Mathematical & Computational Applications, Vol. 4, No. 2, pp. 121-136, 1999.
© Association for Scientific Research

ADVANCED SIGNAL PROCESSING TECHNIQUES FOR FAULT DIAGNOSTICS – A REVIEW

Şeref Naci Engin, Kayhan Gülez

Yıldız Technical University
Electrical & Electronics Faculty
Electrical Engineering Department
80750 Beşiktaş-İstanbul, TURKEY

M.N.M.Badi

University of Hertfordshire
Faculty of Engineering & Information Sc.
Department of Aerospace, Mech. & Civil Eng.
AL10 9AB Hatfield, Herts, UK

Abstract-An overview of major signal processing and analysis methods finding an area of application in machinery fault diagnostics is presented. Among them Wavelet Transform, which has been gaining a special interest in many research and engineering fields, is emphasized with an example application based on numerically obtained vibration-time data representing various health conditions of a spur gear.

1. INTRODUCTION

Data collection and data interpretation are two main aspects in Condition Monitoring (CM) or more specifically in Fault Diagnostics (FD). A majority of the work in the field to date has been focused on the collection of relevant data. However, the capabilities to interpret the data and to determine the condition of a system being monitored is far behind the data acquisition capabilities, and hence recently has been attracting more interests by researchers, Staszewski 1996 [20], Engin et al. 1996, 1997 [7,8]. Dealing with data in complex structure and huge amount, and the need for fast or automated data interpretation makes the subject even more important and challenging.

Experiences with machinery condition monitoring data indicate that machines do not break down without some form of prior warning. An amount of deterioration in the vibration and/or noise pattern usually takes place. Such changes compared to machine's normal or 'healthy' state have been used as an indication of an impending or a developing fault. When a noise or vibration signal is captured for monitoring, it can be difficult to analyze it in its raw state. This is because it is usually submerged by other signals from adjacent components, or the variation in the signal is too small to be detected. In this case special signal processing and analysis techniques have to be utilized to accentuate the changes and to extract more information in order to make correct decisions on the machine's health.

From basic, e.g. statistical and spectral ones, to relatively more advanced, e.g. time-frequency and time-scale analyses, there have been many signal processing methods encountered in the area. The criteria for a successful method is whether it explains and sheds light on the physical phenomena and makes new predictions, Cohen 1995 [5]. On that account, in vibration/noise data analysis the kind of signal analysis method to be used should depend on the complexity of the problem, e.g., if the characteristics of the changes has a spiky nature *kurtosis* as a statistical signal analysis tool can be very useful to identify the fault, Badi *et al.* 1996 [1]. But if there are more than one fault in the system or if the changes are too sudden or too small, other more sophisticated techniques should be investigated.

The signal analysis techniques are in general divided into two groups, the time-domain analysis and the frequency-domain analysis (called spectral analysis), according to which domain the signal is processed. A third one called *quefrency*-domain is studied sometimes amongst the frequency-domain methods, Staszewski 1996 [20], Hunt 1996 [12], since it is a kind of double spectrum or spectrum of spectrum; and sometimes as independent, Rao S.S 1995 [18]. Here, from the point of presentation convenience, the latter is followed. There are also more sophisticated signal analysis methods, which have been finding area of application in CM recently. These are generally called *Time-Frequency* and *Time-Scale* (TFTS) signal processing methods, which analyze the signals by transforming them into time-frequency or time-scale domains.

This paper presents a brief review of CM vibration signal analysis methods, which utilize various types of signal processing techniques and especially the current state of the art in the field. It also discusses some results of an example rotating machinery fault diagnostics application.

2. TIME-DOMAIN METHODS

The objectives of time domain analyses are to determine the *statistical characteristics* of the original function by manipulating the series of discrete numbers. Therefore, they employ some statistical methods to help to describe the spiky nature of the time signals. That is to *assume* that the process is *random* and exhibits a degree of statistical regularity. With this assumption various statistical methods have been developed to compute the probability of failure.

In a random process, since the motion is assumed as random, the precise value of x (amplitude of the motion) at any chosen time $t=t_0$ cannot be precisely predicted. The statistical properties of a random signal are therefore described in terms of probability density function $p(x)$ and this function expresses the probability (a value between zero and one) of x taking a value between x and $x+\Delta x$ (where Δx is a vanishingly small increment). The probability must sum to unity over all possible x values. This is formulated for analog and discrete signals with the following couple of equation,

$$\int_{-\infty}^{\infty} p(x)d(x) = 1\,, \quad \sum_{t=0}^{\infty} p(x_t) = 1 \qquad (1)$$

The probability density function, which is usually assumed to be most applicable is the *Gaussian* or *'Normal' distribution* given by

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

where $\bar{x}$ is the arithmetic mean value (indicates center of the distribution) and σ is standard deviation (indicates spread of distribution) and defined by

$$\sigma = \sqrt{\overline{x^2} - \left(\bar{x}\right)^2} \qquad (3)$$

Figure 1 shows, with parameters normalized, how this function varies for different standard deviations.

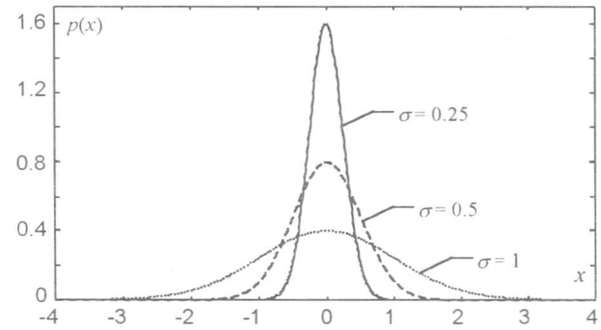

Figure 1. Gaussian or 'normal' probability density functions obtained for various σ (standard deviation) values.

As seen standard deviation (σ) is a measure of the preciseness of the data set: the smaller the σ, the less the spread of values.

2.1 Moments of Probability Density Function

The moments of probability density function $p(x)$, Eq.1., give indication of general features of the distribution of variable x. The moment of order n is the average of x^n. Several significant statistical signal analysis tools, which describe the behavior of x are constructed by the first few moments. The first moment ($n = 1$) gives the average value of x, $\bar{x}$ (or μ_x).

$$\bar{x} = \int_{-\infty}^{\infty} x\, p(x)dx \qquad (4).$$

Likewise the mean-square value is determined from the second moment ($n = 2$)

$$\overline{x^2} = \int_{-\infty}^{\infty} x^2\, p(x)dx$$

The *variance* of x (referred as γ_x, or alternatively σ_x^2, for convenience), is a good measure of concentration and defined as the mean-square value about the average μ_x.

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x-\mu_x)^2\, p(x)dx = \overline{x^2} - \mu_x^2$$

Consequently, the standard deviation is the positive square root of the variance, $\sigma_x = \sqrt{\gamma_x}$ (for hand computations see Eq. 11). When the mean value μ_x is zero, and the standard deviation is equal to the root mean square (rms.) value. If the variance (σ_x^2) and mean value (μ_x) of the signal x are known, then the rms. value of x can be calculated as

$$x_{rms} = \sqrt{\frac{1}{n}\left(\sigma_x^{\,2} + \sum_{i=1}^{n} x_i^{\,2}\right)} \text{ or simply } x_{rms} = \sqrt{\sigma_x^2 + \mu_x^2} \qquad (7).$$

2.1.1 Crest and Form Factors

The Crest Factor is the ratio of the peak to peak (or just peak in some literature) value of a signal to its rms. value. It is an immediate extension to overall level monitoring,

$$CF = x_{pp} / x_{rms} \qquad (8).$$

Therefore the Crest Factor is much more sensitive than the rms. value to the changes in the spiky nature of a vibration signal. Moreover, it is much less likely to give false alarms than using the peak to peak value on its own, since it does a kind of normalization (by means of rms.) that takes changes into account for the overall vibration level. The results of crest factor computations for vibration signals representing healthy and faulty conditions showed that the more peaks the signal has, the higher is the crest factor.

There is another factor, which gives some indication of the wave form of a complex vibration, called *form factor*. It is defined simply as the ratio of the mean value of a signal to its rms. value, μ_x/x_{rms}. In some literature as contrary to the definitions above, these two factors are defined as the ratio of the rms. value to the peak (or peak to peak) value and the rms. value to the mean value, respectively. As the names and definitions suggest, *crest factor* gives rather an indication of peak, hence, it is more sensitive to sudden peaks, whereas *form factor* gives an indication of waveform of the signal.

2.1.2 Kurtosis

The first two statistical moments of a probability density distribution are the mean value and the mean-square value, which were both outlined above. The third statistical moment is the *skewness* of a distribution. It is a measure of the symmetry of the probability density function and is particularly useful for the statistical analysis of many separate experimental results and correlating past results with future outcomes. In the analysis of time series, skewness characterizes the degree of asymmetry of a distribution around its mean.

The fourth statistical moment called *kurtosis* (K) is widely used in machinery diagnostics. It characterizes the relative spikiness or 'flatness' of a time signal compared to its normal state, Eq.9. Positive kurtosis indicates a relatively peaked distribution and negative kurtosis indicates a relatively flat distribution.

$$K = \int_{-\infty}^{+\infty} \frac{(x-\mu_x)^4 p(x)}{\sigma^4} dx \qquad (9).$$

Where as usual, σ is the standard deviation of the signal, $p(x)$ is the probability density function of x, and μ_x (or $\bar{x}$) is the average value of x. The kurtosis is calculated for a set of data or a sampled vibration signal with the following, Eq.10,

$$K = \frac{1}{n\sigma^4} \sum_{i=1}^{n} (x_i - \bar{x})^4 \qquad (10)$$

where n is the number of data and x_i is the ith term in the data series. As it is known standard deviation is a measure of how widely values are dispersed from the average value (mean) and is calculated for a set of data by Eq.11,

$$\sigma = \sqrt{\frac{n\sum x_i^2 - (\sum x_i)^2}{n(n-1)}} \text{ or } \sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (11).$$

The time-domain statistical signal processing methods used in the initial stages of this research are mean, rms., crest factor and kurtosis for doing comparisons between *good* and *faulty* (or *healthy* and *unhealthy*) states, Engin 1998 [9]. These methods have the advantage of producing a single number that is a relative measure of spikiness of an oscillating time series. The numbers computed for vibration signals collected over a period of time can be treated as an indication of the condition of the component and can be plotted against time. The plots can be used to detect any trend i.e. early signs of incipient failures. However it is still difficult to interpret these plots very efficiently even for the signals of dominant faults. Too many samples are needed to make more sound and reliable comments on the results of statistical methods.

It is often necessary to quantify the degree of relation, similarity or interdependence of one set of data upon another. There are several methods or operations to do this, e.g., cross-correlation, covariance and convolution, and they play a prime role in both digital and analog signal processing in establishing further effective signal analysis methodologies. Refer to Engin 1998 [9] for details.

3. FREQUENCY-DOMAIN METHODS

There are generally three main reasons for spectral or frequency analysis. The first is that it simplifies the understanding of the waveform. Secondly, physical properties of the signal, for example its propagation through a medium often depends on the frequency and thirdly it is a mathematical tool for solving equations, Cohen 1992 [4]. The first two reasons make spectral analysis a very significant tool for condition monitoring signal analysis, which is of interest to this study. The third reason is that convolution's presenting a computational convenience: multiplication in the time-domain corresponds to convolution in the frequency-domain and vice versa. Any time series equation can be solved in the frequency-domain conveniently, then the result can be transformed back to time-domain by utilising forward and inverse Fourier transforms.

An overall vibration time signal is usually dominated by a few major frequency components such as machine rotation speed and gear tooth meshing frequencies. These dominant frequencies are important to monitor. However, they will mask frequency components with smaller amplitudes, which are produced by other rotating elements that must be considered. In general this problem is solved by decoupling the vibration time series into all its frequency components so that the components with smaller amplitudes can be observed more clearly. The spectra obtained for low to high frequencies give much earlier warnings of faults produced by frequency components with smaller amplitudes. This is because changes at individual frequencies will be monitored more carefully before the effects become large enough to be detected in the overall vibration time series. Spectrum or frequency-domain analysis also enables fault diagnosis since the frequencies at which changes detected are connected with the rotating machine components that deteriorate. Different changes in the main frequency components and their side bands can be associated with various faults by experience. It can be concluded that even in the case when satisfactory fault detection can be achieved by one of the time-domain methods, spectrum analysis is still often required for detailed fault diagnosis.

4. QUEFRENCY-DOMAIN METHODS

Quefrency domain or cepstral analysis, which is an extension of spectral analysis, is another widely used signal analysis tool in condition monitoring. The *power cepstrum*, first defined and introduced as far back as 1963 by Bogert *et al.* (in Ref. [16]), is originally defined as "the power spectrum of the logarithm of the power spectrum", formulated in Eq.12. Several other definitions can be found for the term cepstrum in the literature such as "inverse Fourier transform of the logarithm of the power spectrum", or "amplitude spectrum of the logarithmic spectrum" Randall 1985 [16], Rao S.S. 1995 [18]. However, this is not critical since they all detect and display distinct peaks in the same location if there is strong periodicity in the spectrum. The name "*cepstrum*" is derived by paraphrasing 'spectrum', the reason being that the cepstrum is obtained by performing a further spectrum analysis on a frequency-spectrum. Several other terms such as "*quefrency*", "*saphe*", "*rahmonics*" and "*gamnitude*" are derived similarly from 'frequency', 'phase', 'harmonics' and 'magnitude' by reversal of some letters.

Using the terminology $F\{\}$ to indicate forward Fourier transform and $F^{-1}\{\}$ inverse Fourier transform of the bracketed function, the original definition of the (power) cepstrum is

$$c_p(\tau)=|F\{\log F_{xx}(x)\}|^2 \quad (12),$$

where $F_{xx}(x)$ is the power spectrum of the time signal $x(t)$ and given by

$$F_{xx}(x)=|F\{x(t)\}|^2 \quad (13).$$

This definition of cepstrum was not found very clear when it was compared with *autocorrelation* function, which was defined as "the inverse Fourier transform of the power spectrum", given below in the Eq.14.

$$R_{xx}(\tau)=F^{-1}\{F_{xx}(x)\} \quad (14).$$

Here, the principal difference with respect to the auto-correlation function is that for the cepstrum the first spectrum is logarithmically converted, see Eq.12.

However, the cepstrum was later defined as "the inverse Fourier transform of the logarithm of the power spectrum", thus making its connection with the autocorrelation clearer. In this case the power cepstrum, $c_p(\tau)$, becomes

$$c_p(\tau)=F^{-1}\{\log F_{xx}(x)\} \quad (15).$$

However, as just mentioned above, in practice the choice of the definition of cepstrum is not critical since they both (Eq.12 and 15) detect and display distinct peaks in the same location if there is *strong periodicity* in the (logarithmic) spectrum.

The other type of cepstrum known as the '*complex cepstrum*', is described by Oppenheim et al. in 1968, Randall 1985b [16]. It was defined as "the inverse Fourier transform of the complex logarithm of the complex spectrum" and was called the '*complex cepstrum*', given by the equation,

$$c_c(\tau)=F^{-1}\{\log F_x(x)\} \quad (16),$$

where $F_x(x)$ is the complex spectrum of $x(t)$.

Despite its name it is a real valued function, but the name indicates that, unlike the power cepstrum, the complex cepstrum is obtained from the complex spectrum, with no loss of phase information. For this reason the process by which it is obtained is reversible, and it is thus possible to return to the original signal after performing filtering operations. This may be useful and desirable in some applications where removal of convolved and multiplied effects by linear filtering techniques is needed, Randall 1977 [17]. Cepstrum is still in use in various signal and image processing applications effectively as a feature extraction tool.

5. TIME-FREQUENCY/SCALE DOMAIN

It has been quite difficult to satisfactorily handle signals carrying nonstationary or transient components, such as small or short lasting vibrations emerging as chirping sounds, using conceptualizations based on stationarity. The production of particular frequencies depends on physical parameters, which may change in time possibly due to an incipient failure originating from various causes. Therefore, examining local behavior of the vibration signal with reasonably precise frequency information may be useful to interpret the signal in a better way. The aim of time-frequency/time-scale (TF/TS) signal analysis is to describe how the frequency or spectral content of a signal evolves and to develop the physical and mathematical ideas needed to understand what a time-varying spectra is, Cohen 1995 [6]. The attempt to represent a signal simultaneously in time and frequency –or a frequency related scale– is full of challenges both physical and mathematical, and there has been an extensive multidisciplinary research on this issue. This section study the basic theory on such methods, which have been found an area of application to CM of rotating machinery.

5.1 Short time Fourier transform (STFT)

Fourier transform gives frequency information which is extracted for the complete duration of a signal $x(t)$. As known, since the integral in Fourier transform, extends from $-\infty$ to $+\infty$, which covers all time the signal lasts, the information obtained is on the average frequency content over the whole length of the signal. If there is an oscillation, which may be due to a fault, somewhere in the duration of the signal, it will change the resultant values of Fourier transform $X(\omega)$, but its location on the time axis will be lost. In other words, the Fourier transform implies that time information is not needed after the transform is applied. There is no attention to *when* the signal components of different frequencies act. Likewise, when the inverse transform is obtained, one is supposed to have no interest in the frequency of the various components of the signal $x(t)$, W.J. Williams 1998 [21]. Time-localization must then be achieved by first windowing the signal, and subsequently calculating the local spectral coefficients. The window is then moved to a new position and the calculation is repeated, which is the basis for the short-time Fourier transform (STFT). To study the properties of the signal at time t, the signal is emphasized at that time and suppressed at other times. This is achieved by multiplying the signal by a window function, $w(t)$, centered at t, to produce a modified signal x_t,

$$x_t(\tau) = x(\tau)\, w(\tau - t) \tag{17}$$

The modified signal is a function of two times, the fixed time being interested (window position) t, and the running time τ. The window function is chosen to leave the signal more or less unaltered around the time t but to suppress the signal for times distant from the time of interest, Cohen 1995 [5].

Summing up it can be said that, STFT works by first dividing a signal into short consecutive segments by a windowing function and then computing the Fourier transform coefficients of each segment. This is a time-frequency localization technique in that it computes the frequencies associated with small portions of the signal as follows,

$$STFT_w x(t,\omega) = \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-j\omega\tau} d\tau \tag{18}$$

The problem with windowing, however, is that the slice of the signal which is extracted is always the same length. Thus, the number of data points in time history used to resolve a low frequency component is the same as the number used to resolve a high frequency component. Using long window can cause to lose the time localization ability of the analysis method. Shortening the window to increase time resolution can result in unacceptable increases in computational effort, especially if the short-duration phenomena being investigated do not occur very often, which is the usual case for rotating machinery faults. Due to the duality of time and frequency, the resolution is restricted by the *uncertainty principle*, Chui 1992 [3],

$$\Delta t \Delta f = 1/(4\pi) \tag{19}$$

Where Δt is duration of the window in time domain and Δf is its frequency bandwidth. Therefore, it would be sensible to search for a time-frequency method employing some sort of "adjustable" windows, which will be addressed later on.

5.1.1 Gabor transform

The limitations of Fourier methods have not by any means been observed recently. Dennis Gabor (1946) was the first to introduce time-frequency analysis, which is called *Gabor transform* for his name, or later interpreted as *time-frequency wavelets*, Meyer 1993 [14], or *Gabor wavelets*, Chui 1992 [3]. Gabor introduced a time localizing window function to broaden the application of Fourier methods. In an attempt to extract local frequency information from the signal x, he proposed a windowed Fourier transform,

$$T_g x(\omega,b) = \int_{-\infty}^{\infty} x(t) g(t-b) e^{-j\omega t} dt \tag{20}$$

This is exactly the same equation as Eq.18, except that t is taken as the running time and b is the position of window function g on the time axis. The signal x and the window function g are both energy signals, i.e. defined in $L^2(\Re)$, Lebesgue space of square-integrable functions, i.e. signals having finite/measurable energy. The windowing function Gabor used was the Gaussian function, which makes an optimal window by minimizing end effects and satisfying the minimum condition of the uncertainty principle. The Gaussian (window) function is defined as,

$$g_\alpha(t) = \frac{1}{2\sqrt{\pi\alpha}} e^{-t^2/(4\alpha)} \tag{21}$$

where α ($\alpha > 0$) determines the width of the window function.
This function is proportional to the standard normal probability density function. Using the Gaussian window function, the Gabor transform is now defined to be,

$$(G_b^\alpha x)(\omega) = \int_{-\infty}^{\infty} x(t) g_\alpha(t-b) e^{-j\omega t} dt \tag{22}$$

which localizes the Fourier transform of the signal x about the time point $t = b$. In other words, it gives a picture of the signal x at time b and angular frequency ω ($=2\pi f$). As with the continuous Fourier transform, there exists an inverse Gabor transform that will recover the original signal from its transform,

$$x(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (G_b^\alpha x)(\omega) g_\alpha(t-b) e^{j\omega t} db\, d\omega \tag{23}$$

In these two similar time-frequency analysis methods the STFT and Gabor transform, Eq.18 and Eq.22, the signal under study is subdivided into a number of small records where it is assumed that each sub-record is stationary. To reduce the effect of leakage (the effect of having finite data), each subrecord is then multiplied by an appropriate window and the Fourier transform is applied to each sub-record. However, there exists signals in nature whose spectral content is changing so rapidly that finding an appropriate short-time window is problematic since there may not be any time interval for which the signal is stationary, Matalgah 1998 [13]. To deal with these time changes properly, it is necessary to keep the length of the time window as short as possible. This, however, will reduce the frequency resolution in the time-frequency plane. Hence, there is a trade-off between time and frequency resolutions, as explained above by the *uncertainty principle*.

5.1.2 Spectrogram and Scalogram

In engineering applications, the time-frequency plot obtained by means of the square of the modulus of the STFT, $|\mathrm{STFT}_w x(t,\omega)|^2$, or Gabor transform, $|(G_b^\alpha x)(\omega)|^2$, is called a "*spectrogram*" (it is also named *sonogram*) For example, in Gabor transform based spectrograms, for each b, i.e. for each position of the window, a different spectrum may be obtained and the total number of these spectra represents a function of b and ω, which gives a time-frequency distribution.

The wavelet analogy of the spectrogram is the wavelet time-scale plot, which can be named as *scalogram*, since with wavelets, use of scale is most common, Ogden 1997 [15]. These plots consist simply of the square of the wavelet transform $|W_\psi f(a,b)|^2$, where b represents the location in time and a represents the scaling factor. In this study, scalograms, i.e. the wavelet contour maps and 3d mesh diagrams are focused in the analysis of condition monitoring machinery vibration signals.

The spectrogram and scalogram (based on respective continuous or discrete transforms) are useful objects in analyzing vibration time signals. As mentioned, both the Gabor transform and wavelet transform divide the time-frequency plane into blocks measuring in time local frequency content of the signal. For the Gabor transform (same as STFT), the analyzing windows all have constant shape with fixed window length and height. Whereas, for the wavelet transform, these blocks are short and wide for analyzing low-frequency (small-scale) content and, tall and narrow for analyzing high-frequency (large-scale) phenomena. In plotting either the spectrogram or the scalogram, each of these blocks in time-frequency (or time-scale) plane are shaded in the color scale / gray scale according to the magnitude of the corresponding coefficient.

5.2 Wigner-Ville Distribution

Similar criticism of the usual Fourier transform was addressed by Jean Ville, as early as Gabor, in 1947, as he applied it to acoustic signals, Meyer 1993 [14]. Ville introduced the physicist Eugene P. Wigner's works (1932) on calculation of joint distribution of position and momentum (for a gas) to signal analysis as joint time-frequency distribution. Ville's revival of the Wigner Distribution (WD) of quantum mechanics led the WD to be known mostly as Wigner-Ville Distribution (WVD) in the signal processing and analysis circle.

The WD or WVD is the main and hence probably the most extensively studied time-frequency distribution. It is defined, Cohen 1992 [3], for the signal $x(t)$ as,

$$W(t,\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty} x^*(t-\frac{1}{2}\tau)\,x(t+\frac{1}{2}\tau)\,e^{-j\tau\omega}\,d\tau \qquad (24),$$

where $x^*(t)$ is the complex conjugate of the original time function $x(t)$ and τ is time slices around time point t.

If the signal is written in terms of its spectrum $X(\omega)$ and substituted into the above equation, WD is obtained in terms of the spectrum,

$$X^*(\omega+\frac{1}{2}\theta)X(\omega-\frac{1}{2}\theta)e^{-jt\theta}d\theta \quad (25),$$

ι dummy variable corresponding to frequency transformation of τ. As it can be seen o equations above, the distribution is derived by generalizing the relationship between *spectrum* and the *autocorrelation* function. The Wigner distribution is said to be the signal because the signal enters twice in its calculation. It must be noted that to ;ner distribution at a particular time, the signal components made up of the product of at a past time multiplied by the signal at a future time are added up. Here the duration ;t equals to that of the future. Therefore, to determine the properties of the Wigner ɔn at a time *t*, the left part of the signal is folded in an imaginary way over to the right to re is any overlap. If there is, then those properties will be present now, at time *t*. This ɔint makes many issues and results regarding Wigner distribution clear. Everything said me-domain holds for the frequency-domain because the Wigner distribution is basically in form in both domains. Another important point, which may be a drawback, is that the distribution gives equal weighting throughout the time domain. That is it weighs the far nes equally to the near times. Hence the Wigner distribution is highly non-local.

·ly with short-time Fourier, Gabor and Wavelet Transforms, the Wigner distribution has ɔroperty of time-frequency localization. The shape of the window of the WD is similar to es used in Gabor transform or STFT, but contrarily, there is no *trade off* problem between .nd frequency resolutions in WD. This unlimited resolution property, however, can lead the ɔution to produce negative values, which present no physical meaning. This can therefore be ɔreted as violation of the uncertainty principle (refer to Eq.19).

Wigner distribution or corresponding Wigner-Ville distribution has been satisfactorily ied in the analysis of nonstationary signals, Boashash and O'Shea 1992b [2], Matalgah *et al.* 8 [13]. This comes from the ability of the WD to detect signal features in both time and Juency domains. As mentioned, one advantage of the WD over the STFT is that it does not fer directly from the time-frequency trade-off problem. On the other hand, the WD has the advantage that it is limited by the appearance of *cross-terms* (viewed as extra peaks on the ıe-frequency distribution plots, which may lead to confusion and misinterpretation). These terference terms are due to the non-linearity property of the WD. One way to remove them is ; smoothing the time-frequency plane, Cohen 1995 [4], but this will be at the expense of ecreased resolution in both time and frequency. In addition, this will increase the computational omplexity and consequently the time necessary for the computation.

A few more time-frequency distributions (called *Cohen class* time-frequency distributions) as extensions of WVD have been developed in order to alleviate the drawbacks that WVD suffers.

5.3 Wavelet Transform

A relatively new method has appeared in signal processing recently called *Wavelet Transform* (WT in short). Although wavelet theory has been studied for many years by mathematicians (before the word *wavelet* was even coined), Haar (1911), Franklin (1928), Calderon and Zygmund (1952), an explosion of applications have been seen in engineering since the mid-eighties. This resurgence was initiated by Morlet *et al.* (1982), in geoexploration studies as a technique for the analysis of seismic signals, and was followed by a detailed mathematical analysis by Grossmann and Morlet (1984), Rioul and Vetterli 1991 [19].

As stated, the basic idea in time-frequency representations is that *two* parameters are needed: one called a, which refers to frequency; the other called b, which indicates the position in the signal. Thus a general time-frequency transform of a signal x will take the form,

$$x(t) \leftrightarrow \psi(a,b) = \int_{-\infty}^{\infty} \overline{\psi_{ab}}\, x(t)\, dt \qquad (26),$$

where ψ_{ab} is the analyzing function and $\overline{\psi_{ab}}$ (indicated also as ψ^{*}) is its complex conjugate. In wavelet transform, the analyzing function ψ is defined as,

$$\psi_{ab}(t) = a^{-1/2}\, \psi\left(\frac{t-b}{a}\right) \qquad (27).$$

Combining the two equations (26 and 27), the basic formula for the continuous WT (CWT) can be obtained as,

$$W_x(a,b) = a^{-1/2} \int_{-\infty}^{\infty} \overline{\psi}\left(\frac{t-b}{a}\right) x(t)\, dt \qquad (28).$$

In discrete WT (DWT), the two parameters a and b which are for scaling and translating, respectively can be defined as functions of level j and position k

$$a = 2^{-j} \quad j \in Z, b = a.k \quad k = 0..n-1 \qquad (29).$$

Then the analyzing function ψ becomes

$$\psi_{j,k} = 2^{j/2}\psi(2^{j}t-k) \qquad (30).$$

where ψ called *mother wavelet* and $\psi_{j,k}$ called *daughter wavelet*. Here the level j determines how many wavelets are needed to cover the mother wavelet, and the number k determines the position of the wavelet and gives the indication of time. It is possible to decompose any arbitrary signal $x(t)$ into its wavelet components. The approach is similar to the harmonic analysis in Fourier transform except that, instead of breaking a signal down into harmonic functions of different frequencies and amplitudes, the signal is broken dawn into wavelets of different scale (level), different positions and the corresponding amplitudes of wavelets.

This can be put into a simpler explanation to stress the similarity of approaches between the two transforms, as follows. The Fourier transform breaks down a signal by frequency, and the wavelet transform breaks down a signal into components of different scales by comparing the signal to wavelets of different sizes. In both cases, this is done by integration: multiplying the signal by the analyzing function (sines and cosines or wavelets) and integrating the product, Hubbard 1996 [11]. FT and WT are both linear and square-integrable functions, derived from group representation theory (from different groups). The essential difference between the two is in the way the frequency (scaling) parameter a is introduced in the analyzing function. In both cases, b is simply a time translation.

5.3.1 The Diagnostics Methodology

The feature extraction scheme proposed for the fault diagnostics methodology here is based on calculating mean-square D20 WT map of the vibration signal to introduce the characterizing mean-square wavelet amplitudes of the critical levels to the ANNs. The block diagram for the methodology is sketched in Figure 2.

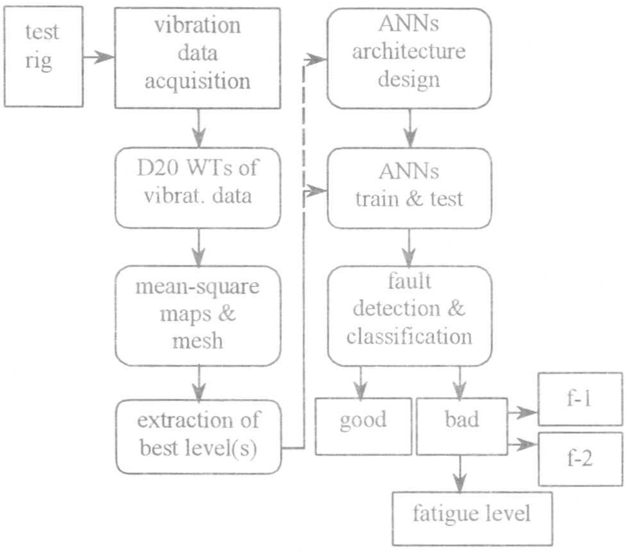

Figure 2. Block diagram of the feature extraction scheme for gear fault diagnosis.

The diagram details the steps of the fault diagnosis methodology as part of the established Condition Monitoring set-up, detailed in Engin 1998 [9]. It combines the two types of fault experiments; the impulsive (blip and shaved faults) and bending fatigue failures. The ANN based classification results of this example problem were presented before, Engin and Gülez 1999 [10]. As noted, the performance of the scheme was tested with standard backpropagation ANNs.

5.3.2 Wavelet Applications

As was the case when introducing other signal processing methods, the customary approach is followed, and the proposed WT–ANN based fault diagnostics methodology was presented with the numerically simulated data. For this purpose a simple MATLAB® program was coded to simulate vibration time signals representing three different health states of a typical 36-tooth spur gear. The resultant vibration signals representing the reference, having fault-1 (similar to gear with one tooth giving a "blip") and fault-2 (similar to gear having a "shaved" tooth) are displayed in Figure 3 (a-c), respectively.

All three kinds of signals carried an amplitude modulated main sine and several other sine functions with higher frequencies giving the meshing frequency and its first three harmonics. The fault-1 and fault-2 were introduced as localized sine functions (enveloped with fast decaying exponential components), i.e. starting with a relatively high amplitude and ending with a very low amplitude at around 205° and 210° of rotation angles. The signal and seeded fault-1, fault-2 and the noise components of the signal are plotted separately in Figure 4 (a-d), respectively.

While the first fault was designed to last 7 or 8 samples (giving a sharp impulse), the second lasted around 15 samples (simulating nearly half of a tooth was shaved), corresponding to the duration of ~2.6 and ~5.2 degrees in Figure 4 (b) and (c), respectively. The embedded faults are tried to be kept fairly small (around half of the signal amplitude) to simulate the conditions of developing faults realistically. Hence, as Figure 3 illustrates the variations between the signals are hardly distinguishable by eye.

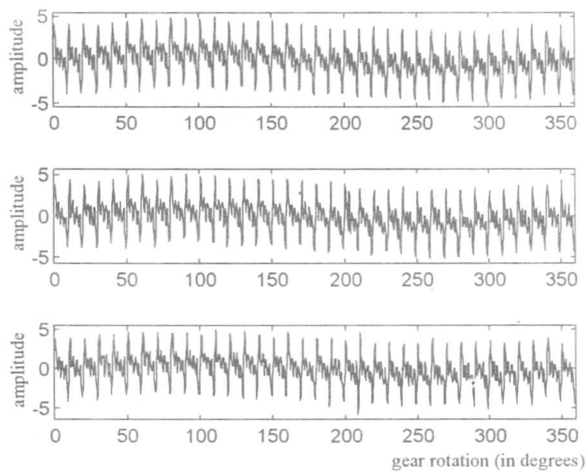

Figure 3 Numerically simulated vibration signals; the reference (a), first fault (b), and second fault (c).

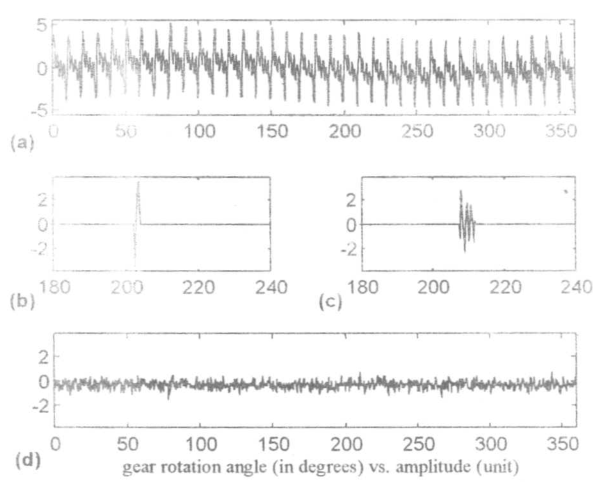

Figure 4. The reference signal (a), and Fault-1 (b), Fault-2 (c), noise components (d).

A number of copies of each health state were distorted by random noise signals as shown in Figure 4 (d). Then their D20 wavelet transform based mean-square wavelet maps were computed for feature extraction. Example 3d mesh diagrams of these maps (for the three signals plotted in Figure 3 (a-c)) are presented in Figure 5 to 7, respectively.

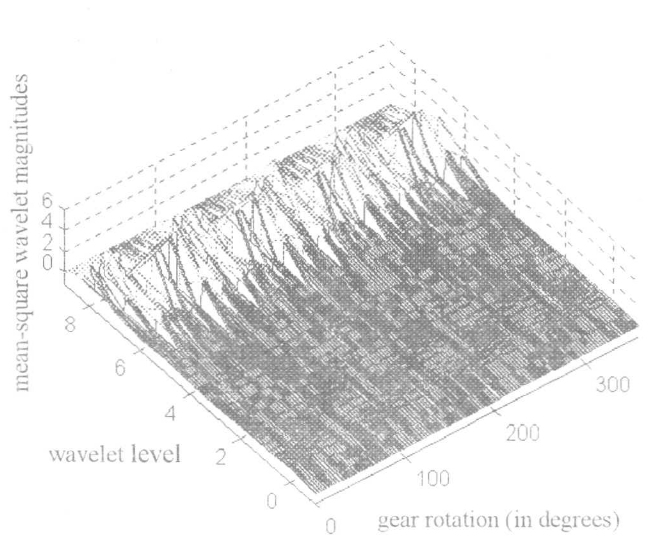

Figure 5. Wavelet mean-square mesh diagram for the numerically simulated reference vibration time signal given in Figure 3 (a).

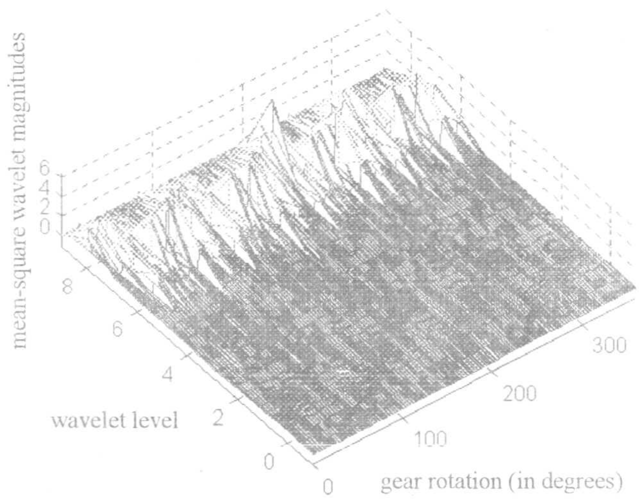

Figure 6. Wavelet mean-square mesh diagram for the numerically simulated fault-1 (similar to the blip fault) vibration time signal given in Figure 3 (b).

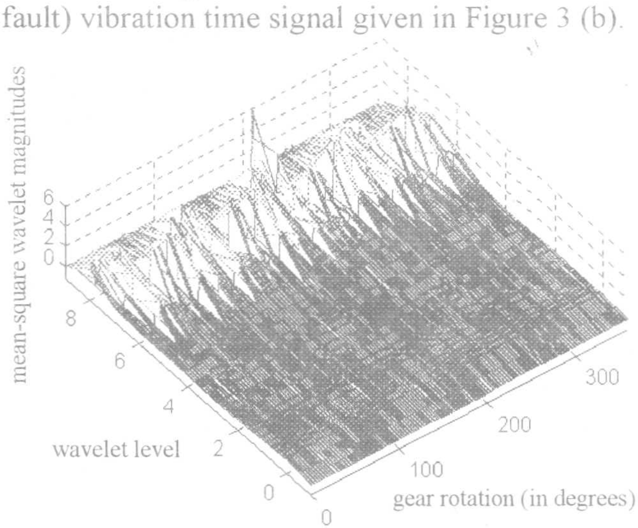

Figure 7. Wavelet mean-square mesh diagram for the numerically simulated fault-2 (similar to the shaved fault) vibration time signal given in Figure 3 (c).

Wavelet magnitudes distributed in high levels detected the changes indicating the health of the signal sufficiently. For this example, when the scalograms are studied it is seen that while levels 7 and 8 are mostly indicative for fault-1, levels 8 and 9 are more indicative for fault-2. This is due to the fact that the fault-1 involved lower fault frequencies compared to the fault-2. Therefore, while levels 7 and 8 are sufficient for fault-1, for fault-2 higher levels, i.e. levels 8 and 9 seemed to be more descriptive. And only several (about six) wavelet magnitudes in those levels reveal the differences between the health patterns.

These successful visual results led the authors to get use of the 3d wavelet plots as a feature extraction tool to incorporate with ANNs to establish an automated Fault Diagnostics methodology, Engin and Gülez 1999 [10].

6. CONCLUSIONS

The statistical or the classical time- and frequency-domain techniques offer more straightforward approaches, they all do some sort of averaging on the signal's duration, which cause loss of time information. This is a real drawback when dealing with signals that have very short-lasting or suddenly occurring components. While, time-frequency methods, e.g. short-time Fourier transform and Wigner-Ville distribution can localize those nonstationary features, they suffer from several inconveniences, which is difficult to compensate, such as time-frequency trade-off problem and returning extra/unwanted terms. The wavelet transform as a time-frequency or time-scale method has proved itself as a powerful analysis method in pinpointing signal features, which may represent a fault. On the other hand, AI based techniques, in particular ANNs, are becoming commonly used pattern classification tools. Therefore, there is a great need to develop effective ANN input data pre-processing algorithms, which exploit advanced signal processing and analysis techniques.

The challenge is therefore to search for a good TFTS-analysis/ANN integration methodology, which holds an effective feature extraction scheme producing compressed and interpretable input data for the ANN for fast and reliable pattern recognition. The system to be devised, on the other hand, should be able to give good fault diagnosis results when dealing with *small faults*, which represent *early stages* of an impending failure.

REFERENCES

[1] Badi, M.N.M., Engin, Ş.N., Schönfeld, D., "Fault classification of a model drive-line using time domain data", Proc. of 9th COMADEM'96 Int., University of Sheffield, UK, 43-51, July 16-18, 1996, Sheffield Academic Press, eds. Rao, B.K.N., Smith, R.A. and Wearing, J.L., 1996

[2] Boashash, B. and O'Shea, P., "Signal detection using time-frequency analysis", Time-frequency signal analysis methods and applications", chap. **12**, 279-304, Longman Cheshire, Melbourne, ed. Boashash, B., 1992.

[3] Chui, C. K., "An Introduction to Wavelets" Academic Press, San Diego, 1992.

[4] Cohen, L., "Introduction: A Primer on time-frequency analysis", Time-frequency signal analysis methods and applications", chap. **1**, 1-42, Longman Cheshire, Melbourne, ed. Boashash, B., 1992.

[5] Cohen, L. "Time-frequency analysis", Englewood Cliffs, NJ: Prentice Hall, 1995.

[6] Cohen, L., "Time-varying spectral analysis – New concepts and unsolved problems", Proceedings of IEEE UK Symp. on Applic of Time-Frequency / Time-Scale Methods - TFTS'95, Univ. of Warwick, UK, 3-10, Aug 30-31, 1995, ed. Lawson, S., 1995.

[7] Engin, Ş.N., Badi, M.N.M., Esat, I.I., "Use of different wavelet transforms in condition monitoring vibration data", Proc. of 9th COMADEM'96 Int., University of Sheffield,UK, 463-472, July 16-18, 1996, Sheffield Academic Press, ed. by Rao, B.K.N., Smith, R.A. and Wearing, J.L.

[8] Engin, Ş.N., Badi, M.N.M., Yeşilyurt, I., "A D20 wavelet transform based feature extraction scheme for automated rotating machinery fatigue failure assessment", Proc. of IEEE UK Symp. on Applic. of Time-Frequency / Time-Scale Methods - TFTS'97, Univ. of Warwick, Coventry, UK, 145-148, Aug 27-29, 1997, ed Lawson, S.

[9] Engin, Ş.N., "Condition monitoring of rotating machinery using wavelets as a pre-processor to Artificial Neural Networks (Artificial Neural Networks)", PhD Thesis, University of Hertfordshire 1998

[10] Engin,Ş.N., Gülez,K. "A wavelet transform artificial neural networks (WT–ANN) based machinery fault diagnostics Methodology", NSIP99, Antalya, Turkey, to appear.

[11] Hubbard, B. B., "The world according to wavelets – the story of a mathematical technique in the making" A K Peters, MA, USA, 1996.

[12] Hunt, T.M., "Condition monitoring of mechanical and hydraulic plant", Chapman & Hall, London, 1996.

[13] Matalgah, M., Knopp, J. and Mawagdeh, S., "Iterative signal processing method using Gabor wavelets and wavelet transform for the analysis of phonocardiogram signals", Time Frequency and Wavelets in Biomedical Signal Processing, IEEE press, M. Akay, .271-304, 1998.

[14] Meyer, Y., "Wavelets – Algorithms & applications", SIAM (Society For Industrial and Applied Mathematics), Pennsylvania, US, 1993 (Translated by Ryan, R.D.).

[15] Ogden, R.T., "Essential wavelets for statistical applications and data analysis", Birkhäuser, Boston, 1997.

[16] Randall, R. B. and Hee, J., "Cepstrum analysis", Digital Signal Analysis using Digital Filters and FFT Techniques, Brüel & Kjær, pp. 161-198, Jan. 1985.

[17] Randall, R. B., "Frequency Analysis", Brüel & Kjær Application Notes, 1977.

[18] Rao, S. S, "Mechanical vibrations", Addison-Wesley Publishing Co., 1995

[19] Rioul, O. and Vetterli, M., "Wavelets and signal processing" in IEEE Signal Processing Magazine, 84(4), pp.14-38, Oct. 1991.

[20] Staszewski, W.J., "Gearbox vibration diagnostics - An overview", Proceedings of the 9th International

[21] Conference on Condition Monitoring and Diagnostic Engineering Management (COMADEM '96), University of Sheffield, UK, pp. 43-51, July 16-18, 1996, Sheffield Academic Press, eds. Rao, B.K.N., Smith, R.A. and Wearing, J.L.

[22] Williams, W. J., "Recent advances in time-frequency representations: some theoretical foundations", Time Frequency and Wavelets in Biomedical Signal Processing, IEEE press, ed. M. Akay, 3-43, 1998.