omega + a^*}\right]$ through $-\pi$.

Thus, the overall all-pass function gives a phase change of -2π (a phase lag of 2π) for the passage of each pole-zero pair. This confirms that the minimum phase function will have minimum phase lag for a given amplitude characteristic (because each pole-zero pair in the all-pass function gives an additional lag of 2π). It is also evident that the minimum phase function will have the shortest possible impulse response, because the multiplication by each all-pass function in the frequency domain corresponds to a convolution with its impulse response in the time domain. Note that the allpass function has no poles in the right half plane, and is therefore causal and stable.

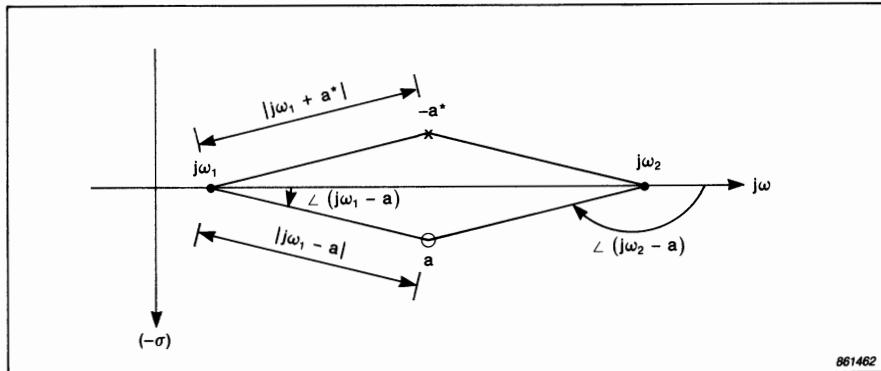


Fig.C8. Illustration of the changes in the frequency response function $\frac{j\omega - a}{j\omega + a^*}$ as ω goes from a small value (ω_1) to a large value (ω_2)

In Chapter 8 it is shown that complex cepstrum analysis is considerably simplified for minimum phase functions, because the cepstrum is causal, and thus the phase spectrum is the Hilbert transform of the log amplitude spectrum and does not have to be separately measured.

In Chapter 8 it is also discussed how the response characteristics of linear systems can be expressed in the cepstrum domain. Because this involves taking the logarithm of the frequency response function, it is better to express the latter as a product of terms (Eqn.(C.14)) rather than as a sum (Eqn.(C.12)). Moreover, it is found to be useful to express the frequency response in terms of the z-transform rather than the Laplace transform for reasons which will become obvious. As mentioned in Section 3.5.1, the z-transform is effectively the equivalent of the Laplace transform when applied to sampled functions. It is shown in Ref.8.17 that the equivalent of Eqn.(C.14) in the z-plane is:

$$H(z) = \frac{|A| \prod_{k=1}^{m_i} (1 - a_k z^{-1}) \prod_{k=1}^{m_o} (1 - b_k z)}{\prod_{k=1}^{p_i} (1 - c_k z^{-1}) \prod_{k=1}^{p_o} (1 - d_k z)} \quad (8.18)$$

where the a_k and b_k represent zeroes, and the c_k and d_k represent poles. The moduli $|a_k|$, $|b_k|$, $|c_k|$ and $|d_k|$ are constrained to be < 1 , and therefore the terms in both the numerator and denominator are divided into two groups, depending on whether the pole or zero is inside or outside the unit circle. It will be seen that the corresponding bracketed term goes to zero when $z = a_k$ or c_k

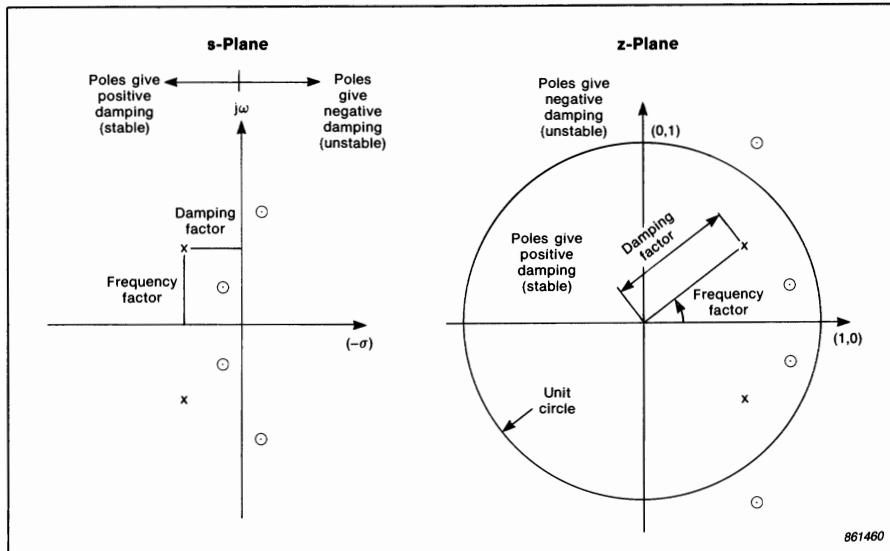


Fig.C9. Equivalence of transfer function characteristics in the s-plane and z-plane

(points inside the unit circle) or $z = 1/b_k$ or $1/d_k$ (points outside the unit circle). Fig.C9 shows how features in the z -plane relate to the equivalent ones in the s -plane. Thus, poles inside the unit circle (equivalent to the left half plane) correspond to positive damping, while poles outside the unit circle (equivalent to the right half plane) correspond to negative damping and cannot be present for stable causal systems.

In Ref.8.17 it is shown how taking logarithms of Eqn.(8.18) changes the products to sums of terms involving $\log(1 - \alpha z^{-1})$. Expanding these as power series gives terms which can be interpreted as the z -transforms of other series which form part of the cepstrum of the response function. For the c_k terms, for example, (the poles within the unit circle), the derivation is as follows:

$$\begin{aligned} \log \left[\frac{1}{\prod_{k=1}^{p_i} (1 - c_k z^{-1})} \right] &= - \sum_{k=1}^{p_i} \log(1 - c_k z^{-1}) \\ &= \sum_{k=1}^{p_i} \left(\sum_{n=1}^{\infty} \frac{c_k^n z^{-n}}{n} \right) = \sum_{n=1}^{\infty} \left(\sum_{k=1}^{p_i} \frac{c_k^n}{n} \right) z^{-n} \end{aligned}$$

which will be seen (Eqn.(3.14)) to be the z -transform of $\sum_{k=1}^{p_i} \frac{c_k^n}{n}$ for $n > 0$.

The final result, for quefrequencies other than zero (where the series cannot be evaluated) is that the cepstrum is given by:

$$C_a(n) = - \sum_{k=1}^{m_i} \frac{a_k^n}{n} + \sum_{k=1}^{p_i} \frac{c_k^n}{n} \quad , n > 0$$

(8.19)

and

$$= \sum_{k=1}^{m_o} \frac{b_k^{-n}}{n} - \sum_{k=1}^{p_o} \frac{d_k^{-n}}{n} \quad , n < 0$$

As mentioned at the end of Section 8.5, each of these terms is a complex exponential (in conjugate pairs a damped sinusoid) further damped by the hyperbolic function $1/n$. This is illustrated by the example of Fig.8.18, where the basic signal rather fortuitously represents the impulse response of a single degree-of-freedom system. Fig.8.18 (e) and (h) are reproduced in Fig.C10, where it can be seen that the cepstrum (Fig.C10(a)), like the impulse response (Fig.C10(b)) is a damped sinusoid with the same frequency. The cepstrum is damped more heavily (by the hyperbolic effect) and there is also a phase shift.

The single degree-of-freedom system has no zeroes and is minimum phase. Thus the phase spectrum is the Hilbert transform of the log amplitude spectrum (Fig.8.18 (f) and (g)) and the cepstrum (like the impulse response) is causal. However, because the log amplitude spectrum is even, the damped sinusoid in the cepstrum is a cosine, whereas that in the impulse response is a sine.

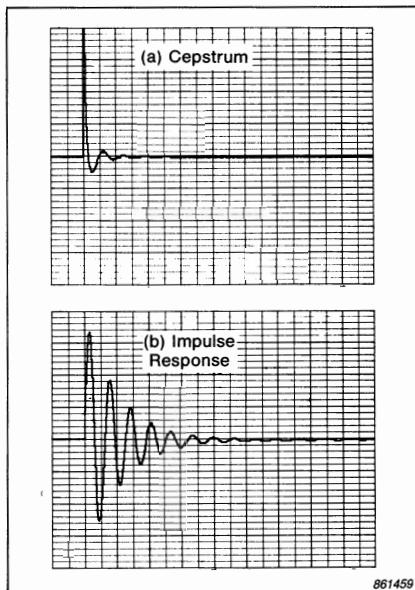


Fig.C10. Comparison of (complex) cepstrum and impulse response for a single degree-of-freedom system

REFERENCES

Chapter 1 — Introduction

- 1.1. A. Papoulis: *The Fourier Integral and its Applications.* McGraw-Hill, 1962.

Chapter 2 — Theoretical Analysis

- 2.1. G.M. Jenkins &
D.G. Watts: *Spectral Analysis and its Applications.* Holden-Day, San Francisco, 1968.
- 2.2. J.S. Bendat &
A.G. Piersol: *Random Data: Analysis and Measurement Procedures.* Wiley — Interscience, 1971.
- 2.3. E.O. Brigham: *The Fast Fourier Transform.* Prentice-Hall, N.J., 1974.
- 2.4. N.Thrane: "The Hilbert Transform", *Brüel & Kjær Technical Review*, No. 3, 1984.
- 2.5. J.T. Broch: "FM Magnetic Tape Recording", *Brüel & Kjær Technical Review*, No. 1, 1967

Chapter 3 — Filter Analysis of Stationary Signals

- 3.1. C.G. Wahrman &
J.T Broch: "On the Averaging Time of RMS Measurements", *B&K Technical Review*, Nos. 2 & 3, 1975.
- 3.2. J.T. Broch &
C.G. Wahrman: "Effective Averaging Time of the Level Recorder Type 2305". *B&K Technical Review*, No. 1, 1961.

- 3.3. A.V. Oppenheim & R.W. Schafer: *Digital Signal Processing*. Prentice-Hall, N.J., 1975.
- 3.4. L.R. Rabiner & B. Gold: *Theory and Application of Digital Signal Processing*. Prentice-Hall, N.J., 1975.

Chapter 4 — Fast Fourier Transform (FFT)

- 4.1. J.W. Cooley & J.W. Tukey: "An Algorithm for the Machine Calculation of Complex Fourier Series", *Math. of Comp.*, Vol. 19, No. 90, pp.297–301, 1965.
- 4.2. "Special Issue on the Fast Fourier Transform", *IEEE Trans. Audio & Electroacoustics*, Vol. AU-15, June 1967.
- 4.3. J.W. Cooley, P.A.W. Lewis & P.D. Welch: "The Fast Fourier Transform Algorithm: Programming Considerations in the Calculation of Sine, Cosine and Laplace Transforms". *J. Sound Vib.* (1970) Vol. 12 No. 3 pp. 315–337.
- 4.4. R.B. Randall & J. Hee: "Cepstrum Analysis", *Brüel & Kjær Technical Review*, No. 3, 1981 (Appendix B)
- 4.5. N. Thrane: "The Discrete Fourier Transform and FFT Analyzers", *Brüel & Kjær Technical Review*, No. 1, 1979
- 4.6. E.O. Brigham: *The Fast Fourier Transform*. Prentice-Hall, N.J., 1974
- 4.7. N. Thrane: "Zoom-FFT", *Brüel & Kjær Technical Review*, No. 2, 1980.
- 4.8. E.A. Sloane: "Comparison of Linearly and Quadratically Modified Spectral Estimates of Gaussian Signals". *IEEE Trans. Audio & Electroacoustics*, Vol. AU-17 No. 2, June 1969, pp. 133–137.
- 4.9. P.D. Welch: "The Use of Fast Transform for the Estimation of Power Spectra: A Method Based on Time Averaging over Short, Modified Periodograms", *IEEE Trans. Audio Electroac.*, Vol. AU-15, No. 2, June 1967, pp.70–73

Chapter 5 — Transient Analysis

- 5.1. N. Thrane: "Zoom-FFT", *Brüel & Kjær Technical Review*, No. 2, 1980.
- 5.2. R.B. Randall & N. Thrane: "Impulse Analysis using a Real-Time Digital Filter Analyzer", *Brüel & Kjær Technical Review*, No. 4, 1979.
- 5.3. J.T. Broch & H.P. Olesen: "On the Frequency Analysis of Mechanical Shocks and Single Impulses". *B & K Technical Review*, No. 3, 1970.
- 5.4. C.G. Wahrman: "Averaging Time of Measurements". *B & K Application Note* (out of print).

Chapter 6 — Non-Stationary Signals

- 6.1. T.A.C.M. Claesen & W.F.G. Mecklenbräuker: "The Wigner Distribution – A Tool for Time-Frequency Signal Analysis", *Philips J. Res.* Vol. 35, Part I, pp. 217–250 (1980), Part II, pp. 276–300 (1980), Part III, pp. 372–389 (1980).
- 6.2. C.P. Janse & A.J.M. Kaizer: "Time-Frequency Distributions of Loudspeakers: The Application of the Wigner Distribution". *J. Audio Eng. Soc.* Vol. 31, No. 4, 1983, pp. 198–222.
- 6.3. H. Herlufsen: "Order Analysis Using Zoom FFT". *Brüel & Kjær Application Note*, No. 012–81.

Chapter 7 — Dual Channel Analysis

- 7.1. H. Herlufsen: "Dual Channel FFT Analysis". *Brüel & Kjær Technical Review*, Part I, No. 1, 1984, Part II, No. 2, 1984.
- 7.2. J.S. Bendat & A.G. Piersol: *Engineering Applications of Correlation and Spectral Analysis*. Wiley, N.Y., 1980.
- 7.3. J.Y. Chung & J. Pope: "Practical measurement of Acoustic Intensity – the two microphone cross spectral method" *Proc. Internoise 1978*, pp. 893–900.

- 7.4. W.G. Halvorsen &
J.S. Bendat:
"Noise Source Identification using Coherent
Output Power Spectra". *Sound and Vibration*,
Vol. 9, p. 15, 1975.
- 7.5. M.H. Richardson:
"Measurement and Analysis of the Dynamics of
Mechanical Structures". Modal 3,0 Manual,
Structural Measurements Systems, CA, USA.
- 7.6. L.D. Mitchell:
"Improved Methods for the FFT calculation of
the Frequency Response Function", *Journal of
Mechanical Design*, April 1982, Vol. 104,
pp. 277-279.
- 7.7. P. Cawley:
"The Reduction Bias Error in Transfer Function
Estimates using FFT Based Analyzers", *MSA –
Session, ASME Conf.*, Dearborn MI. Sept. 1983,
pp. 45-53.
- 7.8. D.J. Ewins:
Modal Testing: Theory and Practice. Research
Studies Press Ltd, Letchworth. Available from
Brüel & Kjær.
- 7.9. D.L. Brown,
R.J. Alleman,
R. Zimmerman
& M. Mergeay:
"Parameter Estimation Techniques for Modal
Analysis". *SAE Paper 790221* from Detroit
meeting Feb./Mar., 1979.

Chapter 8 — Cepstrum Analysis

- 8.1. B.P. Bogert, M.J.R. Healy & J.W. Tukey:
"The Quefrency Alanysis of Time Series for
Echoes: Cepstrum, Pseudo-Autocovariance,
Cross-cepstrum and Saphe Cracking", in *Pro-
ceedings of the Symposium on Time Series
Analysis*, by M. Rosenblatt, (Ed.), Wiley
N.Y. 1963, pp. 209-243.
- 8.2. A.V. Oppenheim, R.W. Schafer & T.G. Stockham:
"Nonlinear Filtering of Multiplied and Con-
volved Signals", *IEEE Trans. Audio & Electro-
acoustics*, Vol. AU-16, No. 3, Sept. 1968.
- 8.3. R.B. Randall & J. Hee:
"Cepstrum Analysis", *Brüel & Kjær Technical
Review*, No. 3, 1981.

- 8.4. J.H. Miles, G.H. Stevens & G.G. Leininger: "Analysis and Correction of Ground Reflection Effects in Measured Narrowband Sound Spectra using Cepstral Techniques". *NASA TM X-71810*, 1975.
- 8.5. A.A. Syed, J.D. Brown, M.J. Oliver & S.A. Hills: "The Cepstrum: A Viable Method for the Removal of Ground Reflections". *J. Sound and Vibration*, Vol.71, No. 2, pp. 299–313, 1980.
- 8.6. J.S. Bolton & E. Gold: "The Application of Cepstral Techniques to the Measurement of Transfer Functions and Acoustical Reflection Coefficients". *J. Sound Vib.*, Vol. 93, No. 2 (1984) pp. 217–233.
- 8.7. A.M. Noll: "Cepstrum Pitch Determination", *J.A.S.A.* Vol. 41, No. 2, 1967, pp. 293–309.
- 8.8. R.W. Schafer & L.R. Rabiner: "Digital Representations of Speech Signals". *Proc. IEEE* Vol. 63, pp. 662–677, 1975.
- 8.9. N. Thrane: "Application of a Long Memory FFT Analyzer in Speech Analysis". Paper presented at the Fourth F.A.S.E. Symposium on Acoustics and Speech, Venezia, Italy, 1981. Available as Brüel & Kjær Application Note No. 066–81.
- 8.10. J.E. Luck: "Automatic Speaker Verification using Cepstral Measurements". *J.A.S.A.* Vol. 46, pp. 1026–1032, 1969.
- 8.11. P. Bradshaw & R.B. Randall: "Early Detection and Diagnosis of Machine Faults on the Trans Alaska Pipeline", *MSA-Session, ASME Conf.*, Dearborn MI, Sept. 1983, pp. 7–17.
- 8.12. G. Sapy: "Une application du traitement numérique des signaux au diagnostic vibratoire de panne: La détection des ruptures d'aubes mobiles de turbines". *Automatisme - Tome XX*, No. 10, pp. 392—399, October 1975.
- 8.13. R.B. Randall: "Machine Health Monitoring by Analysis of Accelerometer-Derived Signals". *Proc. FASE 84*, Sandefjord, August 1984, pp. 13–29.

- 8.14. R.B. Randall: "Cepstrum Analysis and Gearbox Fault Diagnosis." Brüel & Kjær Application Note No. 233-80. Reprinted in *Maintenance Management Int.*, Vol. 3, 1982/1983, pp. 183-208.
- 8.15. J.M. Tribolet: "A New Phase-Unwrapping Algorithm." *IEEE Trans. Acoust., Speech, Signal Process.*, Vol. ASSP-25, pp. 170-177 (1977).
- 8.16. A. Papoulis: *The Fourier Integral and its Applications*. McGraw-Hill, 1962.
- 8.17. A.V. Oppenheim & R.W. Schafer: *Digital Signal Processing*. Prentice-Hall, N.J. 1975.
- 8.18. R.W. Schafer & L.R. Rabiner: "System for Automatic Formant Analysis of Voiced Speech." *J.A.S.A.* Vol. 47, pp. 634-648, 1970.
- 8.19. A.V. Oppenheim: "Speech Analysis-Synthesis System Based on Homomorphic Filtering." *J.A.S.A.* Vol. 45, pp. 458-465, 1969.
- 8.20. W.D. Mark: "Gear Noise Excitation", from *Engine Noise: Excitation, Vibration and Radiation*. (Eds. R. Hickling & M.M. Kamal). Plenum Publishing Corp. N.Y., 1982.
- 8.21. R.G. De Jong & J.E. Manning: "Gear Noise Analysis using Modern Signal Processing and Numerical Modeling Techniques." *SAE Paper 840478* from Int. Congress and Expo. Detroit, 1984.
- 8.22. R.B. Randall: "Separating Excitation and Structural Response Effects in Gearboxes." Paper C305/84. I Mech. E. Conf. on Vib. in Rotating Machines, York, 1984, pp. 101-107.
- 8.23. R.H. Lyon & A. Ordubadi: "Use of Cepstra in Acoustical Signal Analysis." *J. Mech. Des.*, Vol. 104, Apr. 1982, pp. 303-306.

INDEX

- AC recording 101, 127, 129, 138, 140, 141
A/D Converter 117
Aliasing 12, 13, 30, 117, 122, 153, 156, 170, 221
All-pass function 298, 325
Amplitude demodulation 178
Amplitude modulated signal 70
Amplitude modulation 178
Amplitude spectrum 139, 142
Analog analysis 80
Analog filters 3, 78, 192
Analog filters and detectors 12
Analog-to-digital (A/D) converter 79, 109, 169, 219
Analytic signal 3, 11, 62, 70, 178, 184, 274
Anti-aliasing filter 123, 221
Auto-correlation function 13, 228, 253, 256, 262, 271, 279
Autocorrelation coefficient function 257
Autospectrum 228, 236, 252, 253
Averaging 90, 120, 174
Averaging detectors 88
Averaging time 88, 91, 101, 110, 120, 126, 133, 134, 140, 141, 144, 194,
..... 199, 201

Ball-bearings (see also "Rolling element bearings") 273, 288
Band energy level 13
Band-limited white noise 256
Bandpass filtered white noise 256
Bandwidth 33, 83, 115, 173, 186, 307, 309
Bandwidth compensation 132, 133, 142
Beat frequency 42, 126
Beat or modulating frequency 144
Bessel functions 72
Bias error 229, 242, 244
Bit-reversed address 148, 151
Bowtie correction 255
Butterfly operation 151

Calibration	139, 173, 199
Campbell diagram	216
Carrier frequency	70, 178
Cascade plot	215
Causal signal	58, 329
Cepstrum	259
Cepstrum analysis	4, 13, 271
Choice of bandwidth	132
Circular correlation effect	254
Coherence (function)	13, 147, 229 230, 240, 244, 247
Coherent Output Power	234, 242
Coincident (or co-) spectrum	228
Complex cepstrum	13, 273, 274, 297, 299, 326
Complex conjugate	18
Complex exponential function	265, 328
Complex FFT transform	168
Complex logarithm	274
Compliance matrix	319
Conjugate even function	23, 60, 65, 275, 313
Conjugate odd function	65
Convolution	12, 39; 49, 51, 91, 107, 155, 158, 192, 239, 253, 275, 279, 312, 326
Convolution of Spectral Functions	52
Convolution Theorem	12, 39, 53, 61, 70, 153, 158, 239, 253, 275, 310
Correlation coefficient	230
Correlation functions	147
Crest factor	88
Cross correlation function	13, 173, 233, 252, 259, 269
Cross spectrum	13, 228, 240, 253, 255, 269
Curve fitting	265
Cyclic signals	13, 213
D/A converter	109
Damped natural frequency	250, 320
dB peak-to-peak	296
DC recording	103, 127, 128, 137, 140, 141
Deconvolution	14, 276, 279, 303
Delay time	269, 274
Delta functions	19, 32, 50, 51, 53, 91, 263, 275, 314
Demodulated signals	13
Detector	79, 87, 105, 106, 128, 135, 199
Detector response time	99
Deterministic signal	32, 84, 126, 130, 133, 138, 139, 140, 144
DFT (Discrete Fourier Transform)	12, 153, 169, 274
Diesel engine signals	211, 213, 304
Digital analysis	80

Digital averaging	120
Digital filter analysis	109, 143, 202
Digital filters	3, 78, 167, 185
Digital filters (and detectors)	12
Digital zoom	262
Discrete Fourier Transform (DFT)	12, 28, 146
Discrete stepped filter analyzers	104
Dispersive systems	262
Dual channel analysis	227, 239
Dual channel FFT analysis	4, 318
Dwell time	134, 140, 199, 201
Dynamic range	103, 115, 117, 127, 156, 189, 195
Echo detection and removal	279, 282, 299
Echoes and reflections	14, 259, 275, 283
Effective bandwidth (of transient)	198
Effective length (of transient)	198
Effective noise bandwidth	33, 38
Effective record length	190
Eigenvalue problem	319
Eigenvectors	319
Energy spectral density (ESD)	13, 32, 47, 48, 185, 186, 188, 196, 197, 199, 202
Envelope	69, 70, 257, 259
Envelope analysis	181
ESD scaling	190
Euler's relation	17
Even function	60
Exponential averaging	92, 94, 109, 115, 120, 143, 193, 203, 213
Exponential window	187, 302
Fast Fourier Transform (FFT)	3, 12, 28, 78, 115, 119, 123, 130, 131, 146, 186, 192, 197, 291
FFT algorithm	147
FFT analyzer	109, 185, 213
Filter characteristic	36, 55, 85, 119, 122, 128, 133, 135, 138, 156, 157, 159, 222, 292
Filter coefficients	115, 117, 119, 120
Filter impulse response	78, 192, 194, 199
Filter response time	85, 105, 134, 141, 198
Filters	79, 81
Finite element analysis	252
Flat-top weighting	159, 225
Forced response simulation	252
Formant	285, 303

Fourier analysis	11, 19, 71, 91, 305
Fourier integral	167
Fourier integral transform	26, 112
Fourier series	20, 26
Fourier transform	12, 19, 25, 49, 53, 58, 112, 239, 306, 308
Fourier transform pair	26
Frequency modulation	3, 11, 71
Frequency response characteristics	101
Frequency response function	13, 53, 94, 147, 230, 239, 250, 263, 318, 321, 327
Gamplitude	271
Gaussian function	159
Gaussian random signals	44, 145, 229
Gear diagnostics	289
Gear vibrations	14, 289, 303
H ₁ and H ₂	240, 245
Half cosine pulse	45
Hamming Function	158
Hanning weighting window	48, 157, 161, 175, 188, 211, 212, 214, 224, 245, 292, 301
Harmonic pattern	81
Harmonics	14, 21, 41, 130, 179, 182, 236, 273, 275, 284, 288, 303
Heterodyne filter	109
Heterodyne techniques	104
High speed analysis	3
Hilbert transform	3, 11, 58
Hilbert transform techniques	177
Homomorphic signal processing	273
Ideal filter	33, 83, 202, 312
Impedance matrix	318
Impulse response	12, 50, 53, 58, 69, 92, 112, 233, 239, 284, 299, 323, 326, 328
Impulse response functions	13, 263
Impulse train	53, 95, 154
Initial phase angle	19, 54
Input autospectrum	240
Instantaneous frequency	66, 70
Instantaneous phase angle	70
Kaiser-Bessel window	159
Kinetic energy	67

Laplace plane	239
Laplace transform	112, 250, 298, 318, 323, 327
Leakage	12, 155, 232, 245
Level recorder	99, 101, 105, 127
Lifter	271
Lifted spectrum	271
Line number	173
Linear averaging	94, 115, 120, 143, 193
LMS (logarithmic mean square) detector	89, 99
Logarithmic potentiometer	99
Long-pass lifter	271
Loudspeaker characteristics	301
Lowpass filter	109, 117, 156, 222, 303
Lowpass filter characteristic	101, 123, 221
Lowpass filtration	91
 Machine diagnostics	288, 303
Machine vibration signals	13
Magnitude of the analytic signal	266
Matched z-transform	115, 122
Maximum phase	299
Minimum phase characteristic	302
Minimum phase functions	297, 325
Modal analysis	250, 265
Mode shape	67, 251, 319
Modulated signals	70
Modulation	70, 293
Modulation index	72
Multi degree-of-freedom system	265, 322
Multi-channel measurements	147
Multiple channel FFT analysis	13
Multiple coherence	236
 Negative damping	324
Negative frequency	11, 19, 30, 54, 61, 62, 112, 168
Noise bandwidth	83, 95, 159
Non-coherent power	236
Non-destructive zoom	169
Non-linearity	232, 242
Non-stationary signals	4, 48, 80, 177, 185, 211
Nyquist (polar) display, plot	67, 250, 321
Nyquist frequency	30, 112, 122, 153, 156, 167, 172, 221, 222
 1/12-octave analysis	120, 195
1/3-octave analysis	87, 115, 117, 125, 131, 142, 192, 195, 202
1/8-octave spectrum	105

Octave band filtering	119, 130
Octave filters	83
Octave Selectivity	85
Octave-band analysis	130
Odd function	60
One-sided spectrum	65
Order (tracking) analysis	4, 13, 219
Output auto-spectrum	240
Paper speed	127, 137
Parallel analyzers	107
Parseval's theorem	25, 47, 201
Partial coherence	236
Pattern recognition	286
Peak detector	200
Peak detection	206
Periodic function (signal)	20, 41, 197
Phase cepstrum	297
Phase demodulation	182
Phase deviation	72, 183
Phase modulation	3, 11, 71, 178
Phase spectrum, spectra	65, 241
Phase unwrapping	297, 302
Picket fence corrections	173, 225
Picket fence effect	12, 161, 245
Planetary gears	293
Poles	250, 298, 321, 324, 327
Potential energy	67
Power cepstrum	13, 228, 273, 282, 291, 297
Power spectral density (PSD)	19, 32, 42, 142, 173, 314
Power spectrum, spectra	25, 32, 43, 54, 78, 173, 228, 273, 277
Practical filter characteristic	83
Pre-computational delay	233
Probability density	43, 45, 97
Pseudo-random signals	44, 265
Q-factor	43, 81, 115
Quadrature (or quad-) spectrum	228
Quasi-periodic signals	41, 258
Querfency	271, 328
Rahmonics	271
Random error	229
Random noise	95
Random signal	83, 126, 133, 136, 138, 140, 142
RC averaging (integration)	92, 99

RC circuit	92, 95, 98
RC time constant	94, 111
Real-time	115, 217
Real-time analysis	107, 109, 176
Real-time analyzers	107
Real-time frequency	176
Real-time zoom	166
Reciprocating machine vibrations	13, 179, 211, 212
Rectangular pulse	45
Rectangular weighting (window)	48, 54, 157, 186, 187, 255
Recursive digital filtering	79
Resampling	168
Residue	250, 321
Resolution	173, 174, 229, 245
Reverberation time	233, 266
Ripple	92, 94, 96, 112, 141, 175
Rise time	85
RMS detector	200
RMS spectrum	185
Rolling element bearings	179, 289, 293
Root Mean Square (RMS) amplitude, value	25, 88, 99, 110
Rotating vectors	18
Run-ups or run-downs	13, 211
Running average	91
Running linear averaging, integration	92, 99, 193
Sample rate	117
Sampling frequency	26, 30, 109, 110, 112, 119, 165, 219, 221
Sampling ratio	122
Saphe	271
Scaling	173, 184, 185, 196, 296
Seismology	283
Selectivity	83, 84, 212
Sequential filter analysis	197
Sequential or serial analyzers	107
(Shannon's) sampling theorem	30, 110, 118, 120, 167
Shape factor	84, 124, 133, 136, 138
Short-pass lifter, filtering	271, 286, 301
Short-term peak value	201
Short-term RMS value, detection	207
Sidebands	14, 70, 73, 77, 131, 182, 275, 288
Sidelobes	155, 157
Sign function	58
Signal-to-noise ratio	187, 236, 242
Single-degree-of-freedom system	69, 179, 250, 265, 299, 319, 322, 328

Sound intensity	230
Spectral map (3-dimensional)	13, 215
Speech analysis	13, 49, 211, 212, 284, 303
Square root extraction	88, 97
Standard deviation	44, 96, 115, 126, 145, 316
Stationary deterministic signals	41, 173
Stationary random functions, signals	39, 42, 95, 144, 173, 174, 245, 312
Stationary signals	80, 125, 252, 253
Sweeping filter analyzers	106
Swept filter analysis	130
 Tape loop	32, 107, 198
Tape recorder	109
TDS (Time Delay Spectrometry)	301
Third octave filters (see $1/3$ octave analysis)	83
3dB bandwidth	83, 321
3dB-dimensional representation	286
Time compression	3, 130, 131
Time compression analyzers	109
Time constant	268
Time delay	233, 259
Time window, windowing	4, 13, 39, 48, 157, 188
Time-Series	26
Tone burst	32, 36, 45, 200, 202, 203
Toothmesh frequency, frequencies	14, 289, 303
Torsional vibrations	182
Tracking (see also "Order analysis")	184
Tracking lowpass filter	224
Transfer function	112, 239, 250, 298, 321, 324
Transient signals	13, 45, 97, 252, 254
Transients	4, 32, 80, 185, 253
True RMS detectors	88, 104
Truncated Gaussian window	159
 Undamped natural frequency	320
Undersampling	170
Underwater acoustics	283
Unit circle	112, 298, 328
 Voice pitch	285
Wahrman detector	88, 89, 99, 203
Waterfall (or cascade) plot	215
White noise (signals)	33, 233
Whole body vibration	14, 131
Wiener-Khinchin relationship	253

Wigner distribution	211
Window function	232
Writing speed	103, 134, 136, 138, 140, 141
Z-transform	112, 298, 327
Zero padding	173, 254
Zeroes	298, 324, 327
Zoom factors	172
Zoom FFT (analysis)	12, 165, 172, 178, 188, 247, 293
Zoom processor	166, 178
Zoom tracking	222