Automatic Bearing Fault Diagnosis Using Vibration Signal Analysis and Fuzzy Logic

Brahim Harhoud, Messaoud Ramdani
Université Badji-Mokhtar de Annaba
Faculté des Sciences de l'Ingéniorat
Laboratoire d'Automatique et Signaux de Annaba (LASA)
BP. 12, 23000, Annaba, Algeria
messaoud.ramdani@univ-annaba.org

ABSTRACT: *Condition monitoring systems using vibration measurements and supervised classifiers can be used to automate the diagnosis process of rotating machines. In this paper, we describe an automatic diagnosis system for detection and classification of defects in ball bearings using a time varying parametric spectrum estimation method for analyzing nonstationary vibration signals. The classification task is accomplished by an adaptive neural fuzzy inference system. The designed system was developed to be able to classify four types of preestablished defects in ball bearings operating under several shaft speeds and load conditions. The system was tested with experimental data collected from drive end ball bearing of an induction motor driven by mechanical system.*

Keywords: Bearing Fault Diagnosis, AR Modeling, Wavelet Analysis, Fuzzy Classifier

Received: 3 March 2012, Revised 16 April 2012, Accepted 19 April 2012

© 2012 DLINE. All rights reserved

1. Introduction

Identifying the cause of process abnormalities is very important for process supervision. Today's world of highly automated complex machinery requires elaborated decision and advanced condition monitoring systems to truly fulfil the goals of Computer Aided Manufacturing (CAM). Machine failure occurs when a component, structure, or system is unable to accomplish its intended task, resulting in its retirement from usable use. Condition-based maintenance involves the collection and interpretation of data relating to the operating condition of critical components of the equipment, predicting the occurrence of failure, and consequently the determination of appropriate maintenance strategies.

Despite the progress that has been made in the maintenance area, there still a need for further improvements in order to increase the diagnosis accuracy and to reduce the human errors. A considerable amount of research has been carried out previously for the development of many vibration analysis techniques. Most of them use either time or frequency domain representation of vibration signals, on the basis of which many specific features are defined, allowing the recognition with a classification scheme between various operating faulty states.

In recent years, soft computing techniques such as fuzzy logic, neural and neuro-fuzzy networks have been used successfully in pattern recognition tasks and their suitability for fault diagnosis purposes has also been demonstrated. Statistical pattern classification is a traditional technique in which the parameters of the distributions are computed using all training data. Many researchers have paid attention to neural networks classifiers because of the capability of approximation and trainable systems, neural networks with the abilities of real-time learning, parallel computation and self-organisation make pattern classification more suitable to handle complex classification problems. In order to predict the remaining service life of a machine, once a defect is identified, it is further important for the monitoring system to provide an estimate of the defect's severity level. To this end, both feed-forward and recurrent networks have been investigated for artificially induced defects [10].

This paper focuses on the development of a neuro-fuzzy system as a condition monitoring tool. The features are derived from a time-varying parametric spectrum representation. The neuro-fuzzy system is applied to faults recognition. Whatever the adopted approach, the problem to be solved is reduced to the construction of an optimal partition space in which the various classes can be well separated.

The structure of the paper is as follows: Section II gives a brief overview of vibration analysis. Section III presents signal processing methods and feature extraction. Then in section IV, the general scheme for condition monitoring is described. The application of the proposed methodology to the design of ball bearing diagnosis system is presented in section VI. Finally, some concluding remarks are also given.

2. Vibration Analysis

The vibration analysis is based on the processing of the vibration signals; in order to examine the health of the revolving machines, that lies within the scope of the estimated maintenance of the production equipments, the objective of the monitoring of these systems are: (i) Reduction of the number of system stops; (ii) Reliability of the production equipments; (iii) Increase in the availability of the system; (vi) Better managing of the stock.

The mechanical vibrations are oscillating movements around an average position of balance; these movements can be periodic or not periodic (transients or random). The periodic vibrations can correspond to a pure sinusoidal movement or a periodic movement which one breaks up with a nap of elementary sinusoidal movements (harmonic components). Transient vibrations are generated by discontinuous forces (shocks); as for the random vibrations are characterized by a random movement oscillating these vibrations are represented mathematically only by probabilistic relations. The measuring equipment of the vibrations includes:

- Transformation of the mechanical vibration into electric signal,
- Recording of the signal in numerical form,
- Signal analysis.

The chipping of a track of bearing causes shocks and a resonance, and this phenomenon appears in high frequencies.

3. Signal Processing for Condition Monitoring

Since the morphology of a vibration signal varies from incipient faults to complete failure, it is necessary to find out some advanced signal processing techniques, which allow the extraction of a compact feature set, which can still capture most of the useful information inherent in the original signal, is thus very important. Suitable feature extraction methods highlight the important discriminating characteristics of the data, while simultaneously ignoring the irrelevant attributes, i.e., the noise. The common parameters computed in the time domain include the root mean square value (RMS), kurtosis, skew, and crest factor [13], [3], [8]. Since the skew is generally zero for raw vibration data, rectified skew has also been used. This consists of rectifying the vibration signal and then computing the skew for the data. The autoregressive modeling have also been applied to analyzing the spectra of defective bearings for classification purpose [2]. Other bearing characteristic parameters such as BPFO (Ball-pass frequency of outer ring), BPFI (Ball-pass frequency of inner ring) and BSF (Ball spin frequency) have also been used extensively as frequency domain characteristics [7]. They represent the amplitudes at specific frequencies only, thus do not provide a wholesome measure of bearing health. Since time and frequency domain parameters alone have been found to of limited effectiveness for detection of incipient faults, a multi-domain representation combined with an appropriate dimensionnality redution seems to be an effective way to capture vital fault information. In the present study, the features are derived from time-domain, frequency domain and wavelet domain.

3.1 Statistical time-domain features

The time-domain parameters are able to effectively indicate early faults occuring in rotating machinery. Here, seven of them, which are usually used for the fault diagnosis of rotaing machinery, are shown as follows.

1) RMS value x_{rms}

$$x_{rms} = \sqrt{\sum_{n=1}^{N} x(n)^2 / N} \tag{1}$$

2) Skewness value (SK)

$$SK = \frac{\sum_{n=1}^{N} (x(n) - \bar{x})^3}{(N-1)\,\sigma^3} \tag{2}$$

where $\bar{x}$ is the mean value of time series X defined as:

$$\bar{x} = \frac{\sum_{n=1}^{N} x(n)}{N}$$

and σ is the stabdard deviation of the sample X defined as:

$$\sigma = \sqrt{\frac{1}{(N-1)} \sum_{n=1}^{N} (x(n) - \bar{x})^2}$$

3) Kurtosis (KU)

$$KU = \frac{\sum_{n=1}^{N} (x(n) - \bar{x})^4}{(N-1)\,\sigma^4} \tag{3}$$

4) Crest factor (CF)

$$CF = \frac{\max|x(n)|}{\sqrt{\sum_{n=1}^{N} x(n)^2 / N}} \tag{4}$$

5) Clearance indicator (CLI)

$$CLI = \frac{\max|x(n)|}{\left(\sum_{n=1}^{N} \sqrt{|x(n)|} / N\right)^2} \tag{5}$$

3.2 Statistical frequency-domain features

The spectral analysis of a signal can reveal some information that cannot be found in time-domain. The conventional approach using the fast Fourier transform (FFT) cannot handle arbirary and more complex signals. Therefore, the high-resolution spectral estimation can be achieved by the parametric model-based technique, which involves designing a parametric model based on the vibration signal recorded. A frequency spectrum is then generated from this model. In this study, the autoregressive (AR) model is used to estimate the power spectrum density (PSD) of a process and to extract some frequency-domain features. The AR modeling consists in representing the signal $x(n)$ as follows:

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + e(n) \tag{6}$$

where $e(n)$ is white noise with zero mean and variance σ^2, p is the order of the model and a_k are known as the autoregressive coefficients.

The Power Spectral Density (PSD) of an AR process is given by:

$$P(f) = \frac{\sigma^2 / f_s}{\left| 1 + \sum_{k=1}^{p} a_k e^{-j2\pi f_k / f_s} \right|} \tag{7}$$

where f_s is the sampling frequency. The following spectral characteristics are chosen to represent the vibration pattern:

(1) Center of gravity frequency: It is defined as

$$CGF = \frac{\sum_k P(f_k) \times f_k}{\sum_k P(f_k)}$$

where f_k is frequency and $P(f_k)$ is the estimated power spectral density.

2) Spectral Skewness value (SSK)

$$SK = \frac{\sum_{k=1}^{M} (P(k) - \bar{P})^3}{(M-1)\,\sigma^3} \tag{8}$$

where $\bar{P}$ is the mean value of time series P defined as:

$$\bar{P} = \frac{\sum_{k=1}^{N} P(k)}{M}$$

and σ is the stabdard deviation of the sample X defined as:

$$\sigma = \sqrt{\frac{1}{(M-1)} \sum_{k=1}^{M} (P(k) - \bar{P})^2}$$

where M is the number of the frequency points.

3) Spectral Kurtosis (SKU)

$$SKU = \frac{\sum_{n=1}^{M} (P(k) - \bar{x})^4}{(M-1)\,\sigma^4} \tag{9}$$

4) Normalized subband energies : the frequency domain is divided into eight equally spaced frequency subdands B_i, i = = 1, 2, ··· ,8 and the subband energies are computed as

$$E_i = \sum_{f_k \in B_i} |P(f_k)|^2$$

and the normalized subband energies are given by

$$\bar{E}_i = E_i / E_T$$

where the total energy is given by: $E_T = \sum_{i=1}^{8} E_i$

5) The Shannon entropy is computed by

$$ENT = -\sum_{i=1}^{8} \bar{E}_i \ln \bar{E}_i$$

As a result, 12 statistical frequency-domain features are computed.

3.3 Wavelet analysis

The theory of wavelet transform is a coherent mathematical framework for several signal processing techniques, such as the multi-resolution analysis applied in the field of computer vision and data compression [12]. In this work, the wavelet transform will be used for time-frequency analysis, because it allows to separate the signals of high frequency such as the impacts of bearing from the vibrations, and low frequency, such as the rotational frequency of the machine. The wavelet transform is defined as the integral of the signal $s(t)$ multiplied by scaled, shifted versions of a basic wavelet function $\psi(t)$:

$$c(a, b) = \int_R s(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) dt, \; a \in R^+ - \{0\}, \; b \in R \tag{10}$$

Where a is the so-called scaling parameters, b is the time localisation parameter. Associated with wavelet ψ, which is used to define the details in the decomposition, a scaling function ϕ is used to define the approximations. To avoid intractable computations of the continuous wavelet transform (CWT), scales and positions can be chosen based on a power of two, i.e., dyadic scales and positions. The discrete wavelet transform (DWT) analysis is more efficient and accurate [4]. In this scheme a and b are given by:

$$(j, k) \in Z^2 : a = 2^j, b = k2^j, Z = \{0, \pm 1, \pm 2, \ldots\}$$

This allows us to define the scaled and shifted versions of the mother wavet and teh mother scaling function, respectively.

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k) \; \phi_{j,k}(t) = 2^{-j/2} \phi(2^{-j}t - k)$$

A wavelet filter with impulse response g, plays the role of the wavelet ψ, and a scaling filter with impulse response h, plays the role of scaling function ϕ. Thus, the DWT can be described mathematically as:

$$c(j, k) = \sum_{n \in Z} s(n) \, g_{j,k}(n)$$

$$a = 2^j, b = k\,2^j, j \in N, k \in N$$

The details at level j is defined as:

$$D_j(t) = \sum_{k \in Z} c(j, k) \, \psi_{j,k}(t)$$

and the approximation at level L:

$$A_{L-1} = \sum_{j > L} D_j$$

In practice, the decomposition can be determined iteratively, with successive approximations being computed, such that the analysed signal is decomposed into many lowerresolution components. At each level of the decomposition, the signal can be reconstructed by using the reconstruction filters and upsampling [4], [12].

In the present study, the Daubechies wavelet "*db*1" is used to compute a five level DWT. For a given signal $s(t)$, the DWT feature vector is defined as

$$v = [v_1, v_2, \ldots, v_6]^T$$

where $v_i = \sigma_i / \sigma_{ir}$, $i = 1, 2, \ldots, 6$, corresponds to $D1, \ldots, D5, A5$, respectively and σ_i is the standard deviation of the i^{th} decomposition; σ_{ir} is the standard deviation of the i^{th} decomposition of a reference signal.

4. General Scheme for Condition Monitoring

The neuro-fuzzy fault diagnosis can be viewed as a pattern recognition problem. The spirit of pattern recognition techniques is to solve the problem via the most significant features [5]. Therefore, in neural fault diagnosis the *symptoms* serve as the input patterns to the recognition system which identifies the faults by the analysis of the input features. The classifier is carried out by an an adaptive fuzzy inference system.

4.1 Neuro fuzzy system

Assume we have a complex nonlinear multi-input and multi-output (MIMO) relationship where $x = [x_1, \ldots, x_n]^T \in \mathscr{X} \subset \mathscr{R}^n$ is the vector of input variables and $y \in \mathscr{Y} \subset \mathscr{R}^m$ is the vector of ouput variables. In the multi-input and multioutput (MIMO) neuro-fuzzy system given in Figure 1, the overal output is defined as:

$$\hat{y}_j(x) = \sum_{i=1}^{H} f_{ij}\phi_i \Big/ \sum_{i=1}^{H} \phi_i \qquad (11)$$

where, $j = 1, 2, \ldots, m$; $i = 1, 2, \ldots, H$ and $l = 1, 2, \ldots, n$;

$$\phi_i = \prod_{l=1}^{n} \exp\{-(x_l - c_{il})^2/(\sigma_{il})^2\} \qquad (12)$$

Here, we assume that $c_{il} \in \mathscr{X}_i$, $\sigma_{il} > 0$ and $f_{ij} \in \mathscr{Y}_j$, where $\mathscr{X}_i$, and $\mathscr{Y}_j$ are the variation domains of the input x_i and output y_j, respectively.

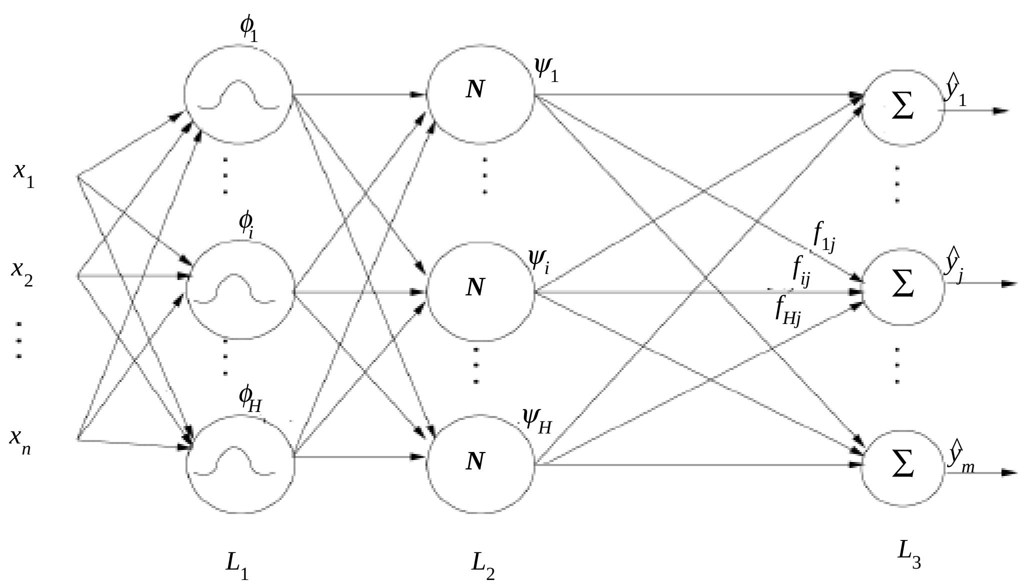

Figure 1. Architecture of the neuro-fuzzy classifier

5. Hybrid Learning Scheme

The learning process is performed in two phases. Firstly, a clustering algorithm is used to find a coarse model that roughly approximates the underlying input-output relationship. Secondly, parameter optimization procedure is performed for a better tuning of the initial structure. In principle, once an appropriate structure is identified, the learning task can be acomplished by any suitable training algorithm such as the standard backpropagation algorithm (BPA). However, because of slow convergence speed of pure BPA, in the following a more efficient training method, namely the combination of gradient descent with least squares optimization procedure will be used.

5.1 Structure identification by clustering

From the available training data that contain N inputoutput samples, a regression matrix X and an output matrix Y are constructed

$$X = [x_1, \cdots, x_N]^T, Y = [y_1, \cdots, y_N]^T \tag{13}$$

Since the study under consideration deals with a *classification* task, the clustering process uses only the input portion of the data is used to discover an underlying structure generating the data. Clustering is a technique to partition a set of samples into exactly c disjoint subsets. Samples in the same cluster are somehow similar than other samples in other clusters. One way to make this problem a well defined one is to define a criterion function that measures the clustering quality of any partition of the data.

Among the clustering methods, the Fuzzy c-means algorithm is one of the most popular. In FCM method, the loss (objective) function is defined as follows:

$$J_m = \sum_{i=1}^{c} \sum_{k=1}^{N} u_{ik}^M d_{ik}^2 \tag{14}$$

where, $d_{ik} = \| x_k - v_i \|$ and u_{ik} denotes the grade of membership input vector k to fuzzy cluster i, v_i is interpreted as prototype of cluster i defined by $\{u_{ik}\}$, and weighting exponent m controls the extent of membership sharing between fuzzy clusters. For $m = 1$, FCM converges in theory to the traditional c - means solution. To minimize Equation (14) subject to normalized condition:

$$\sum_{i=1}^{c} u_{ik} = 1 \tag{15}$$

for each input vector k, using the Lagrangian multiplier method, for $m > 1$, local minimum of Equation (14) was demontrated if and only if

$$u_{ik} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^{2/(m-1)}} \tag{16}$$

$$v_i = \frac{\sum_{k=1}^{N} u_{ik}^m x_k}{\sum_{k=1}^{N} u_{ik}^m} \quad \forall i \tag{17}$$

The FCM algorithm is characterized the parameter m that determines the behavior of the clustering algorithm. The larger m is, the fuzzier is the partition.

5.2 Parameter optimization procedure

The parameters obtained by the identification procedure can be optimized or fine tuned by a variant of gradient descent optimization techniques. This is achieved by an iterative two stage forward-backward optimization algorithm. In the forward stage, with the ellipsoidal basis functions being constant, the weights of the last layer, i.e the functional models f_{ij} $i = 1, \cdots, H$ and $j = 1, \cdots, m$ are identified by solving a least squares problem. Then, in the backward stage, the functional models are fixed and the parameters of the ellipsoidal functions c_{il}, σ_{il}, $i = 1, \cdots, H$; $l = 1, \cdots, n$ are updated by an effective nonlinear gradient-descent (GD) optimization technique, which requires the computation of the derivatives of the objective function to be minimized with respect to the parameters c_{il} and σ_{il} .

The optimization algorithm uses a variable step learning rates. Given a set $\mathscr{D} = \{(x^p, d^p)\}_{p=1}^{N}$, such that $x^p \in \mathscr{X} \subset \mathscr{R}^r$, $d^p \in \mathscr{Y} \subset \mathscr{R}^m$; the objective is to find sub-systems $\hat{y}_j(x^p)$ occurred. The parameters are then updated by in the form of (11), such that the mean squared error (MSE) function

$$E = \frac{1}{2} \sum_{\substack{i=1 \\ x^p \in D}}^{m} \left(\hat{y}_j - d_j^p\right)^2 \tag{18}$$

is minimized. The problem is reduced to the adjustment of the f_{ij}, and the mean (c_{ij}) and variance σ_{il} of the ellipsoidal functions, so that the MSE is minimized.

Now it can be seen that the network output $\hat{y}_j$, and hence E, depends on c_{il} and σ_{il} only through ϕ_i, where $\hat{y}_j$, f_{ij} , b and ψ_i are represented by the following equations:

$$\hat{y}_j = \sum_{i=1}^{H} f_{ij}\, \psi_i \tag{19}$$

$$\psi_i = (\phi_i / b) \text{ and } b = \sum_{i=1}^{m} \phi_i \tag{20}$$

Derivatives of E w.r.t c_{il} and σ_{il}

$$\frac{\partial E}{\partial c_{il}} = \frac{\partial E}{\partial \phi_i}\frac{\partial \phi_i}{\partial c_{il}} = \sum_{j=1}^{m}\left(\frac{\partial E}{\partial \hat{y}_j}\frac{\partial \hat{y}_j}{\partial \phi_i}\right)\frac{\partial \phi_i}{\partial c_{il}} \tag{21}$$

$$\frac{\partial E}{\partial \sigma_{il}} = \frac{\partial E}{\partial \phi_i}\frac{\partial \phi_i}{\partial \sigma_{il}} = \sum_{j=1}^{m}\left(\frac{\partial E}{\partial \hat{y}_j}\frac{\partial \hat{y}_j}{\partial \phi_i}\right)\frac{\partial \phi_i}{\partial \sigma_{il}} \tag{22}$$

Finally, the results of the chain rules are written as follows:

$$\frac{\partial E}{\partial \sigma_{il}} = A \cdot \{\, 2 \cdot \phi_i \cdot (x_l - c_{il}) \,/(\sigma_{il})^2 \,\} \tag{23}$$

$$\frac{\partial E}{\partial \sigma_{il}} = A \cdot \{\, 2 \cdot \phi_i \cdot (x_l - c_{il}) \,/\, (\sigma)^3 \,\} \tag{24}$$

with $A = \left(\sum_{j=1}^{m} (\hat{y}_j - d_j).(f_{ij} - \hat{y}_j) / b \right)$

Let us assume that the task is to find an optimal vector of parameters w which minimizes some objective function $J(w)$. In the case of the neural system under consideration, all the parameters defining the ellipsoidal functions are stacked in a single vector w. The optimization algorithm is a variant of gradient descent in which each parameter w_j has its own step size η_j , and the step sizes are adapted during the optimization process, depending on the learning performance and more specifically on the progress of the objective function and on the sign of its derivatives at successive iterations. Let t be the iteration index. Then, if the objective function has decreased between iteration t -1 and t, the following rule is applied to update each step size η_j

$$\eta_j(t) = \begin{cases} \beta\eta_j(t-1), \; if \; \dfrac{\partial J}{\partial w_j}(t-1) \cdot \dfrac{\partial J}{\partial w_j}(t) > 0 \\ \gamma\eta_j(t-1), \quad otherwise \end{cases}$$

where $\beta > 1$ and $\gamma < 1$ are two coefficients. Hence, the step size is increased if the derivatives have kept the same sign during two iterations, and it is increased if the sign of the derivative has changed, because a jump over a minimum has occured. The parameters are then updated by

$$w_j(t+1) = w_j(t) - \eta_j(t)\frac{\partial J}{\partial w_j}(t)$$

If now the objective function has increased between iterations t -1 and t, all step sizes are decreased simultaneously

$$\eta_j(t) = \delta\eta_j(t-1) \; \forall j$$

$$w_j(t+1) = w_j(t-1) - \eta_j(t) \; \frac{\partial J}{\partial w_j}(t)$$

For the study under consideration, the following numerical empirical values of the coefficients are used:

$$\beta = 1.2 \; \gamma = 0.8 \; \delta = 0.5$$

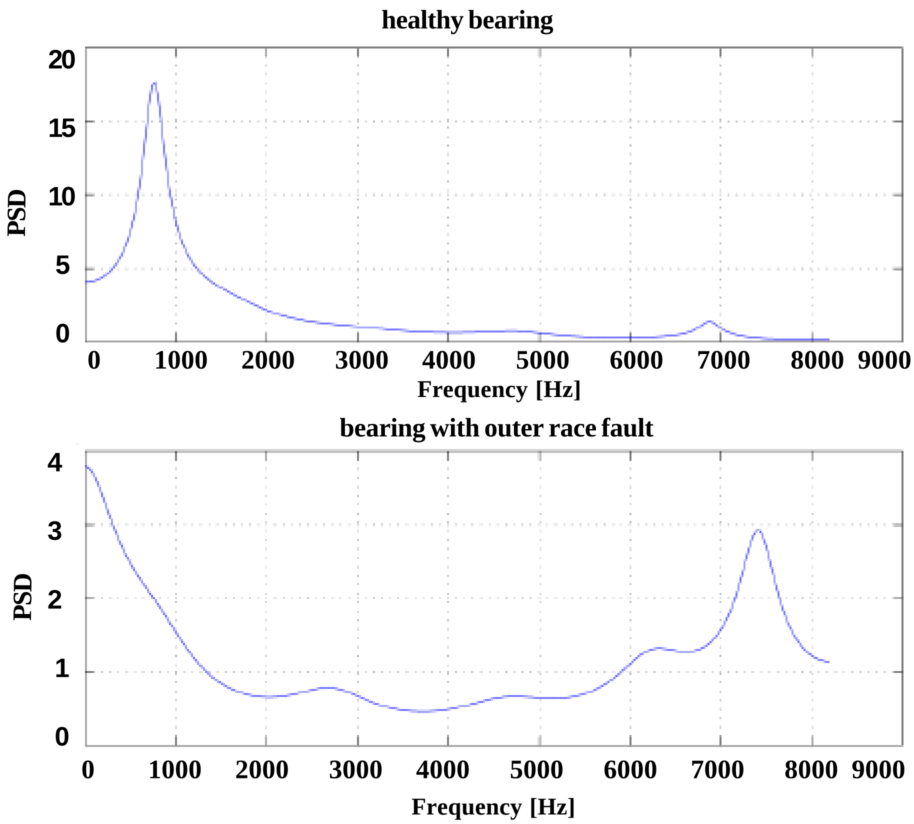

Figure 2. Parametric spectral analysis of vibration signals

Type	size
time-domain	5
frequency-domain	12
wavelet-domain	6
multi-domain	23

Table 1. The Various Parameters Extracted From a Vibration Signal

Method	Training	Testing
time-domain	0.7	0.85
frequency-domain	0.8	1.1
wavelet-domain	0.8	1.1
multi-domain	0.54	0.6

Table 2. Average Error in % For Training and Testing Dataset

6. Experimental Results

To demonstrate the effectiveness of the proposed condition monitoring scheme, data was collected experimentally from both healthy and defective bearings. The database treated represents five classes one of operation normal and four represent faults in bearings.

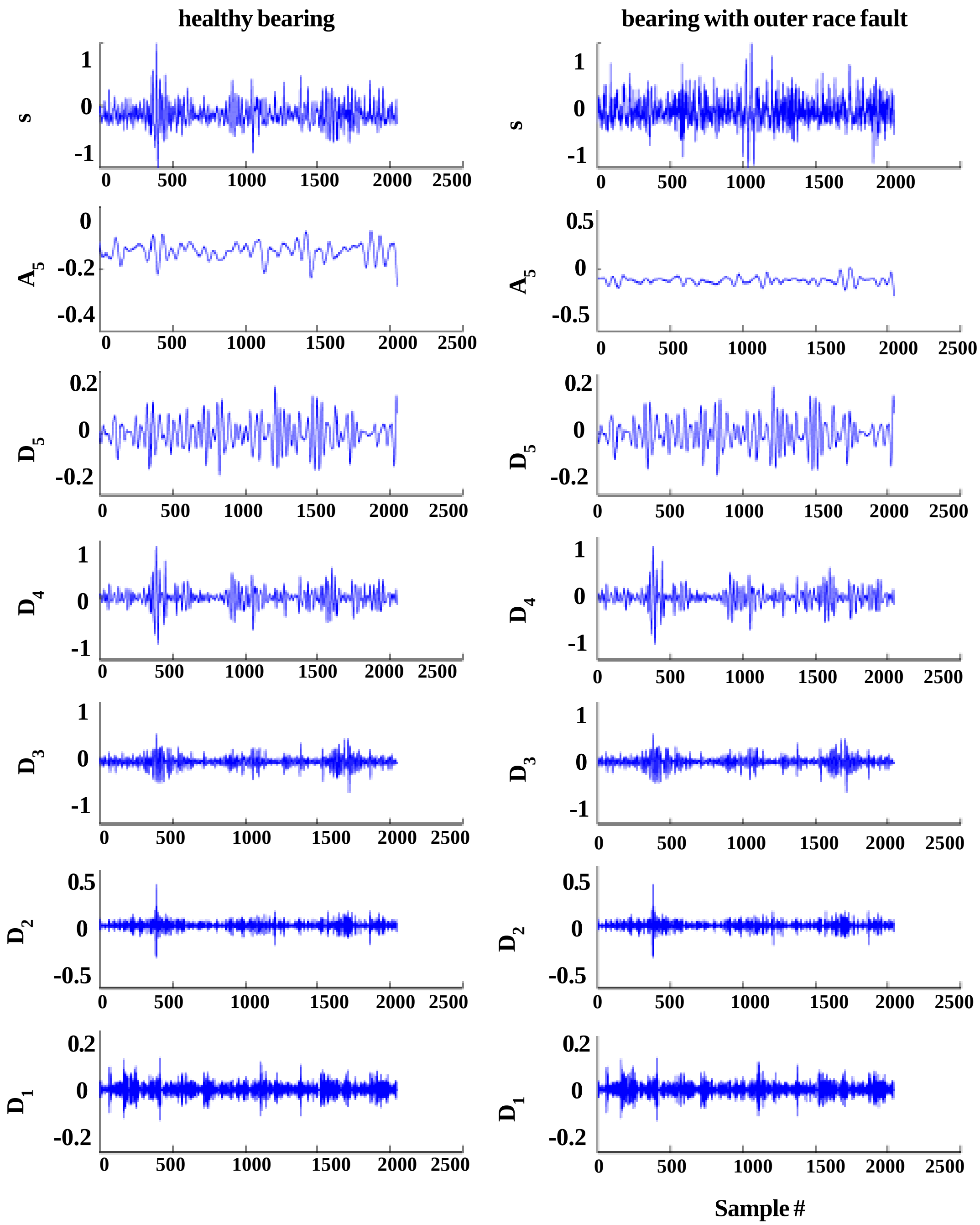

Figure 3. Wavelet decomposition of vibration signals

l1 - new ball-bearing

l2 - outer race completely broken

l3 - broken cage with one loose element

l4 - damaged cage, four loose elements

l5 - no evident damage; badly warned ball-bearing

Ball-bearing type 6204 (steel cage)

Rotational frequency = 24.5625 Hz (Tacho-signal used for the measurement)

Measured signal: acceleration (m/s2)

Sampling frequency: 16384 Hz

Minimum frequency: 0.7 Hz

Data acquisition system (B & K analyser)

Each file consists of 10-12 vectors including 2048 samples.

Figure 2 gives the parametric spectral for two vibration signals. The difference between different conditions indicate the efficiency of the features that are derived from the frequency domain. Figure 3 shows the results of wavelet decomposition of vibration signals. As a comparison, the neuro-fuzzy system was trained using features derived from time domain, parametric spectral analysis and wavelet transform (WLT) as indicated in table I. The available input vectors being classified have been divided in two groups: one used for the learning purposes and the other for testing the network. This was done to test the generalization ability of the network as in practical case it is not possible to have defect vibration data for all possible defect and operating conditions.

The results of the II show the relative effectiveness of the various feature sets for the fuzzy classifier. The best classification rate of 99.46% is obtained for composite feature vector combining the three domain parameters

7. Conclusions

In this paper, an automatic diagnosis system for condition monitoring of rotating machines has been developed. Various defects, which have mechanical origin, are detected by the analysis of the vibration signals recorded for these bearings under different operating conditions. Neuro-fuzzy systems can be used successfully to detect faults in rotating machinery, using statistical estimates of the vibration signals as input features.

Referenses

[1] Alguindigue, I., Loskiewics-Buczak, A., Uhrig, R. (1993). Monitoring and diagnosis of rolling element bearings using artificial neural networks, *IEEE Trans on Industrial Electronics*, 40(2) 209-216.

[2] Baillie, D. C., Mathew, J. (1996). A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings, *Mech. Syst. Signal Process.*, 10, 1-17.

[3] Li, Y., Billington, S., Zhang, C., Kurfess, T., Danyluk, S., Liang, S. (1996). Adaptive prognostics for rolling element bearing condition, *Mech. Syst. Signal Process.*, 10, 1-17.

[4] Daubechies, I. (1992). Ten Lectures on Wavelets. Philadelphia, PA: Society for Industrial and Applied Mathematics.

[5] Duda, R., Hart, P., Stork, D. (2001). Pattern Classification, 2nd ed: Wiley- Interscience

[6] Haykin, S. (1998). Neural networks: A Comprehensive Foundation, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall.

[7] Harris, T. (1991). Rolling bearing analysis, 3rd ed., New York: Wiley.

[8] Honarvar, F., Martin, H. (1997). New statistical moments for diagnosis of rolling element bearings, *ASME Journal of Manufacturing Science and Engineering*, 119, 452-432.

[9] Liu, T., Mengel, J. (1992). Intelligent monitoring of ball bearings, *Mechanical Systems and Signal Processing*, 6 (5) 419-431.

[10] Li, C., Fan, Y. (1999). Recurrent neural networks for fault diagnosis and severity assessment of a screw compressor, *ASME J. Dynamic Syst. Contr.*,121, 724-728.

[11] Lou, X., Loparo, K. A. (2004). Bearing fault diagnosis based on wavelet transform and fuzzy inference, *Mech. Syst. Signal Process.*, 18, 1077-1095.

[12] Mallat, S. (1998). A wavelet tour of signal processing, 2nd ed., San Diego, CA: Academic.

[13] Shao, Y., Nezu, K. (2000). Prognosis of remaining bearing life using neural networks, *In*: *Proc. Inst. Mech. Eng.-J. Syst. Contr. Eng.*, 214 (3) 217-230.