Journal of Sound and Vibration (1998) **211**(5), 735–760

WAVELET BASED COMPRESSION AND FEATURE SELECTION FOR VIBRATION ANALYSIS

W. J. Staszewski

Dynamics Research Group, Department of Mechanical Engineering, University of Sheffield, Mappin Street, Sheffield S1 3JD, England

(*Received* 14 *January* 1997, *and in final form* 21 *July* 1997)

This paper is concerned with wavelet based linear transformations for data compression and feature selection in vibration analysis. Recent developments in wavelet data compression are summarized. A discussion of various types of data including periodic, continuous non-stationary and transient non-stationary signals, are used to show practical aspects of wavelet compression. The analysis employs smooth wavelets and compactly supported wavelets. It has been shown that compression in vibration analysis can be used not only for effective storage and transmission of the data but also for feature selection. A number of different approaches have been presented to show coefficient selection procedures. This includes procedures based on truncated wavelet coefficients according to their amplitude, position and frequency location and a data compression technique based on optimal wavelet coefficients.

1. INTRODUCTION

Many vibration analysis methods require a significant amount of data to store, transmit or process. Further it often appears that data used for vibration analysis is highly correlated. This includes spatial correlation where the values of the signal can be predicted from the neighbouring points in the space–time domain, and spectral correlation where the frequency domain can be used for prediction. The correlations present in the data can be removed using compression. Data compression has been the subject of extensive research for the last three decades. Research and developments are particularly extensive in the area of image processing and information theory. A vast amount of literature can be found summarizing these developments (e.g., reference [1]). There exist many different methods of data compression including: predictive coding, transform coding, pyramid techniques, entropy coding, vector quantization and hybrid techniques. Transform coding is one of the standard techniques used for data compression. A number of transforms can be used for image compression. This includes [1, 2]: the Fourier, cosine and sine, Haar, Walsh–Hadamard, Slant and Karhunen-Loeve transforms.

Statistically, the Karhunen–Loeve expansion is the optimal transform for data compression. The optimal basis is given by the eigenvectors of the correlation matrix. In practice the correlation matrix is not known and the transform is computationally expensive. Additionally, the basis of the transform depends on the data used for compression. Recent years have seen some developments in wavelet and fractal analysis for compression [3–9]. The wavelet transform can give a compression basis which is independent of the data set and can reveal local temporal correlations in the data. Also there exist fast algorithms for wavelet transform calculations. The majority of applications are in the area of image compression [4–9] and acoustical signals [3, 6].

0022–460X/98/150735 + 26 $25.00/0/sv971380

All compression methods fall into two basic schemes: lossless and lossy. Lossless compression does not introduce any distortion and loss of information; the original and reconstructed (decoded) data are identical. Lossy compression allows a certain level of distortion between the original and reconstructed data as long as the error is acceptable. It is obvious that textual data requires lossless compression. On the other hand, lossy compression, with significant error, can be used for image data. Here the acceptable error is when a decoded image is visually the same as the original image. The data used in vibration analysis may require different compression schemes with various levels of acceptable errors. However, the error level is usually smaller for vibration data than for image data. Thus, simple compression based on subsampling is good enough for images but very often not acceptable for vibration data.

It appears in practice that the wavelet based procedure is not widely used for compression in vibration analysis. It has to be underlined that data compression in vibration analysis is not only an effective storage and transmission procedure, but also, if not predominantly, a feature selection procedure. Rough vibration data is very often not appropriate for further analysis and initial feature selection procedure is required. This is for example the case in fault detection where advanced detection procedures like neural networks very often need pre-processing analysis. The feature selection procedure may be a lossy compression scheme with a large level of acceptable error. This in practice leads to high compression ratios.

The aim of this paper is: to bring together recent developments in wavelet based data compression, to give some practical guidance on how to compress different types of vibration data, and to show possible applications in vibration analysis, in particular feature selection procedures based on wavelet compression algorithms. The paper is addressed to the mechanical engineering community and will hopefully provide help and understanding of wavelet based compression in this area. For detailed analysis about compression and wavelets the reader is referred to other publications.

The structure of the paper is as follows. Section 2 defines a linear transformation for data compression. The theoretical background of wavelet analysis is given briefly in section 3. Section 4 summarizes a basic theory of wavelet compression, and gives a number of compression examples for different types of data. A compression procedure with optimal wavelet coefficient is presented in section 5. Section 6 gives two possible feature selection procedures in the area of fault detection. Finally, the paper is concluded in section 7.

2. LINEAR TRANSFORMATION FOR DATA COMPRESSION

Transform coding is a standard compression procedure which leads to a generalized frequency domain. The theory of linear discrete transforms for compression can be found in reference [2]. In what follows, the basic idea is given. The data is represented in the form of a vector $\bar{\mathbf{X}}$, which usually has equally distributed energy among its elements. The basic idea is to move $\bar{\mathbf{X}}$ to a vector $\bar{\mathbf{Y}}$ using a linear transformation $\mathscr{R}$:

$$\bar{\mathbf{Y}} = \mathscr{R}(\bar{\mathbf{X}}). \tag{1}$$

The new vector $\bar{\mathbf{Y}}$ concentrates most of the energy in only a few vector elements. Thus, data compression can be achieved by setting some vector elements below a threshold to zero and discarding them. The data can then be reconstructed using an inverse transformation. The linear transformation of a vector is based on the decomposition of this vector in terms of a basis of elementary vectors. A simple example can be given by

Fourier analysis, where any periodic function $x(t)$ has a Fourier series representation given by,

$$x(t) = \sum_{n=-\infty}^{n=+\infty} a_n \, e^{i2\pi nt}, \tag{2}$$

where a_n are the Fourier coefficients defined as,

$$a_n = \frac{1}{2\pi} \int_0^{2\pi} x(t) \, e^{-i2\pi nt} \, dt. \tag{3}$$

Clearly, function $x(t)$ is decomposed into an infinite sum of elementary functions $f_n = e^{i2\pi nt}$ which form the orthogonal basis. The orthogonality condition means that the inner product $\langle f_m, f_n \rangle$ satisfies

$$\langle f_m, f_n \rangle = \frac{1}{2\pi} \int_0^{2\pi} f_m(t) f_n^*(t) \, dt = 0 \tag{4}$$

for all integers $m \neq n$. It is clear that the elementary functions $e^{i2\pi nt}$ are generated by iteration or dilation of a single function $e^{i2\pi t}$.

Other examples of transform coding include transforms already mentioned in the previous section. The compression based on the Fourier transform does not depend on the data set. The algorithm of calculations is simple and fast. However, the method very often fails to remove correlations in non-stationary data. This is due to the nature of the basis functions, which are local in the frequency domain but non-local in the time domain. Additional locality in the time domain can be obtained using wavelets.

3. WAVELET ANALYSIS

3.1. THEORETICAL BACKGROUND

For the sake of completeness, a brief introduction to the relevant wavelet theory is given in this section. More detailed analysis can be found in reference [10].

By analogy to the Fourier transform, the wavelet transform is a linear transformation that decomposes a given function $x(t)$ into a superposition of elementary functions $g_{a,b}(t)$ derived from an analyzing wavelet $g(t)$ by scaling and translation i.e.,

$$g_{a,b}(t) = g^*\left(\frac{t-b}{a}\right), \tag{5}$$

where * denotes complex conjugation, b is a translation parameter indicating the locality and a is a dilation or scale parameter.

The discrete wavelet transform refers to a discrete time-scale framework. Within this framework, when a binary dilation and dyadic translation is used, the orthogonal wavelet transform can be defined. A function $g(t)$ is called an orthogonal wavelet if the family

$$g_{m,k}(t) = 2^{m/2} g(2^m t - k) \tag{6}$$

forms an orthonormal basis, that is

$$\langle g_{m,k}, g_{n,l} \rangle = \delta_{m,n} \cdot \delta_{k,l} \tag{7}$$

for all integer m, n, k, l, where $\langle , \rangle$ is the usual inner product and $\delta_{m,n}$ is the Kronecker symbol. The orthogonal wavelet transform can now be defined as

$$x_k^m = \int_{-\infty}^{+\infty} x(t) g_{m,k}(t)\, \mathrm{d}t. \tag{8}$$

The transform can be interpreted as a filter bank decomposition. The scale (frequency) partitioning leads to a partitioning in the time domain that is finer in the higher frequency bands. The wavelet synthesis formula is given by

$$x(t) = \sum_m \sum_k x_k^m\, g_{m,k}(t). \tag{9}$$

3.2. ANALYSING WAVELET

A number of different bases have been proposed to construct orthogonal wavelets. The simplest basis can be given by the well-known Haar function $h(t)$ that is equal to 1 on $\langle 0, 1/2)$, -1 on $\langle 1/2, 1)$ and 0 outside the interval $(0, 1)$. Much more effective analysis and synthesis can be obtained with the wavelets $g_r(t)$ of Daubechies [10], which are orthogonal and have the following properties: (1) the support of $g_r(t)$ is the interval $[0, 2r+1]$; (2) $g_r(t)$ have r vanishing moments, i.e., $\int_{-\infty}^{+\infty} t^r g_r(t)\, \mathrm{d}t = 0$; (3) $g_r(t)$ has γr (γ is about 0·2) continuous derivatives.

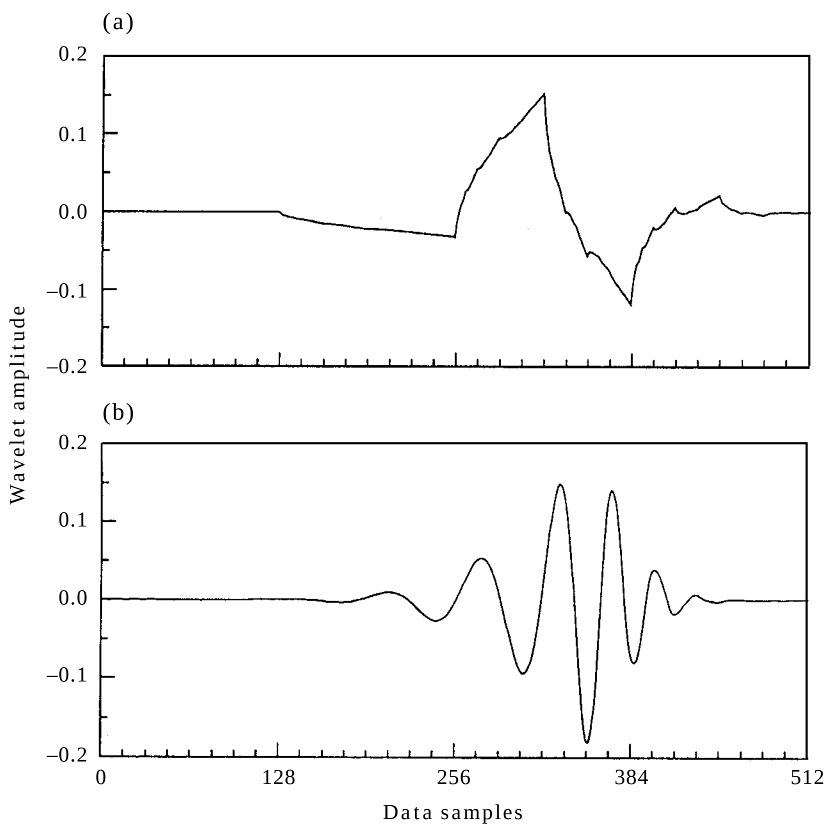

Figure 1. An example of Daubechies' wavelets: (a) fourth order; (b) twentieth order.

Figure 1 shows the fourth and twentieth order Daubechies' basis functions. It can be shown that when $r = 0$, this basis reduces to the Haar wavelet analysis. The orthogonal wavelet decomposition based on the Daubechies' wavelets can be obtained using the pyramidal Mallat algorithm. For more details about the calculation procedure, the reader is referred to references [11, 12]. Other popular wavelets which can be used for compression are the Lemarié, Coiflet and Mexican Hat wavelets [13, 14].

4. WAVELETS FOR DATA COMPRESSION

A wavelet based compression is based on a linear transformation given by equations (1), (8) and (9). The algorithm essentially consists of four major steps: transformation, thresholding, quantization/encoding and reconstruction.

4.1. TRANSFORMATION

The wavelet transform given by equation (8) is computed using the original data. The operation requires a proper choice of the wavelet analyzing function. This is very often a trade-off between the smoothness (differentiability) and compact support of the wavelet functions. The smoothness of a function corresponds to the decay of its Fourier transform. It can be measured using the Sobolev norm obtained from the derivative norms

$$\| g \|^2_{W^N} = \sum_k \| g^{(k)} \|^2. \tag{10}$$

The support of the function means the smallest closed set outside which the function vanishes identically. In general more compactly supported,* and therefore less smooth wavelet functions, are better for non-stationary data with discontinuities, impulses or transients. This class includes, for example, lower order Daubechies' wavelets. Less compactly supported, and therefore more smooth wavelet functions, are better for stationary, regular data or in cases where low level of compression error is required. In such situations, higher order of Daubechies' wavelets or very smooth Lemarié's wavelets [13] can be used. In practice, the number of vanishing moments of the wavelet functions is also important. For regular, smooth, stationary data, more vanishing moments lead to smaller wavelet coefficients. However, for non-stationary, unregular data more vanishing moments lead to more large wavelet coefficients.

4.2. THRESHOLDING

Transient data can be represented by a smaller number of coefficients compared with regular, stationary data which needs more wavelet coefficients for time-scale representation. In the time-scale domain, data compression can be achieved by setting wavelet coefficients x^k_m below a threshold t to zero by applying a threshold function

$$F(t, x^k_m) = \begin{cases} 0 & |x^k_m| < t \\ 1 & |x^k_m| \geqslant t \end{cases} \tag{11}$$

to these coefficients. Thus, the absolute amplitude and not the position of the coefficients is important in the procedure. It is obvious that the bigger the threshold, the higher the compression ratio that can be achieved.

* Strictly speaking a set is either compact or not. Therefore a function cannot be more or less compactly supported than another. However, if one has a measure, as in this case on $L^2(\mathscr{R})$, one can compare the sizes of the compact supports and introduce an order relation.

4.3. QUANTIZATION AND ENCODING

Quantization is the procedure which restricts the values of chosen wavelet coefficients to a limited number of levels. Encoding is an operation in which the whole scale of values is divided into intervals represented by coded symbols. A simple scalar quantization represents the coefficient in terms of these symbols. A more advanced method called vector quantization replaces groups of coefficients with one symbol [15]. Encoding is an operation in which a quantized vector is replaced by a bit stream. Two standard methods used for this operation are: fixed length and variable length coding. Other methods based on entropy coding concepts can also be used. These techniques relate the information about possible coefficient amplitude levels with the probability of their occurrence. The idea is to assign shorter symbols to the more frequent amplitude levels. Here the amount of information (entropy) required to code a symbol of probability p is given by

$$I = \log_2 (1/p). \tag{12}$$

The performance of the encoder can be improved when self-similarity analysis between wavelet sub bands is used [7].

4.4. RECONSTRUCTION

The data can be recovered from its compressed form using encoding and quantization algorithms and the formula is given by equation (9). The compression procedure described in this section can be performed well for the data which can be represented by a small number of wavelet coefficients. This can be assessed by a function M [7]

$$\left(\sum_{i,j} |\langle x, \psi_{j,k} \rangle|^p \right)^{1/p}, \tag{13}$$

by finding the smallest value of p for which M is bounded. The smaller the value of p, the fewer wavelet coefficients are needed to represent the data and thus a better performance of the algorithm can be obtained.

In practice, the wavelet based compression algorithm is more suitable for non-stationary type data. Thus, any parameter which reveals the level of non-stationarity can be used to assess the performance of the algorithm. As an example, the parameter which measures the extent to which a function $x(t)$ is narrow-banded. This parameter can be defined by [16]

$$q = \sqrt{1 - \frac{m_1^2}{m_0\, m_2}}, \tag{14}$$

where m_n is the nth spectral moment of the single sided spectrum of $x(t)$ defined as

$$m_n = \int_0^\infty \omega^n S_{xx}(\omega)\, \mathrm{d}\omega. \tag{15}$$

The smaller the value of q, the more narrow-banded the analyzed process. Simple examples of the wavelet based compression are given in the next section.

5. COMPRESSION EXAMPLES

In order to show some of the properties of the wavelet based compression discussed in the previous section, three examples with different types of real vibration data from rotational machinery were analyzed. The data included 512 samples. The data was compressed using three different transformations: the Fourier transform (FFT) and the wavelet transform with the fourth (DAUB4) and twentieth (DAUB20) Daubechies' wavelet functions. A simple scalar quantization procedure using a six-bit representation was applied.

5.1. COMPRESSION MEASURES

Wavelet based compression represents a lossy compression scheme and allows a certain level of distortion between the original and reconstructed data. This error can be controlled using different measures. In order to compare different compression algorithms and estimate the performance of the methods, two parameters have been used. The first one is the compression ratio defined as the ratio between the number of bits of the original data and the number of bits of the compressed data. The second parameter is the normalized Mean Square Error (MSE) given as

$$\mathrm{MSE}(x) = \frac{100}{N\sigma_x^2} \sum_{i=1}^{N} (x_i - \hat{x}_i)^2, \tag{16}$$

where x_i are the samples of the original function, $\hat{x}_i$ are the samples of the reconstructed (encoded) function, σ_x is the standard deviation of the function and N is the number of sample points in the analyzed data.

5.2. PERIODIC DATA

The first analyzed example involved the stationary periodic data given in Figure 2(a). The frequency domain representation in a form of the power spectrum is represented in Figure 2(b). The value of q^2 for this data is equal to 0·13. Here, three predominant frequency components in the spectrum can be clearly observed. A few sidebands together with some level of noise can also be seen. The solid line in Figure 2(c), displays wavelet transform coefficients calculated for the analyzed data. Here the twentieth order of the Daubechies' wavelet function (DAUB20) has been used. The same coefficients, but plotted according to decreasing order of amplitude, are, presented in Figure 2(c) and denoted by the dashed line. It can be seen that the 150th largest wavelet coefficient is about 10 times smaller than the largest coefficients. This suggests that the wavelet domain concentrates energy better than the time domain, since a smaller number of samples is required to represent the data. However, one can already notice that the frequency domain (Figure 2(b)) requires even less coefficients to represents the data than the wavelet domain.

The compression results for different numbers of wavelet coefficients used for compression and different compression ratios are presented in Figures 3(a) and (b), respectively. It can be clearly seen that the Fourier transform based compression gives much smaller error than the wavelet transform compression. As already mentioned in section 2, the twentieth order Daubechies' wavelet (DAUB20) is more suitable for stationary, regular data.

5.3. CONTINUOUS NON-STATIONARY DATA

Continuous non-stationary data was analyzed for the second example. The data consists of a series of impulses embedded in the noise. Figure 4(a) shows the data in the time domain. The power spectrum represented in Figure 4(b), displays much more frequency components than the previously analyzed data in Figure 2(b). The value of q^2 for this data is equal to 0·23. The wavelet coefficients for this data are represented in Figure 4(c) by the solid line, and in decreasing order of amplitude by the dashed line. It can be seen that in this example, the wavelet domain concentrates the energy better than the Fourier domain. The decay of the dashed line in Figure 4(c) is also faster than in Figure 2(c). This

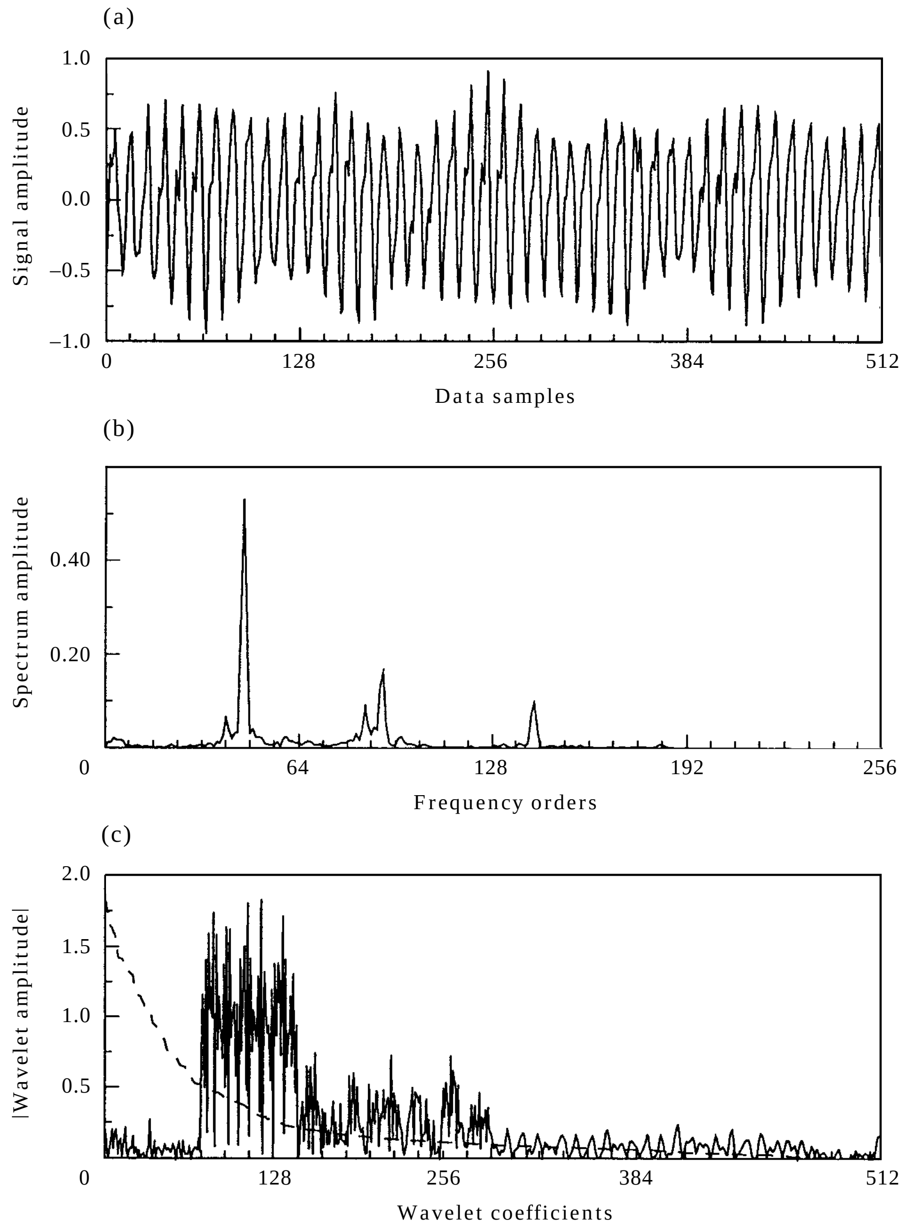

Figure 2. Periodic data used for compression in: (a) time domain, (b) frequency domain, (c) wavelet domain; dashed line indicates wavelet coefficients decreasing according to the amplitude level.

means that a smaller number of wavelet coefficients than in the previous example can be used to represent the data with the same accuracy. The compression results are given in Figure 5. Here the wavelet based compression performs better than the Fourier analysis. In contrast to the previous example, the fourth order function (DAUB4) gives better results than the twentieth order function out of the two Daubechies' wavelets used in the analysis.

5.4. TRANSIENT DATA

The third analyzed example involved non-stationary transient data. Figure 6(a) shows a transient in the time domain. Its frequency domain characteristic is given in Figure 6(b).

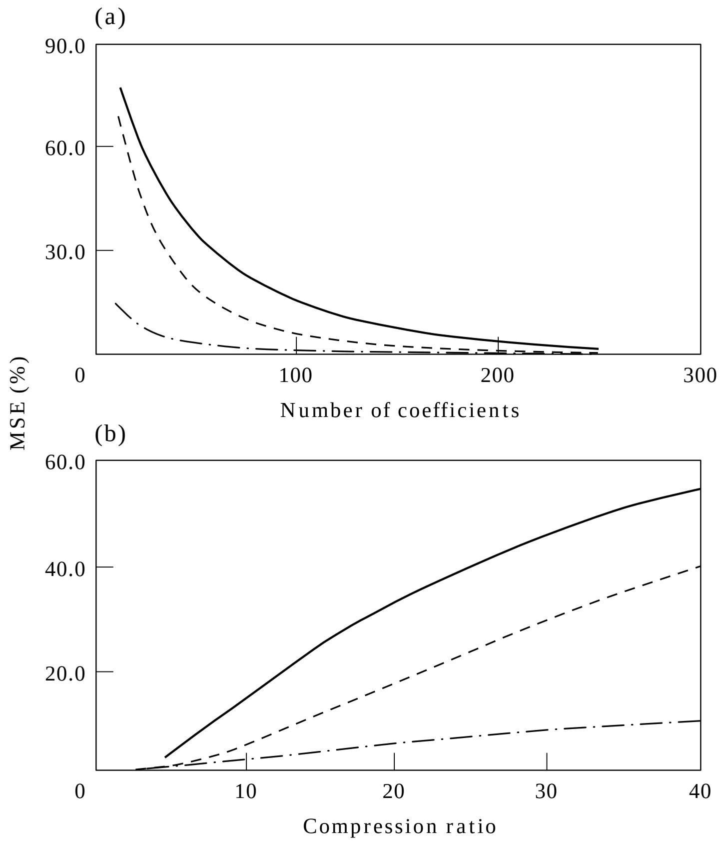

Figure 3. Compression performance for periodic data. MSE plotted as a function of: (a) wavelet coefficients used for compression, (b) compression ratio. ——, DAUB4; – – – –, DAUB20; —·—·—, FFT.

Here, the broad-band nature of the power spectrum with one predominant spectral component can be clearly observed. The value of q^2 for this data is equal to 0·33. The solid line in Figure 6(c) gives the wavelet coefficients for the data. These coefficients are plotted in decreasing order of amplitude using the dashed lines. It can be seen that in contrast to the Fourier domain (Figure 6(b)), the wavelet domain (Figure 6(c)) needs only a few coefficients to represent the transient data. The compression results are given in Figure 7. It can be clearly seen that the wavelet based compression method outperforms the Fourier based technique. Both wavelet functions used in the analysis give similar results.

5.5. DISCUSSION AND SUMMARY

An example showing the comparison between the original data used and the reconstructed (encoded) data is given in Figure 8. Here, three types of data used in sections 5.2–5.4 can be seen. The analysis involved the twentieth order Daubechies' wavelet in Figure 8(a) and the fourth order Daubechies' wavelet in Figures 8(b) and (c). The compression ratio 6:1 gives MSEs equal to 2·02, 5·03 and 0·2% for the periodic, continuous non-stationary and transient data, respectively. These results show good performance of the wavelet based compression algorithms especially for transient data.

The performance of the compression algorithms strongly depends on the data used. Thus, the presented results should be considered only as guidance for different types of data and various compression algorithms. It has to be mentioned that neither the linear

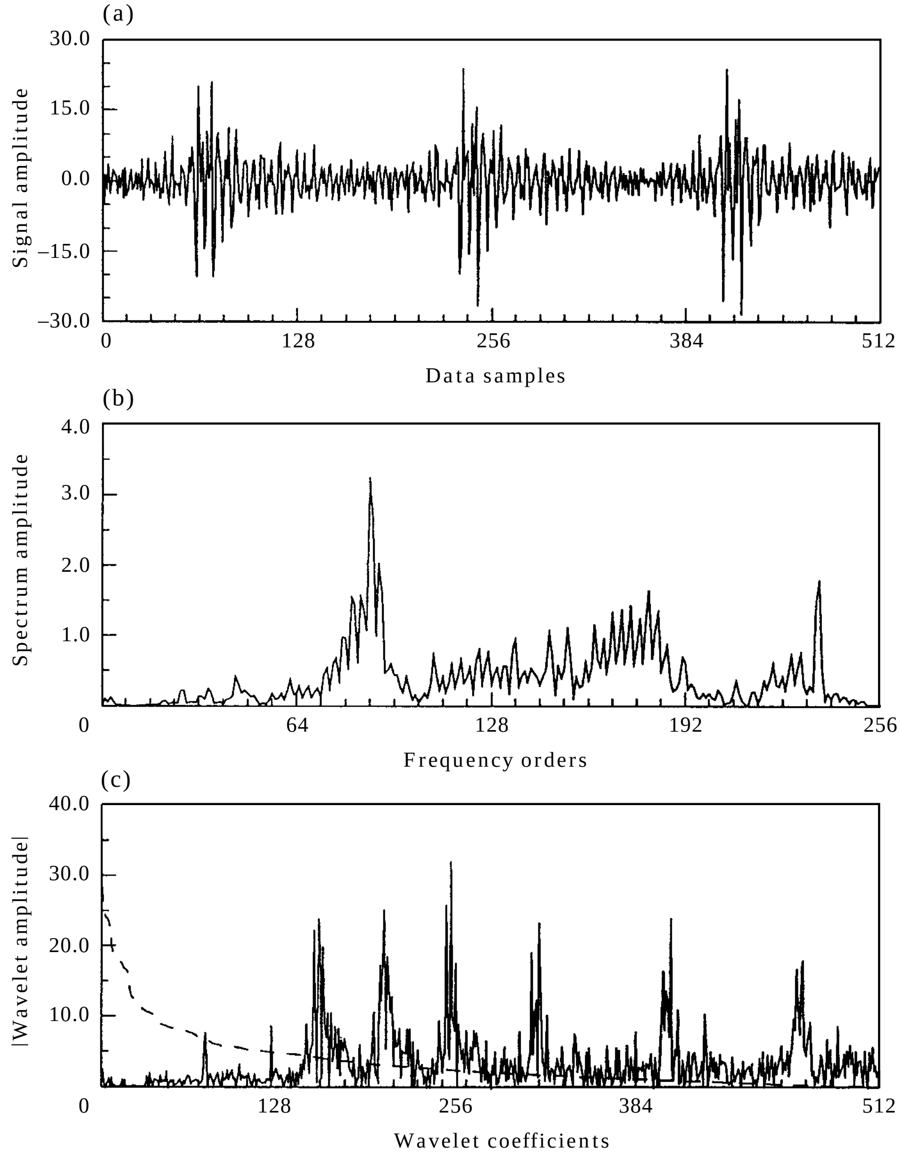

Figure 4. Continuous non-stationary data used for compression in: (a) time domain, (b) frequency domain, (c) wavelet domain; dashed line indicates wavelet coefficients decreasing according to the amplitude level.

transformation based algorithms nor the quantization are optimal for the analysis used. This was not the aim of the study. The examples show that in general: (1) the compression is a trade off between the compression ratio and the quality of compression; (2) the non-stationary data is more suitable for wavelet based compression; (3) the more non-stationary or transient the data, the better the performance of the wavelet compression; (4) the smooth wavelets are better for regular, stationary, periodic data; (5) the compactly supported wavelets are better for non-stationary, transient data; (6) the q used to measure the extent to which the spectrum is narrow-banded can be used to select a type of linear transformation and wavelet function for compression.

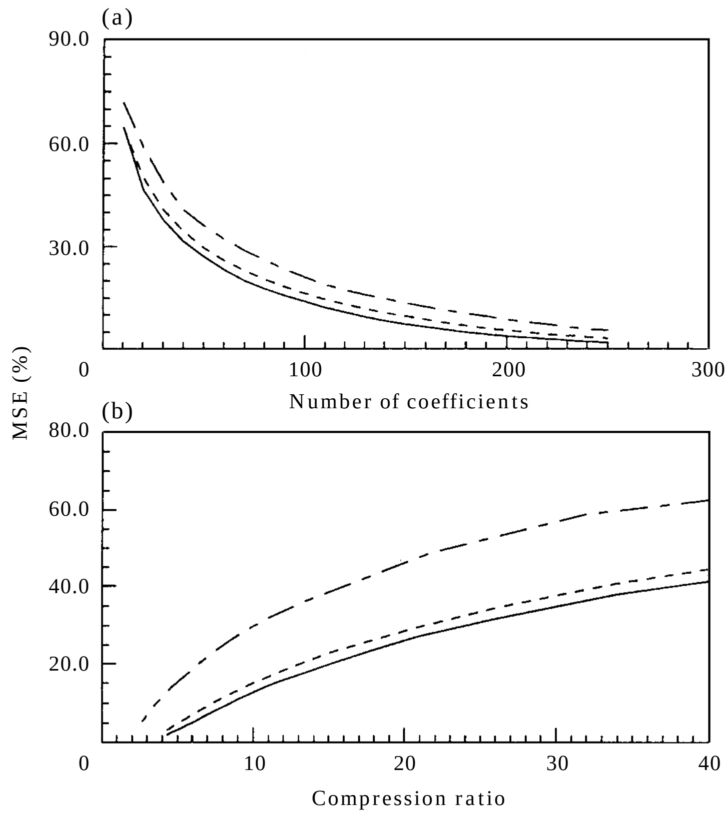

Figure 5. Compression performance for continuous non-stationary data. MSE plotted as a function of: (a) wavelet coefficients used for compression, (b) compression ratio. ——, DAUB4; – – – –, DAUB20; —·—·—, FFT.

6. WAVELET BASED FEATURE SELECTION FOR VIBRATION ANALYSIS

The aim of compression is to reduce the amount of information for effective storage and transmission. However, data compression in many areas of vibration analysis is used for feature selection. This is the case in machinery and structural fault detection. The problem of fault detection can be regarded as one of pattern recognition. Different machine conditions can be described by patterns of features. Any fault detection procedure is therefore a method which can identify fault features and determine a relationship between these features and different machine conditions. All possible machine conditions form classes which have to be known in advance. Neural networks have been established as a powerful aid to pattern recognition. They have the potential to form internal representations of fault classes through training on raw data. It is often impractical to use directly all data values for training; a simple decimation procedure can be used to reduce the training feature space. This reduction does not necessarily mean that the fault symptoms will be preserved or enhanced. An alternative approach can be offered by the wavelet compression: compressed data can be used for the network training. The compression procedure significantly reduces the feature space, removes irrelevant information and enhances features which exhibit faults.

Wavelet based compression for feature selection does not require good quality of accuracy of compression but extracts some specific features which can characterize the data. If these features are not represented by wavelet coefficients with the highest

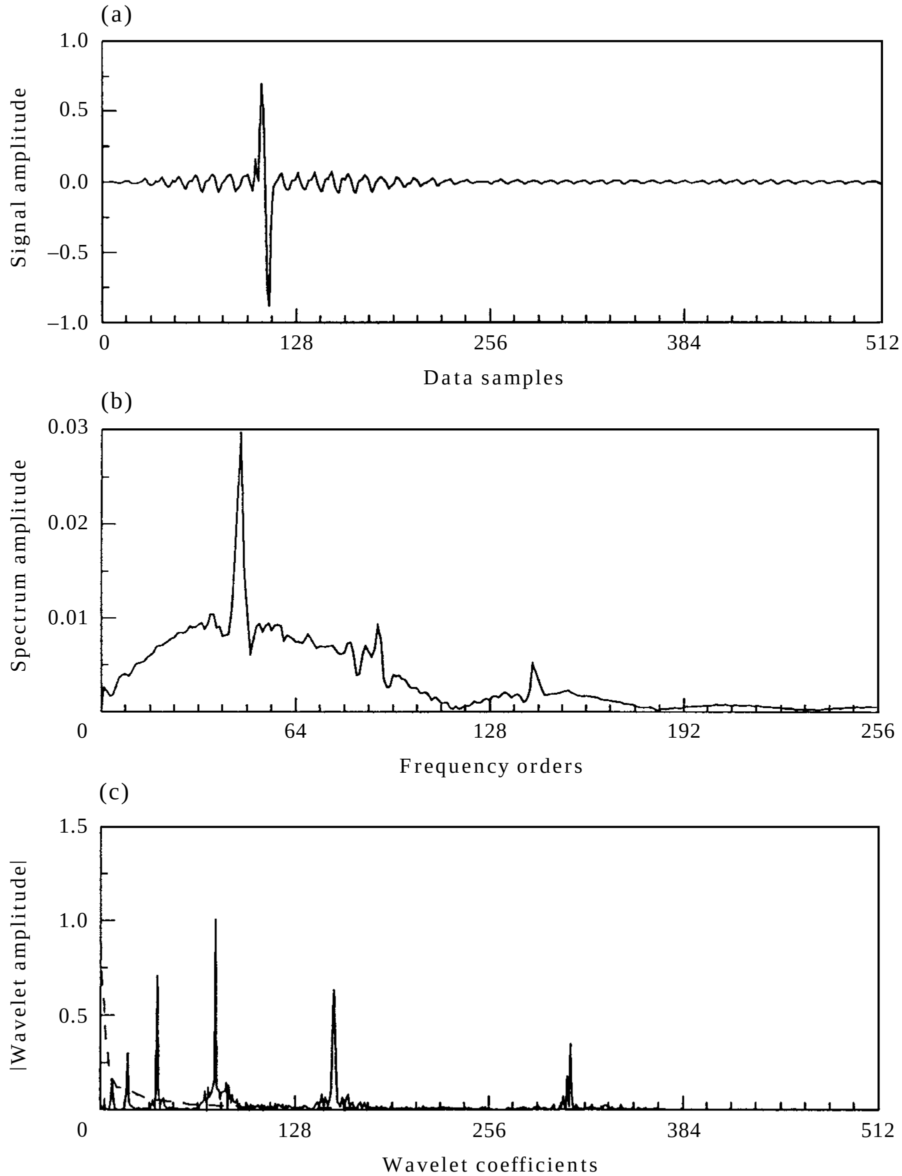

Figure 6. Transient data used for compression in: (a) time domain, (b) frequency domain, (c) wavelet domain; dashed line indicates wavelet coefficients decreasing according to the amplitude level.

amplitude, the wavelet based compression algorithm described in section 4 is often not suitable to keep the features with good accuracy. The question is which wavelet coefficients represent the features, or in other words, which coefficients to choose for feature selection. If the answer was known, the compression could be based on these coefficients. In what follows, a number of different approaches are presented to show coefficient selection procedures and present wavelet based compression for feature selection. Applications of these procedures for pattern recognition analysis can be found in references [17, 18].

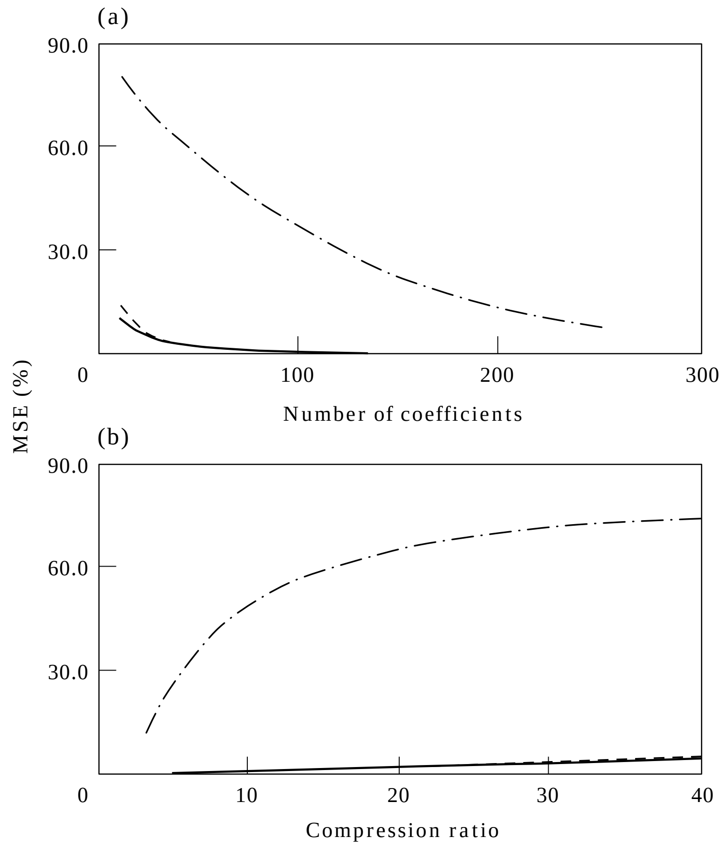

Figure 7. Compression performance for transient data. MSE plotted as a function of: (a) wavelet coefficients used for compression, (b) compression ratio. ——, DAUB4; – – – –, DAUB20; —·—·—, FFT.

6.1. THRESHOLDING PROCEDURE

A simple thresholding procedure, described in section 4.2, is often still sufficient to select required features in vibration data. These features are based on the highest wavelet coefficients chosen for data compression. An example presented in this section involves an ultrasonic acoustic data.

Ultrasonic Lamb waves are used for detection of various damages in composite materials. The detection of gross defects can be accomplished with relative ease. However, some defects, such as delaminations, require enhancement techniques involving feature selection procedures. Figure 9 shows an example of ultrasonic Lamb waves given by 512 data samples. The upper wave shows a damage-free region of the composite plate while the lower wave is recorded directly in front of the delamination. The amplitude of the impulse is about 30 μs depends on the coupling between Lamb wave transducer and the analyzed plate and thus is not a feature of the fault. The reflection from the damage would be expected between 50 and 140 μs. A comparison between upper and lower waves does not clearly show any indication of the fault and allows for considerable scope of improvement in the damage detection procedure. The Lamb wave data was compressed using the procedure described in section 4. The normalized data was decomposed using the orthogonal wavelet transform. Only the 17 highest wavelet coefficients were kept for further analysis; the remaining coefficients were set to zero. The inverse wavelet transform

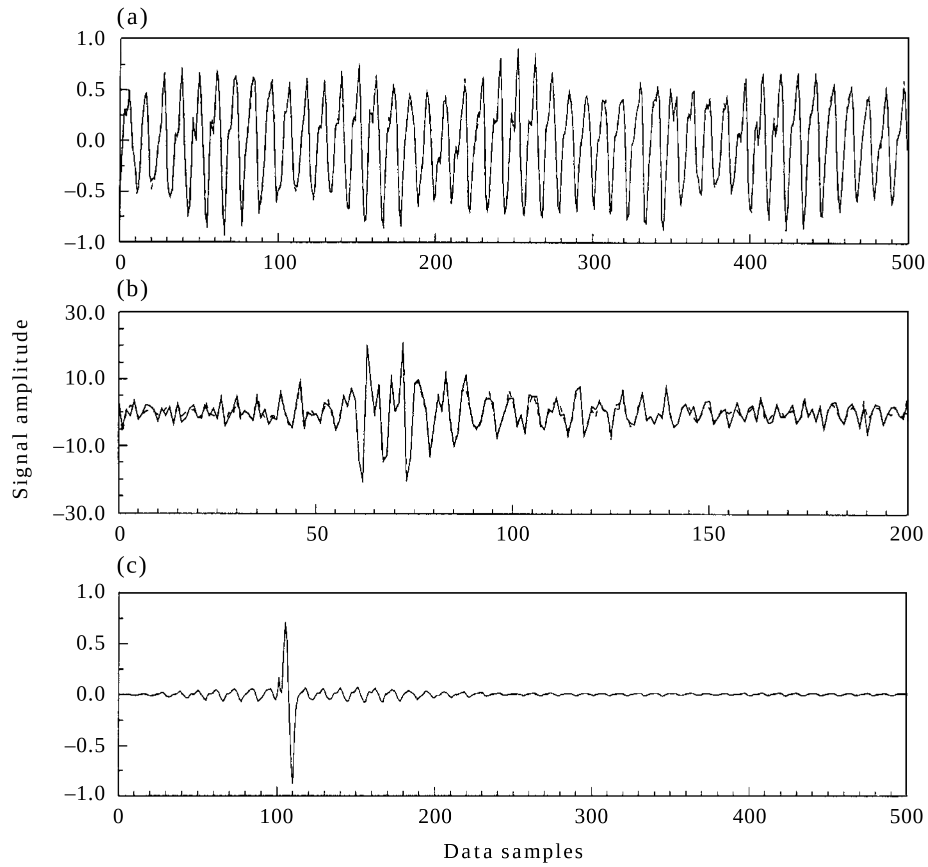

Figure 8. Comparison between original (~~~~) and reconstructed (– – – –) data. Compression ratio used in the analysis, 6:1. (a) Periodic data, MSE = 2·02%; (b) continuous non-stationary data, MSE = 5·03%; (c) transient data, MSE = 0·20%.

was then applied to form the data presented in Figure 10. Here the data is dominated by the impulse which, as previously mentioned, does not indicate the fault. However, the feature of reflection due to delamination can be seen at about 100 μs in Figure 10(b). This feature is not visible in Figure 10(a), where the defect-free section of the plate was analyzed. The feature of reflection can be used for damage detection procedure based on pattern recognition. This procedure requires a significant amount of data for training. It is now obvious that the data, represented by 17 wavelet coefficients, and given in Figure 10 is more suitable for training than the 512 sample vectors given in Figure 9. Further details about the wavelet based feature selection procedure for damage detection in composite materials can be found in references [19, 20].

6.2. A PRIORI KNOWLEDGE OF FEATURES

The wavelet synthesis formula given by equation (9) clearly shows that the analyzed signal can be represented as a sum of m so-called wavelet levels [12]:

$$x_m(t) = \sum_k x_k^m \, g_{m,k}(t). \tag{17}$$

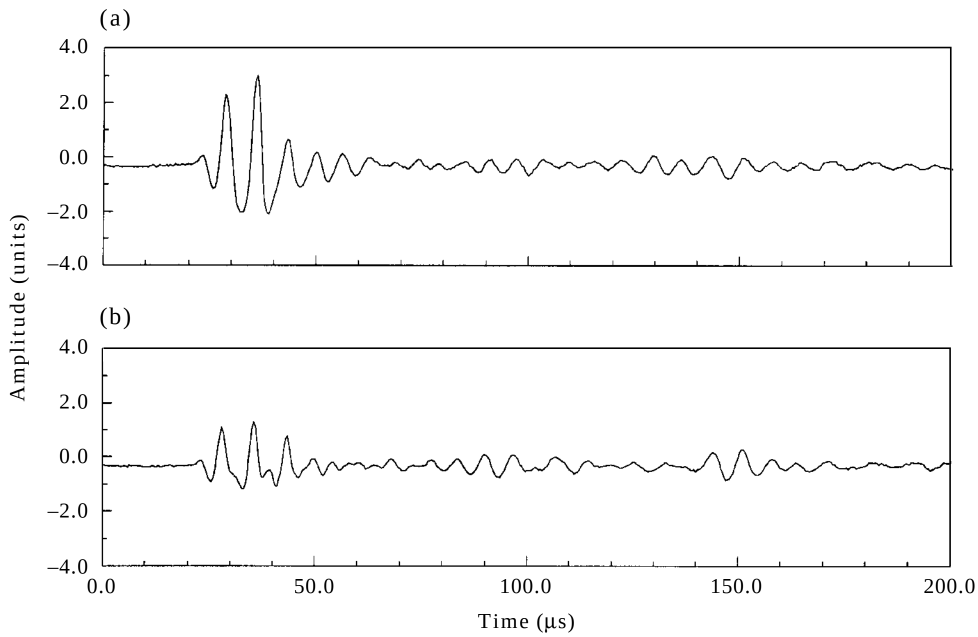

Figure 9. Ultrasonic Lamb wave data: (a) normal condition; (b) delamination.

Each of these levels represents the time behaviour of the signal within different frequency bands. The analysis of wavelet levels and some prior knowledge of a frequency content of required data features can be useful when choosing proper wavelet coefficients for

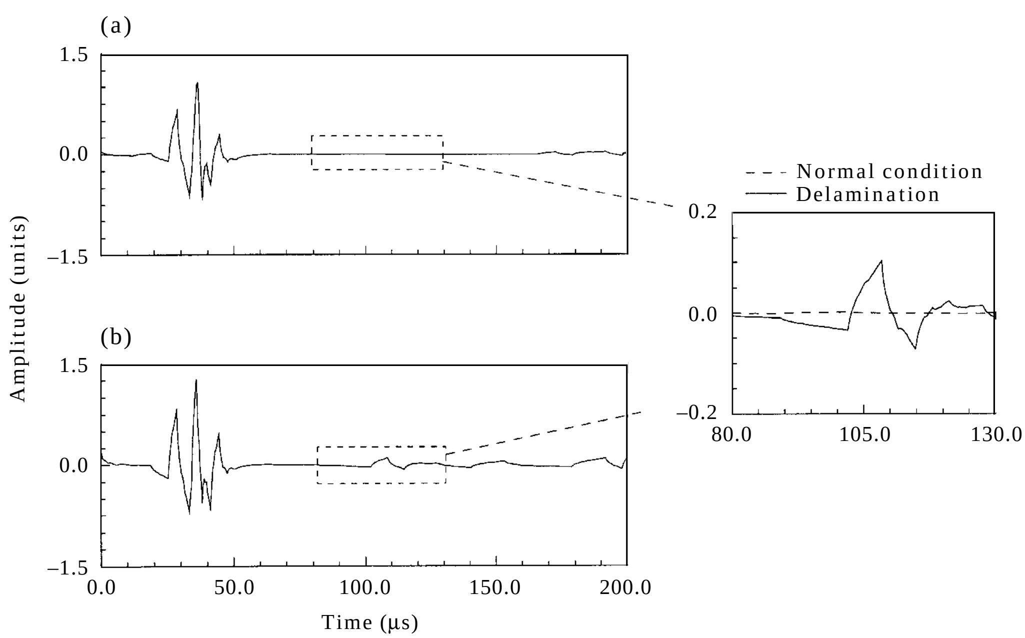

Figure 10. Feature selection in a compressed form of ultrasonic Lamb wave data: (a) normal condition, (b) delamination.

the compression based feature selection procedure. In what follows, an example of fault detection using gearbox data is presented.

It is well known that gearbox vibration data is dominated by the meshing vibration. Local tooth faults (surface wear, cracked tooth) create impulses in vibration data [21, 22]. For a simple fault, an impact appears regularly once per revolution of the damaged wheel. A priori knowledge of rotational and meshing vibration frequencies together with orthogonal decomposition based on wavelet levels, allows for an effective selection procedure for damage detection in gearboxes.

Figure 11 shows spur gear vibration data representing normal (no fault) and fault (broken tooth) conditions. More details about the gearbox data and faults can be found in reference [23]. The wheel rotational and meshing vibration frequencies were equal to 37·5 and 600 Hz, respectively. One gear rotation is given by 172 samples in Figure 11. The impulses created by the damaged tooth are visible at about 120–140° in Figure 11(b). These features can be enhanced using the wavelet based compression procedure. The gearbox data was decomposed using the twentieth order Daubechies' wavelet. Figure 12 shows the result for the data representing the normal condition of the gearbox. Here, the upper vector gives the original data used for decomposition. Figure 13 gives the wavelet decomposed levels in the frequency domain. The frequency analysis was based on 256 samples. Here the upper vector shows the power spectrum of the original data. Four meshing harmonics

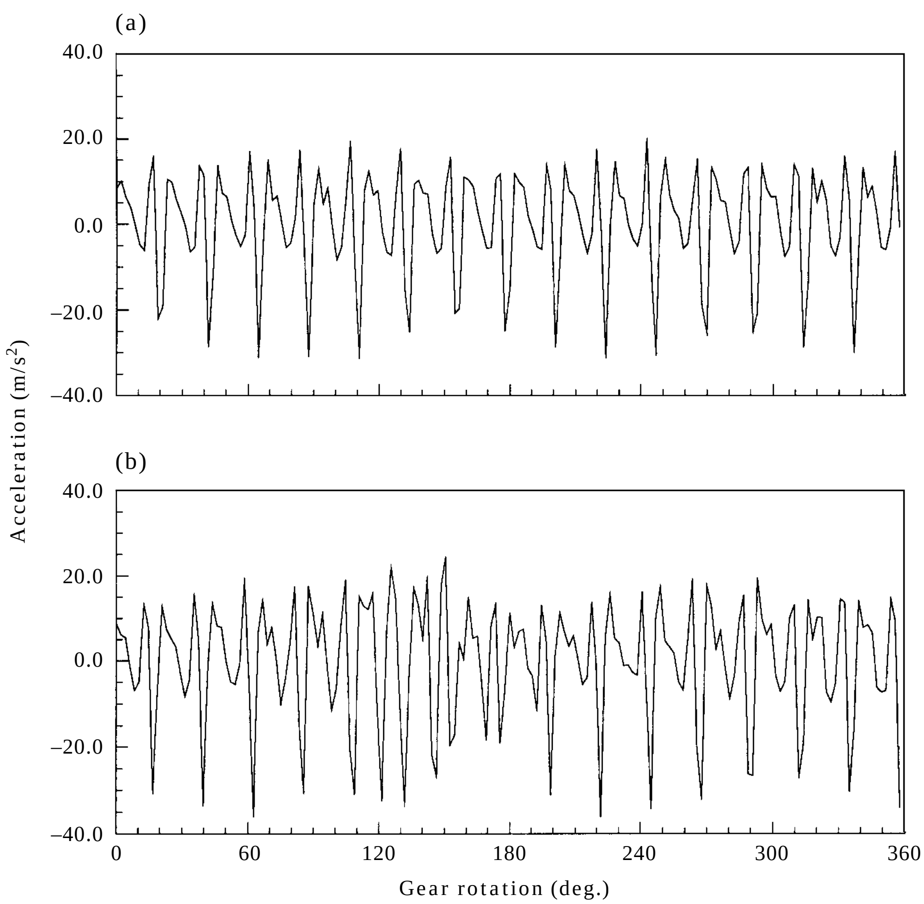

Figure 11. Gearbox vibration data: (a) normal condition, (b) local tooth fault.

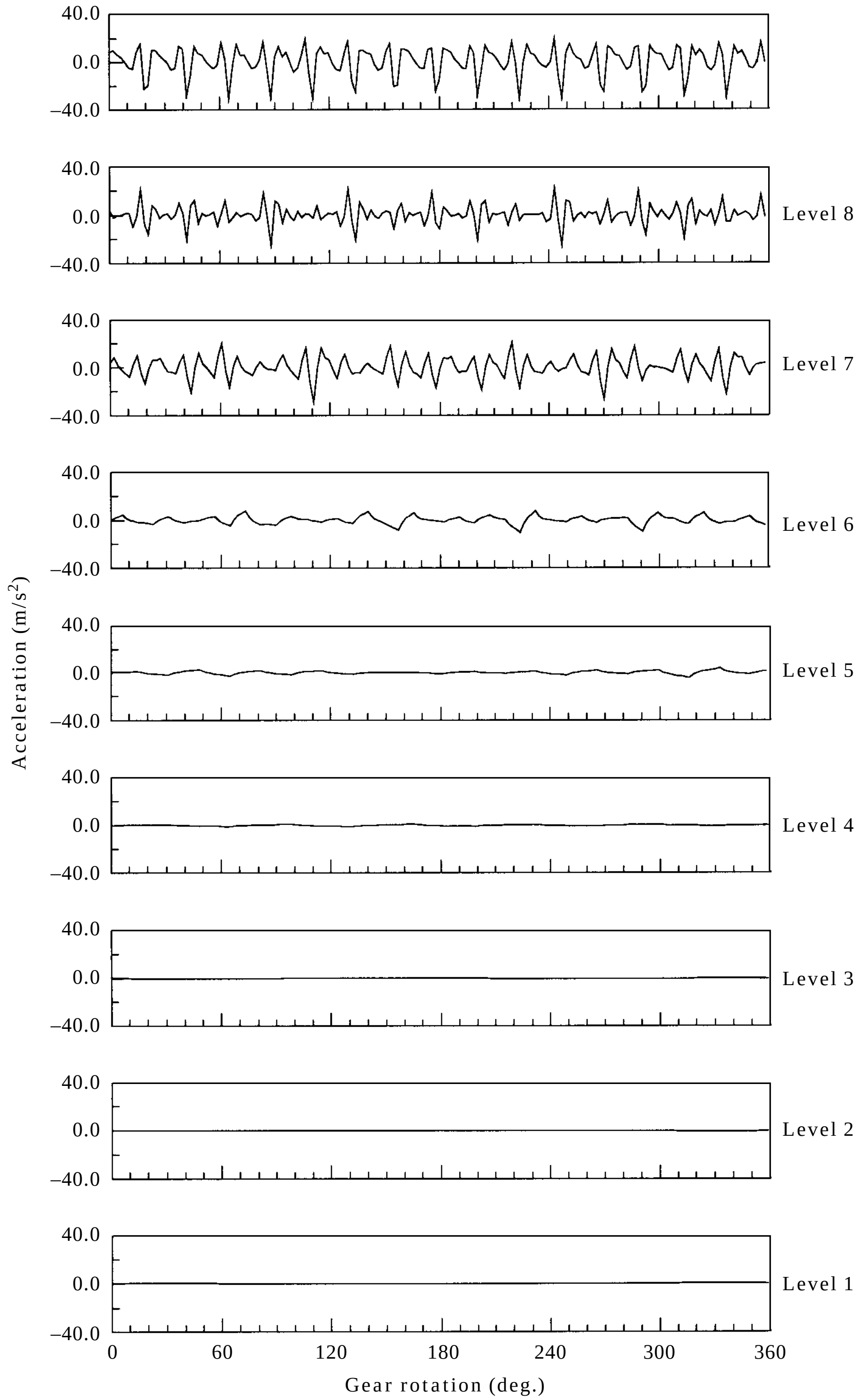

Figure 12. Wavelet decomposition of gearbox vibration data representing normal condition.

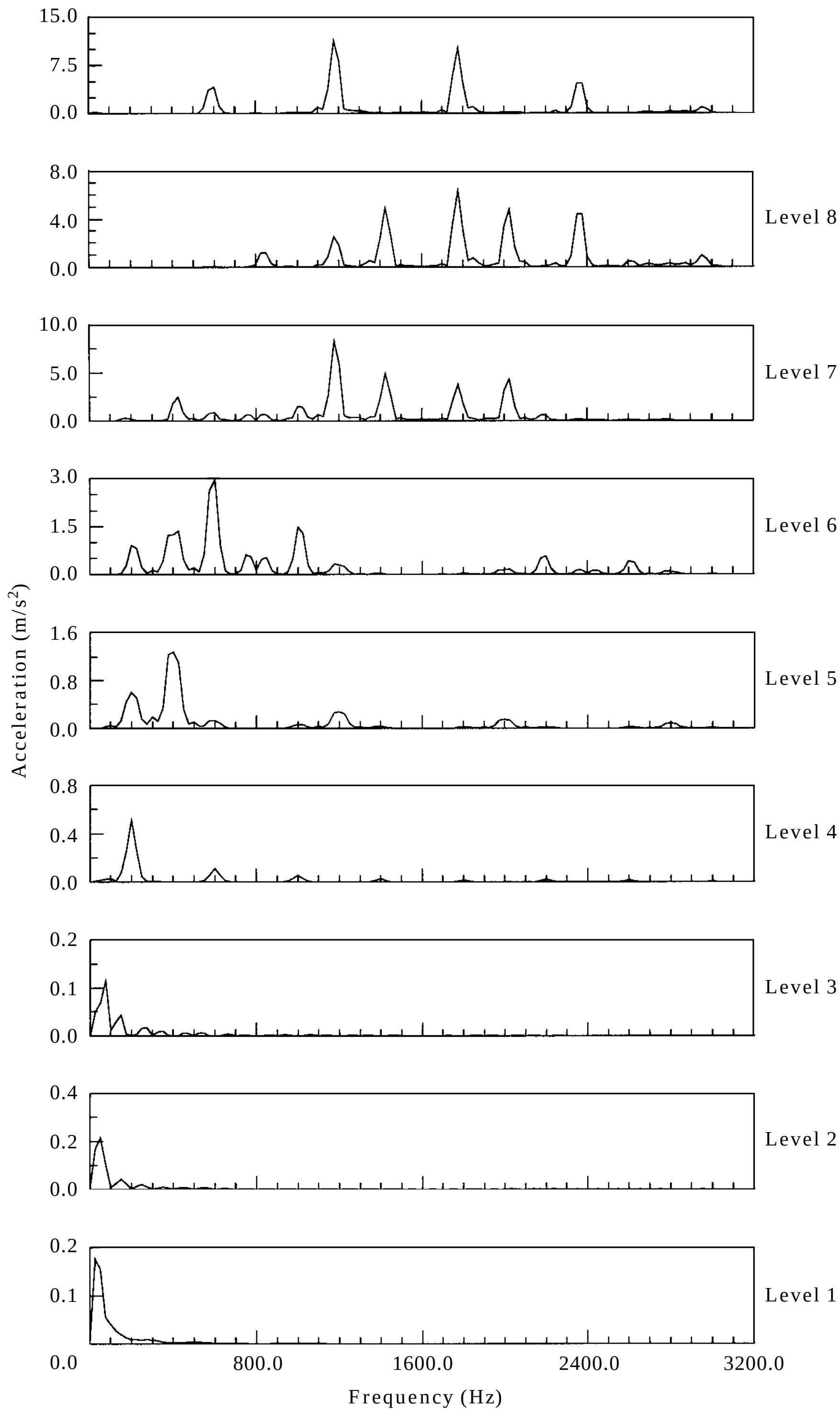

Figure 13. Power spectra of wavelet levels shown in Figure 12.

can be observed in this spectrum. Figure 13 shows that the first meshing harmonics (600 Hz) is predominant in the sixth level of decomposition. Any impact due to the local tooth fault will result in an impulse in gearbox vibration data. These impulses will appear with the rotational frequency harmonics of the damaged wheel (37·5 Hz). As seen in Figure 13, the rotational vibration dominates the three lowest levels of wavelet decomposition. This suggests that any possible local tooth fault will be exhibited by these three levels. Thus the five wavelet coefficients representing the first, second and third levels, together with the maximum five coefficients from the sixth wavelet level, were chosen to form ten coefficients of compressed data for feature selection. On the basis of these coefficients, the data was reconstructed. An example of reconstructed data for analyzed gear conditions is given in Figure 14. Here, the vectors are represented by ten wavelet coefficients. If one compares Figures 11 and 14, the enhancement of fault features and feature selection procedure based on the wavelet compression is now clearly visible. The additional advantage of the procedure is also the fact that the compressed vectors given in Figure 14 are represented by only ten wavelet coefficients in contrast to 172 samples needed to represent the vectors in Figure 11.

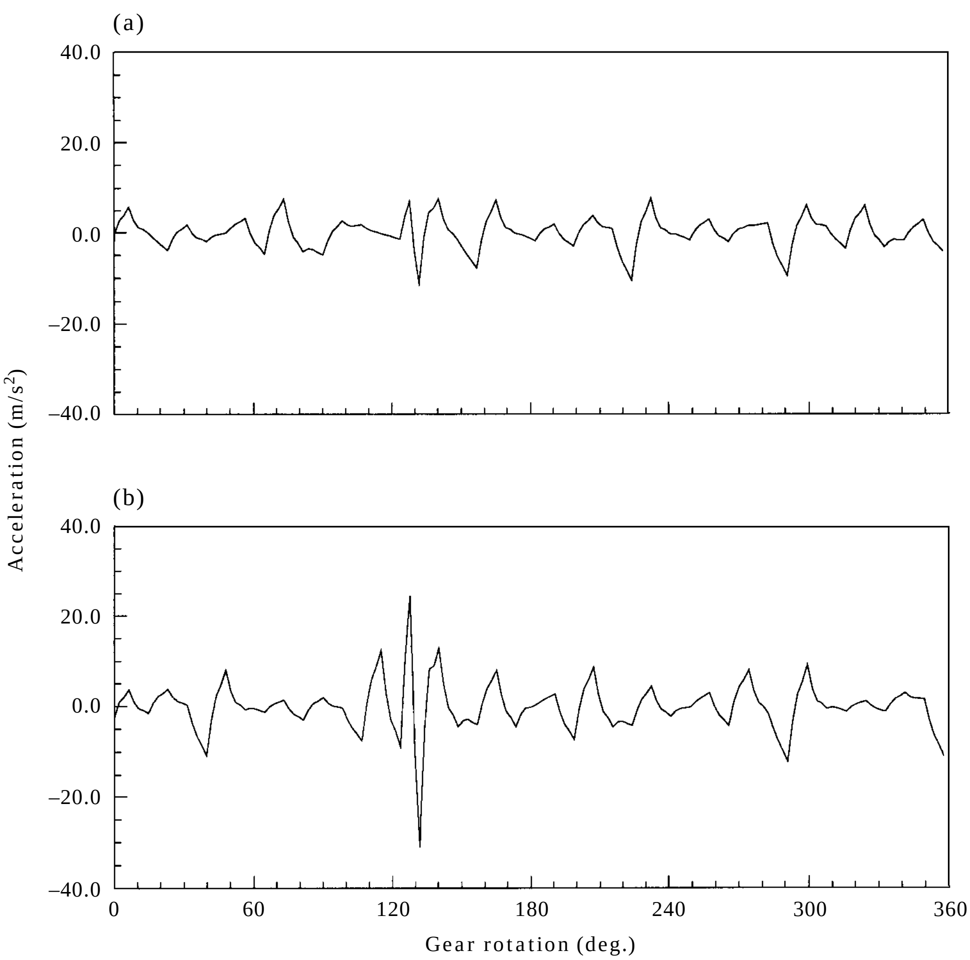

Figure 14. Feature selection in a compressed form of the gearbox vibration data: (a) normal condition, (b) local tooth fault.

6.3. OPTIMAL WAVELET COEFFICIENTS

A simple example of compression with optimal wavelet coefficients involving the spectrum of the gearbox data will be given in this section.

Local tooth faults can display sidebands around meshing harmonics in the power spectrum. Thus, detection of sidebands is important for fault detection in gearboxes. The meshing vibration will be represented by the highest wavelet coefficients and the wavelet based compression algorithm described in section 4 will not keep sidebands with good accuracy. The solution to the problem can be found using the optimization procedure. Thus, instead of thresholding, a genetic algorithm (GA) can be used to select wavelet coefficients for compression. GAs are search procedures based on the mechanisms of natural selection and natural genetics. These procedures use random selection algorithms to do a highly exploitative search through a parameter space. More details about GAs can be found in reference [24].

The simple GA used in this paper employed integer coding, applied reproduction, crossover and mutation operations, and involved new blood and elite chromosomes. An example of compression with optimal wavelet coefficients involves the spectrum of a sample of spur gear data used in the previous section. Figure 15(a) shows the original spectrum of the gearbox data plotted using the solid line. Four predominant meshing harmonics together with the sidebands around the second and the fourth harmonics can be seen. This spectrum was compressed using the algorithm based on the fourth order Daubechies' wavelet and the 25 maximum wavelet coefficients. The aim of the compression was to preserve the amplitude of the sidebands around the second meshing harmonics. The reconstructed data is shown by the long dashed line in Figure 15(a). It can be seen that the thresholding wavelet coefficients procedure does not compress the data properly; the MSE estimated for the frequency bandwidth 800–1150 Hz is equal to 82·0%. The data was

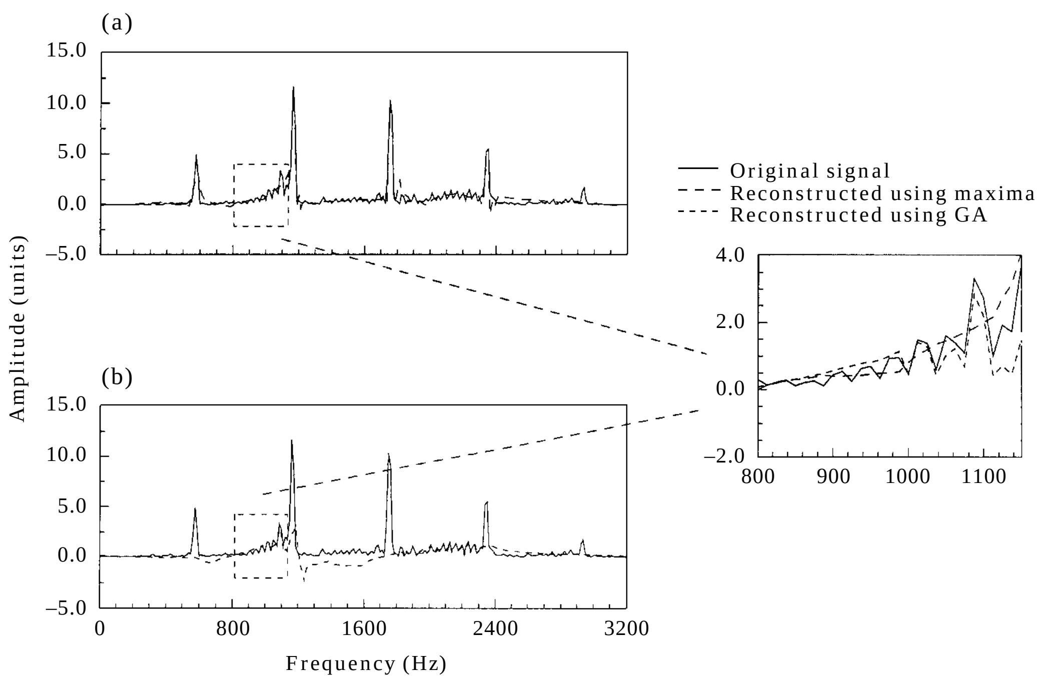

Figure 15. Wavelet compression for fault features selection in gearbox vibration spectra: (a) algorithm based on thresholding, (b) algorithm based on genetic algorithms.

compressed using the same wavelet function and 25 wavelet coefficients determined by a GA. Figure 15(b) gives a comparison between the original data and the reconstructed data using the GA. It can be clearly seen that a simple GA selects wavelet coefficients in such a way that the sidebands around the second components are quite well preserved; the MSE is equal to 24·7%. Finally, Figure 16 shows the comparison between wavelet coefficients used in both compression algorithms. The 25 coefficients selected by a GA in Figure 16(b) are completely different from the 25 maximum coefficients used by thresholding in Figure 16(a).

The simple example presented in this section has shown that vibration analysis often requires a different choice of wavelet coefficients for compression: the position not the absolute amplitude is important. It has to be mentioned that the trade-off between the time required to perform the GA optimization is not substantial. For a given gearbox system, the optimal coefficients can be established in advance from the data representing the normal condition and then used to detect possible faults in the system. More details about compression with optimal wavelet coefficients can be found in reference [25].

6.4. WAVELET COEFFICIENTS TRUNCATED ACCORDING TO THEIR TIME LOCATION

As already explained in section 3, the wavelet transform can be interpreted as a bank of filters, where the frequency partitioning leads to a partitioning in the time domain that is finer in the higher frequency bands. Each wavelet level m given is represented by 2^{m-1} wavelet coefficients which cover different frequency bands. Within each level, the coefficients have time localization properties. This allows one to allocate the wavelet coefficients with given locations in the time domain. Thus, the wavelet coefficients can be truncated not only according to the amplitude but also according to their position in the vector. Keeping the coefficients with properly chosen positions can give good quality compression of specific parts of the data.

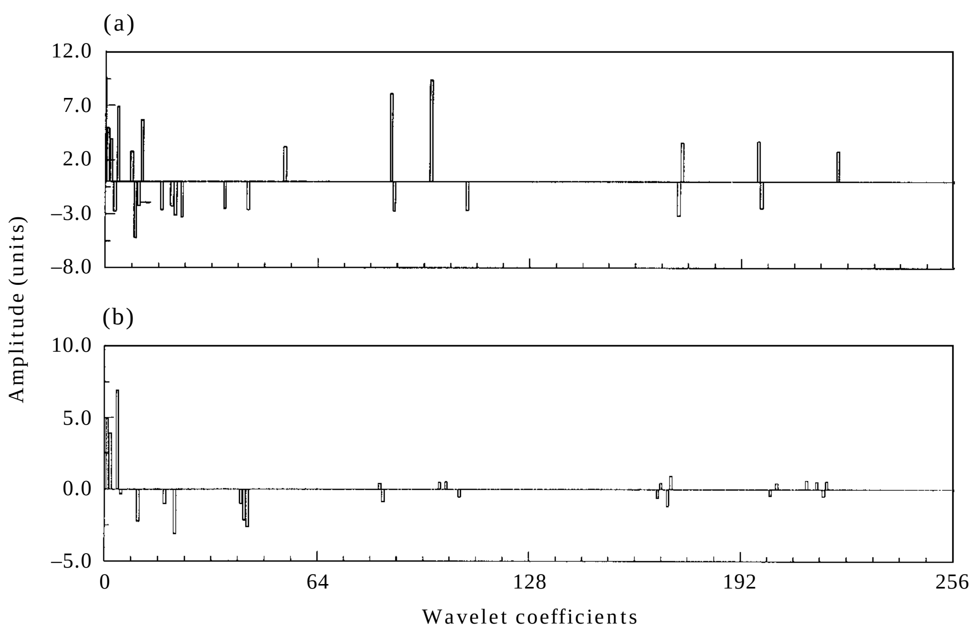

Figure 16. Wavelet coefficients chosen to compress data presented in Figure 15: (a) algorithm based on thresholding, (b) algorithm based on genetic algorithms.

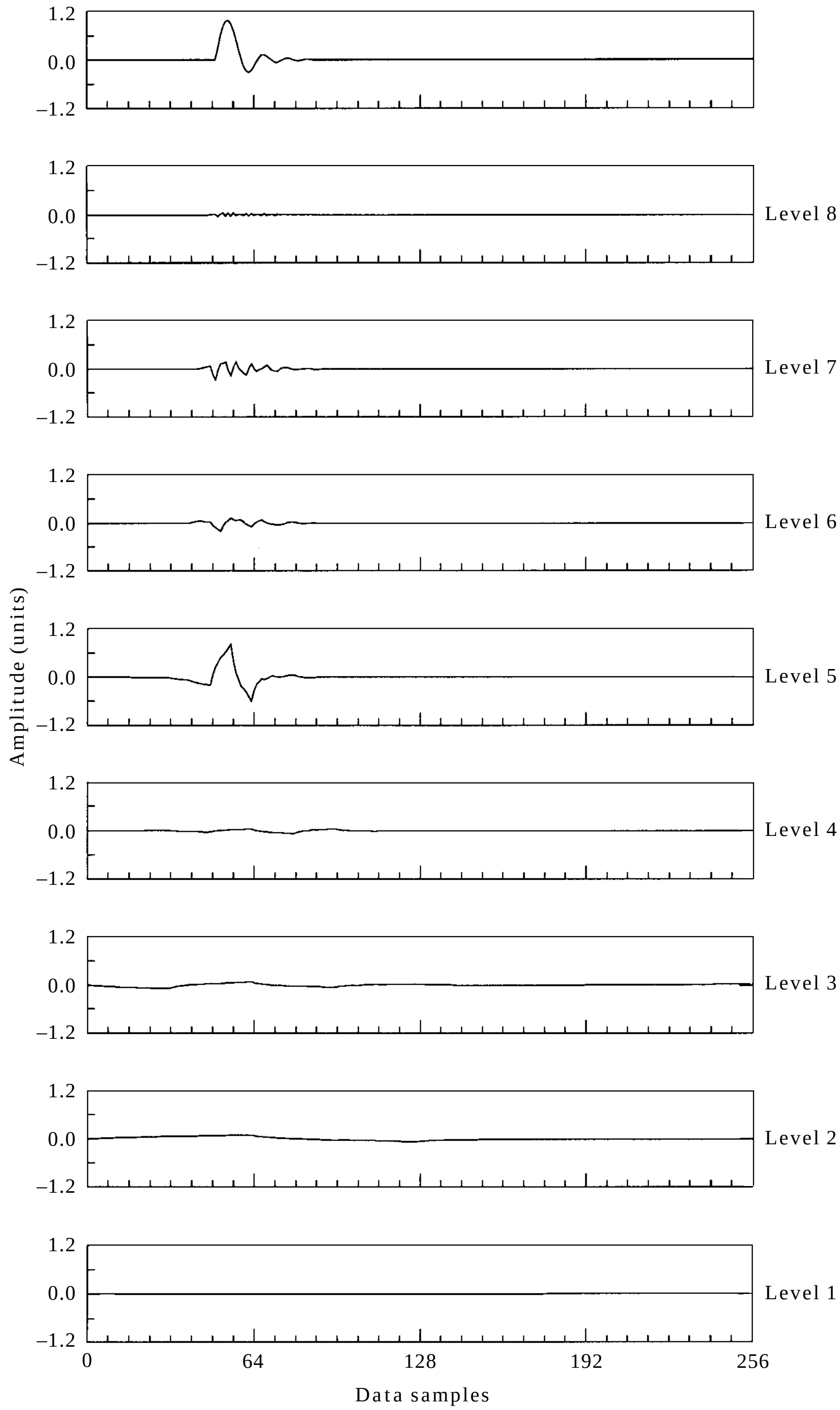

Figure 17. Wavelet decomposition of a logarithmically dampened sine wave transient.

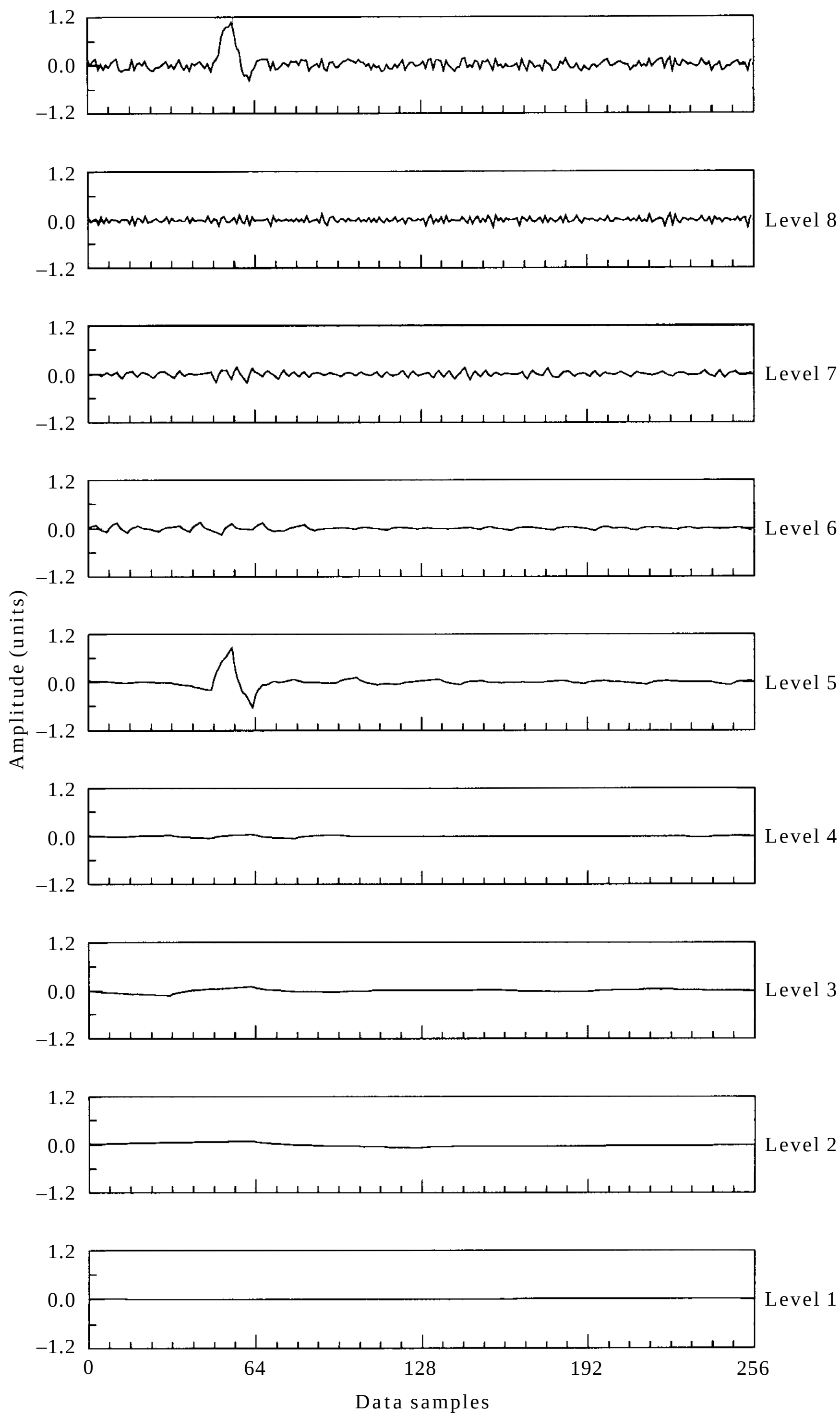

Figure 18. Wavelet decomposition of a logarithmically dampened sine wave transient embedded in the white Gaussian noise.

A simple example is used to show how to compress specific parts of data. This example involves the transient data represented by a logarithmically dampened sine wave. Figure 17 shows the wavelet decomposition of the analyzed transient. Here the plot of the transient is given at the upper part of the figure. It can be seen that the transient is significantly localized in time in the fifth, sixth and seventh levels. The same data was analyzed in the presence of a white Gaussian noise with the standard deviation of $\sigma = 0{\cdot}1$ and $SNR = 10$ dB. The results can be seen in Figure 18, where the noise corrupted transient is given in the upper part of the figure. Figures 17 and 18 show that the transient is significantly localized in time in the fifth wavelet level. Most of the noise can be smoothed out at large levels by amplitude based truncation of wavelet coefficients. However, it is also possible to smooth out the noise at other levels by truncation of wavelet coefficients. It is obvious that the wavelet coefficients representing the transient have to be established. This can be done as follows. The decomposed vector of 256 wavelet coefficients is organized in the hierarchical order, i.e., the eighth level is given by coefficients 129–256, the seventh level is given by coefficients 65–128, etc. The maximum wavelet coefficient, established as number 19, indicates the main feature in the data, namely the transient.

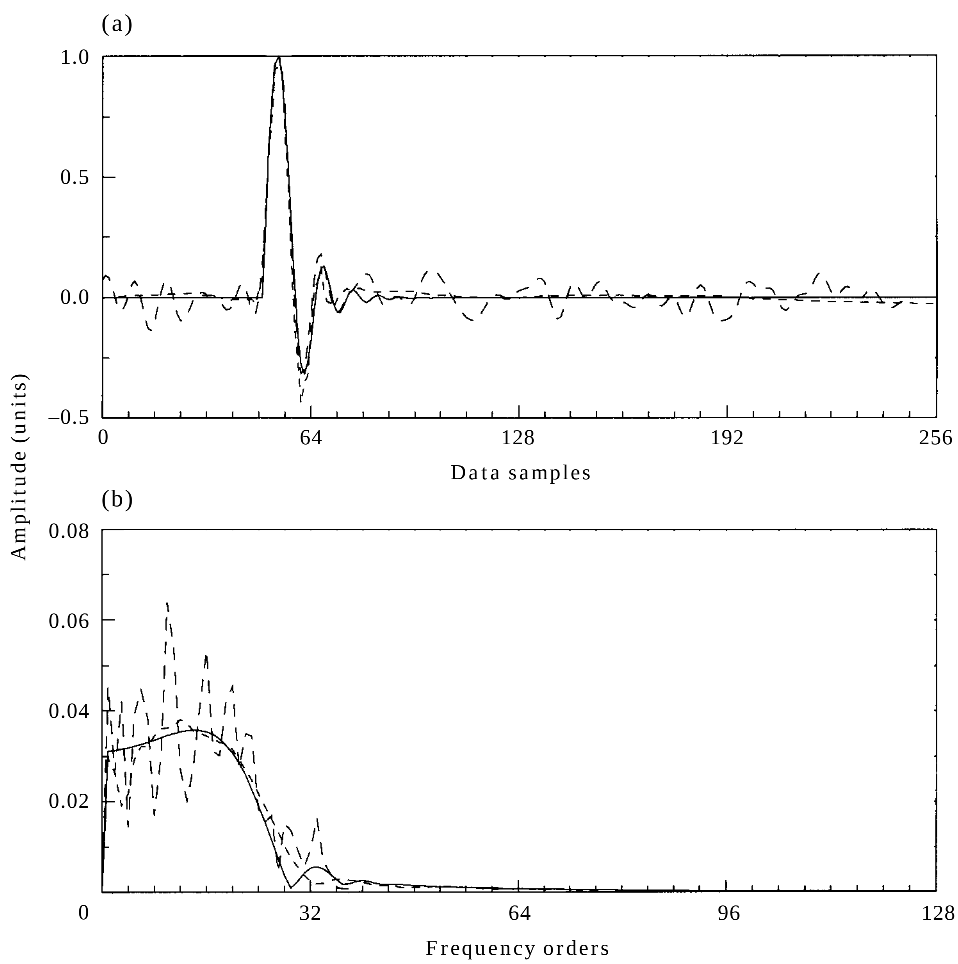

Figure 19. Wavelet compression for fault features selection in transient data: (a) time domain, (b) frequency domain. ——, Original transient; – – – –, reconstructed transient using wavelet-based procedure; — — —, reconstructed transient using filter-based procedure.

The maximum coefficient also localizes the time position of the transient among the coefficients 17–32 which represent the fifth level of decomposition. This time position can also be obtained for other levels, e.g., 38th and 39th coefficient among 33–64 coefficients representing level 6, 76th, 77th, 78th and 79th coefficients among 65–128 coefficients representing level 7, etc. Simply, the higher the level, the bigger the number of coefficients representing the transient and concentrated at the location of the transient. This allows one to select the coefficients representing the transient in different wavelet levels. Following this study, eight wavelet coefficients, namely 9, 19, 38, 39, 76, 77, 78 and 79 were chosen to compress the data and smooth out the noise. Figure 19 shows the result of this procedure. Here, the comparison between the original transient given by the solid line, and the reconstructed transient given by the dashed line is given in the time and frequency domain. It can be seen in Figure 19(a) that the eight chosen wavelet coefficients represent the transient remarkably well; the MSE is equal to 4·9%. One can claim that the similar denoising procedure can be obtained using classical time domain filtration. Figure 19 shows that the tenth order Butterworth filter used does not have the power to remove the noise from the data due to the lack of localization property in the time domain (Figure 19(a)); the MSE is equal to 23·6%. Unwanted low frequency components can still be observed in the filtered data in Figure 19(b). This simple example again shows that very often the position of the wavelet coefficients is important for the feature selection.

7. CONCLUSIONS

Recent developments in wavelet based compression have been reviewed. This includes possible applications of compression in vibration analysis. A number of simple examples have been used to give practical guidance on how to compress different types of data. This analysis has revealed that wavelet based compression is especially effective for non-stationary data. It has been shown that compression in vibration analysis can be used not only for effective storage and transmission of the data, but also for feature selection. A number of different approaches have been presented to show coefficient selection procedures and present wavelet based compression for feature selection. The examples given show that, in contrast to wavelet based compression, the feature selection procedures often use the position together with the amplitude of the wavelet coefficients.

ACKNOWLEDGEMENT

The gearbox vibration data was provided by VSEL Barrow-in-Furness, Cumbria. The Ultrasonic Lamb wave data was captured by Drs Gareth Pierce and Wayne Philp from Strathclyde University, Glasgow. The author is also grateful to Dr Keith Worden for a stimulating discussion and many valuable comments especially in the part of the work involving the genetic algorithm study.

REFERENCES

1. M. Rabbani 1991 *Digital Image Compression*. Bellingham, WA: SPIE Publication, Optical Engineering Press.
2. S. K. Pratt 1991 *Digital Image Processing*. New York: John Wiley and Sons; second edition.
3. R. R. Coifman, Y. Meyer, S. Quake and M. V. Wickerhauser 1989 in *International Conference on Wavelets*, *Marseille*, (Y. Mayer, editor). Paris: Masson. Signal processing and compression with wave packets.
4. T. Ebrahimi and T. R. Reed and M. Kunt 1990 in *Visual Communications and Image Processing*

'90, *Lausanne*, *Switzerland*, 489–502. SPIE Publication. Video coding using a pyramidal Gabor expansion.
5. W. R. Zettler, J. Huffman and D. C. P. Linden 1990 in *Image Processing Algorithms and Techniques*, *Santa Clara*, *CA*, 150–160. SPIE Publication. Application of compactly supported wavelets to image compression.
6. M. V. Wickerhauser 1992 *Wavelets*: *A Tutorial in Theory and Applications* (Ch. K. Chui, editor). Boston: Academic Press, Inc. Acoustic signal compression with wavelet packets.
7. W. Sweldens 1995 in *Wavelets and their Applications in Computer Graphics—SIGGRAPH*; 95 *Course Notes* (A. Fournier, editor). Vancouver: University of British Columbia. Wavelets, signal compression and image processing.
8. M. L. Hilton, B. D. Jawreth and A. Sengupta 1995 *Multimedia systems* **2**. Compressing still and moving images with wavelets.
9. J. Bradley, C. Brislawn and T. Hopper 1993 *Los Alamos National Laboratories*, *Technical reports lq-ur*-1659. The FBI wavelet/scalar quantization standard for gray-scale fingerprint image compression.
10. Ch. K. Chui 1992 *An Introduction to Wavelets*. *Wavelet Analysis and its Applications*, Volume 1. Boston: Academic Press, Inc.
11. S. Mallat 1989 *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI*-**2**. A theory for multiresolution signal decomposition: the wavelet representation.
12. D. E. Newland 1993 *Random Vibration*, *Spectral and Wavelet Analysis*. New York: Longman, Harlow and John Wiley; 3rd edition.
13. I. Daubechies 1992 *Ten Lectures on Wavelets*. Philadelphia, PA: SIAM.
14. Y. Meyer 1993 *Wavelets*. *Algorithms & Applications*. Philadelphia, PA: SIAM.
15. J. M. Shapiro 1993 *IEEE Transactions on Signal Processing* **41**, 3445–3462. Embedded image coding using zerotrees of wavelet coefficients.
16. R. S. Langley 1986 *Journal of Sound and Vibration* Vol. 105, pp. 503–512. On various definition of the envelope of a random process.
17. W. J. Staszewski 1997 in *A Symposium on Time–Frequency and Wavelet Analysis at the Sixteenth Biennal Conference on Mechanical Vibration and Noise*, *ASME Design Engineering Conferences*, *Sacramento*, *CA*. Wavelet novelty measure for machinery diagnostics.
18. W. J. Staszewski, K. Worden and R. Jacob 1997 in *International Conference on Engineering Applications of Neural Networks* (*EANN* '97), *Stockholm*, *Sweden*. Effect on data compression on fault detection networks.
19. S. G. Pierce, W. J. Staszewski, A. Gachagan, I. R. James, W. R. Philip, K. Worden, B. Culshaw, A. McNab, G. R. Tomlinson and G. Hayward in *SPIE's 4th Annual Symposium on Smart Structures and Materials*: *Smart Structures and Integrated Systems Conference*, *San Diego*, *CA*. Ultrasonic condition monitoring of composite structures using a low profile acoustic source and an embedded optical fibre sensor.
20. W. J. Staszewski, S. G. Pierce, K. Worden, W. R. Philp, G. R. Tomlinson and B. Culshaw 1997 Wavelet signal processing for enhanced Lamb wave defect detection in composite plates using optical fibre detection. *Optical Engineering* **36,** 1877–1888.
21. R. B. Randall 1982 *Transactions of the ASME*, *Journal of Mechanical Design* **104**, 259–267. A new method of modelling gear faults.
22. W. J. Staszewski 1994 *PhD Thesis*, *University of Manchester*. The application of time-variant analysis to gearbox fault detection.
23. W. J. Staszewski and G. R. Tomlinson 1994 *Mechanical Systems and Signal Processing* **8**, 289–307, Application of the wavelet transform to fault detection in a spur gear.
24. D. E. Goldberg 1989 *Genetic Algorithms in Search*, *Optimization*, *and Machine Learning*. Reading, MA: Addison-Wesley.
25. W. J. Staszewski 1997 in 2*nd IEE*/*IEEE International Conference on Genetic Algorithms in Engineering Systems*: *Innovations and Applications*, *Glasgow*, pp. 186–190. Data compression with optimal wavelet coefficients.