Asian Journal of Information Technology 6 (3): 342-347, 2007

Diagnosis of Ball Bearing Faults Using Wavelet Analysis and Hidden Markov Models (HMM)

Salah Chenikher, Messaoud Ramdani and Bedda Mouldi
LASA Laboratory, Badji Moukhtar University, BP. 12, Annaba 23000, Algeria

Abstract: In this study, a condition monitoring system for fault diagnosis of ball bearings in rotating machines was developed. Features extraction is based on the relevant information calculated from the vibration signal by wavelet transform. The faults diagnosis procedure is achieved by Hidden Markov Models and uses the wavelet feature as inputs to the HMM. This procedure includes training of the HMM and faults recognition by choosing the model that gives maximum probability of the observation. The designed system was developed to be able to classify four types of pre-established faults in ball bearings and the normal condition. The system was trained and tested by experimental data collected from drive end ball bearing of an induction motor, operating under several shaft speeds and load conditions. The method was applied successfully. It permits the separation of different faults with high recognition rate, almost all fault samples of the database were assigned to the appropriate classes.

Key words: Bearing faults diagnosis, fourier transform, wavlet transform, HMM

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 the 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. Artificial neural networks models have been applied to the domain of fault diagnosis. Thy have the advantage of learning any type of data, capability of noise filtering and parallel computing (Duda *et al*., 2001; Hay kin, 1998; Li, *et al*., 1996). Howevertheir performances depend on the convenient selection of the type of the structure and the quantities of the training data, which are not always available in sufficient quantities. Among various stochastic approaches, the HMM have proven very effective in modelling both dynamic and static signals. The success of the HMM in the speech recognition domain (Rabiner, 1989) motivate its extension to others domains like fault bearing diagnosis (Miao and Makis, 2007; Purushotham *et al*., 2005).

In this study a hidden Markov model based faults diagnosis of bearing is developed. The features extraction is based on wavelet transform. Various defects, which have mechanical origin, are detected by the analysis of the vibration signals recorded for these bearings under different operating conditions. The performances of the proposed method based on wavelet features, are compared to those based on temporal and frequency features.

VIBRATION SIGNALS AND FAULTS CHARACTERISTICS

The vibration signals analysis is applied in order to examine the health of the rotating machines. The monitoring of these systems aims at:

- Reducing the number of system stops.
- Improving reliability of the production equipments.

Corresponding Author: Salah Chenikher, LASA Laboratory, Badji Moukhtar University, BP. 12, Annaba 23000, Algeria

- Increasing the availability of the system.
- Helping to better manage the stock of the spare parts.

The mechanical vibrations are oscillating movements around an average position; these movements can be periodic or not periodic (transients or random). The periodic vibrations correspond to the rotation movement of the machine. Transient vibrations are generated by discontinuous forces (shocks); correspond to the impact of balls and the race defect. The chipping of a track of bearing causes shocks and a resonance of the stage. The shock frequency is characteristic of the type of the ball bearing damage and can be calculated if the bearing geometry and the rotating speed are known.

Depending on the location of the damage, (outer ring, inner ring and rolling elements) by spectral analysis this phenomenon appears in high frequencies.

A defect of ball bearing appears by the continual repetition of the defective contact with the outer race of the bearing like the inner one. For that, the characteristic frequency of the bearing is the double of the rotational frequency and we note it, f_b (Algunindigue *et al.*, 1993; Harris, 1991).

$$f_b = \frac{D_c}{D_b} f_r . \left[1 - \left(\frac{D_b}{D_c} \cos\theta \right)^2 \right] \qquad (1)$$

Where D_b is the balls diameter, D_c is the bearing pitch diameter Fig. 1, f_r mechanical rotor speed in hertz.

Vibrations are generated by stator currents at frequencies given by:

$f_{bng} = |fs \pm m\ f_{int>ext}|$

Where, fs is electrical supply frequency.

Repetition frequencies for the inner f_{int} and outer race f_{ext} are described by:

$$f_{int} = \frac{n_b}{2} f_r . \left[1 - \frac{D_b}{D_c} \cos\ \theta \right]$$
$$f_{ext} = \frac{n_b}{2} f_r . \left[1 + \frac{D_b}{D_c} \cos\ \theta \right] \qquad (2)$$

Where, n_b is the number of bearing balls and è is the contact angle between race and balls.

It should be noted that specific information concerning the bearing construction are required for calculating the exact characteristic frequencies. However, these frequencies can be approximated for most bearings containing between six and twelve balls Fig. 1.

$$f_{int} = 0.4.n_b.f_r,\ f_{ext} = 0.6.n_b.f_r$$

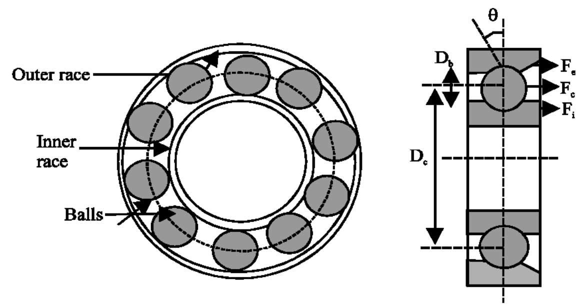

Fig. 1: Schematic of a ball bearing

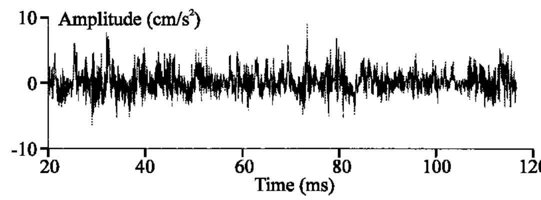

Fig. 2: Example of a faulty vibration signal

This generalization allows for the definition of frequency bounds where the bearing race frequencies are likely to show up without requiring explicit knowledge of the bearing construction.

The measuring of the vibration signals consists on the transformation of the mechanical vibration (acceleration) into electric signal Fig. 2.

FEATURE EXTRACTION

Before the application of the feature extraction procedure from vibration signals, they are normalised by the following expression.

$$x_i = \frac{s_i - \bar{s}}{\sigma} \qquad (3)$$

Where, s_i is the sample i of the signal and σ are respectively the average and the standard deviation of the signal, x_i signal after normalisation.

Frequency analysis: The use of Fourier transform is to search for the periodically repeated peaks in the power spectrum. Frequency features are very informative for rotating components like ball bearings, since well defined frequency components are associated with them. Every defect in bearings is expressed by high frequency components.

The resonance frequency oscillation of the shocks and the possibly suppressed shock periodicity are two modulations of the vibrations that both reduce the amplitude of the power spectrum peaks, which therefore are more likely to be suppressed below the overall noise level. The applied method is to remove the resonance frequency modulation with the envelope method, which consists of a band-pass filter including the resonance frequency followed by a demodulation and a fast Fourier transformation. Both these methods use a band-pass filter to focus on a range of frequencies which must be wide enough to include the resonance frequency. Thus, it is likely that also oscillation frequencies where bearing shock oscillations not are dominating are included in the analysis, with consequences such as lower signal to noise ratio and more sensitivity to possible suppressions of the impact periodicity (Ericsson *et al*., 2005).

Wavelet analysis: The theory of wavelet transform is a coherent mathematical framework for analyzing signals. It was developed as an alternative to the Short Time Fourier Transform (STFT) to overcome problems related to its frequency and time resolution properties. It allows separating signals of high frequency transitions, from low frequency vibrations. The wavelet transform is defined as the integral of the signal x(t) multiplied by scaled, shifted versions of a basic wavelet transform function ø(t), a real valued function whose Fourier transform satisfies the admissibility criteria (Daubechies, 1992; Lou and Loparo, 2004; Mallat, 1998).

The wavelet transform of continuous time signal, x(t), is defined as:

$$W(a,b) = \int_R x(t)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)dt, \quad a \in R^+ - \{0\}, b \in R \qquad (4)$$

Where, a is called the scaling parameter, b is the time localisation parameter. Both, a and b can be continuous or discrete variables.

The Continuous Wavelet Transform (CWT) is a time frequency analysis method which differs from the more traditional Short Time Fourier Transform (STFT) by allowing arbitrarily high localization in time of high frequency signal features by a variable window width, which is related to the scale of observation. This flexibility allows the isolation of the high frequency features. Another important distinction from the STFT is that the CWT is not limited to using sinusoidal analyzing functions. A large selection of localized waveforms called wavelet mother can be utilized.

Multiplying each coefficient by an appropriately scaled and shifted wavelet yields the constituent wavelets of the original signal. For signals of finite energy, continuous wavelet synthesis provides the reconstruction formula:

$$x(t) = \frac{1}{C_\psi}\int_R\int_{R^+} W(a,b)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)\frac{da}{a^2}db. \qquad (5)$$

Associated with the wavelet **Ψ**, which is used to define the details (high scale/low frequency content) in the decomposition, a scaling function to avoid intractable computations when operating at every scale of the CWT. Scales and positions can be chosen based on a power of two. The discrete wavelet transform DWT analysis is more efficient and just as accurate. In this scheme, a and b are given by: $a=2^j, b = k2^j$, m, n∈Z.
Let us define:

$$\Psi_l\ (t) = 2^{-j/2}ø(2^{-j}t\text{-}k), \qquad (6)$$

$$\phi_{j,k}(t) = 2^{-j/2}\phi\ (2^{-j}t\text{-}k), \qquad (7)$$

A wavelet filter with impulse g (high frequency), plays the role of wavelet ø and scaling filter with impulse response h (low frequency), plays the role of scaling function ϕ. Then the discrete wavelet analysis can be described mathematically as:

$$w(j,k) = \sum_{i\in Z} x(i)g_{j,k}(i) \qquad (8)$$

And discrete synthesis:

$$x(t) = \sum_{j\in Z}\sum_{k\in Z} w(j,k)\psi_{j,k}(t) \qquad (9)$$

The detail at level j is defined as:

$$D_j(t) = \sum_{n\in Z} w(j,k)\psi_{j,k}(t) \qquad (10)$$

Features extraction is carried out by the wavelet decomposition at level 5. After reconstitution of decompositions, we choose the feature vectors as the observations in different decomposition levels.

HIDDEN MORKOV MODELS (HMM)

HMM models estimation: The HMM is a Markovian-based model, extended from the concept of Markov chain, whose states cannot be observed directly. Usually, it contains finite number of states, where each state generates an observation at certain time point. The hidden state is characterised by two sets of probabilities: a transition probability and an observation probability distribution. In addition, the third probability distribution has to be computed for an HMM is the distribution of the initial hidden state. In summary, the complete specification of an HMM includes the following elements (Robiner, 1989).

Set of hidden states: $S = \{S_1, S_2, \ldots, S_N\}$,

Where N is the number of states in the model;

State transition probability distribution: $A=\{a_{ij}\}$,

Where $a_{ij} = P[q_{t+1}= S_j|q_t=S_i]$, for $1\leq i,j \leq N$, q_t represents the hidden state at time t;

Set of observation symbols: $V = \{v_1, v_2, \ldots, v_M\}$,

Where M is the number of observation symbols per state.

Observation symbol probability distribution is given by: $B =\{b_j(k)\}$,

Where $b_j(k) = P[v_k \text{ at } t|q_t=S_j]$, for $1\leq, j\leq N$, $1\leq k \leq M$;

Initial state probability distribution: $B = \{\pi_i\}$,

Where $\pi_i = P[q_1=S_i]$, for $1\leq i,j\leq N$.

For convenience, an HMM can be represented by the compact notation: $\lambda = (A, B,\pi)$ to indicate the complete parameter set of the model.

The HMMs training is based on the data from different fault conditions. In this phase input features are modelled by a set of parameters $\lambda = (A, B,\pi)$. The initial parameters of the HMMs are chosen randomly. Different initial parameters generally produce different HMM models.

HMM based diagnosis: The HMM based diagnosis consists on finding the best path or state sequence in each trained model and selecting the one that maximises the path probability for a given input observation. However, in real application, like bearing condition monitoring, it is more efficient to establish several HMM models corresponding to different conditions in consideration Fig. 3.

In that case, the hidden states of the model do not have physical meaning and decision regarding the current machine state is made by choosing the model that gives maximum probability of the observation.

That is, bearing condition C can be selected by

$$C = \text{argmax}\{P(O/\lambda_j)\},\ 1\leq j\leq Nc \qquad (11)$$

Where, Nc is the number of bearing conditions considered in a classification system, which is equal to the number of HMM models in the system. The Hmm model consists of a model of five states.

RESULTS AND DISCUSSION

The vibration data used for validation is composed from four different faults states of ball bearings and one class representing the normal state.

The class c1 represents a new ball bearing, c2 an outer race completely broken, c3 a broken cage with one

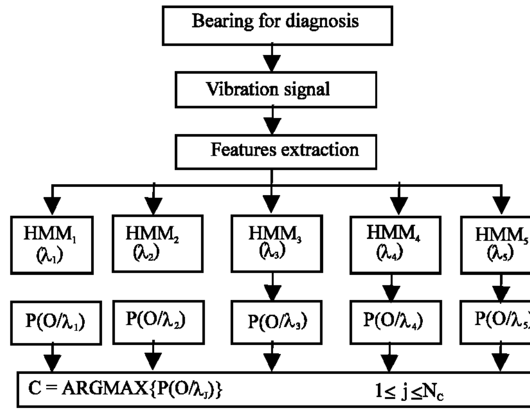

Fig. 3: HMM diagnosis scheme

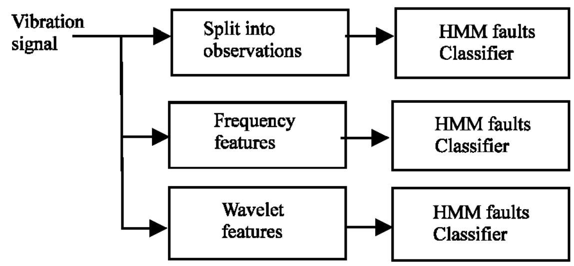

Fig. 4: Features extration and faults diagnosis based on HMM

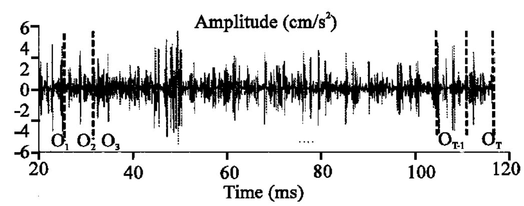

Fig. 5: Feature extraction (split into a sequence of non-overlapping observation1

loose element, c4 a damaged cage with four loose elements and c5 a no evident damage (badly warned ball bearings).

The rotational frequency of the bearing is equal to 24.5625 Hz. The Measured signal is acceleration (cm/s^2). Signals are sampled at frequency of 16384 Hz. The minimum frequency considered is 0.7 Hz. The recorded signals include 2048 samples.

Three kinds of features are extracted to characterise each vibration signal.

The extracted temporal feature vectors are composed from temporal observations Fig. 5.

The Fourier transform is best suited for the analysis of stationary signals. In such case, Frequency features are very informative. Every defect associated with the ball bearings is expressed by high frequency components.

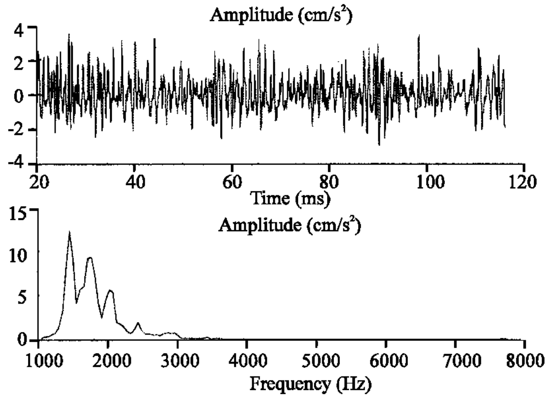

Fig. 6: Vibration signal and specturm of a normal ball bearing

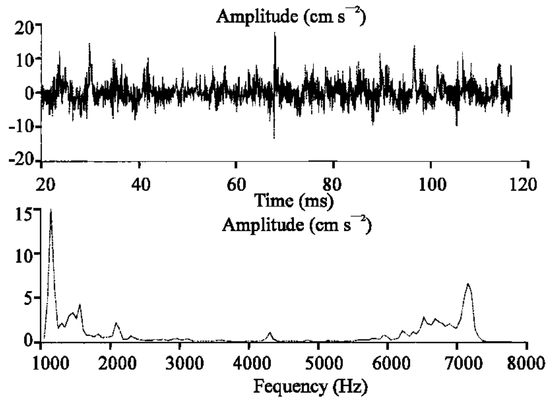

Fig. 7: Vibration signal and specturm of defective ball bearing

For each of the four default classes Fourier transform was carried out. Figure 6 and 7, show the spectrum of the signals it exhibits the normal frequency characteristics and those related to one kind of defect. Based on visual inspection, or simple threshold test, it was observed that for the normal condition, the frequency components were concentred in low frequency and defect condition frequency components were observed in high frequencies some spectral frequency features were computed in high frequency bound (maximum frequency, median frequency and mean amplitude).

In many cases it is difficult to distinguish the bearing conditions by Fourier transform method. However, wavelets transform permits achieving best results for this specific application. After data normalization, the Daubechies-1 wavelet was used to achieve the wavelet transform.

The vibration signals and the wavelet decompositions of the signal are split into a sequence of non-overlapping observations. The length of the observation was chosen to reduce the computation time and at the same time containing enough information to

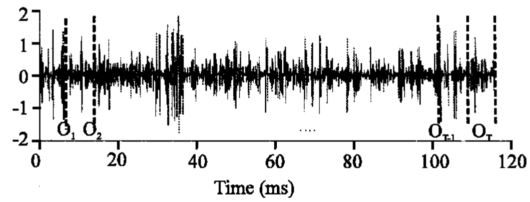

Fig. 8: Example of feature extraction from second level wavelet decompostion (detail D2)

Table 1: Ratio of good classification for temporal and wavelet features, given for training and testing data.

Method	Training (%)	Testing (%)
Temporal features	96,21	94,47
Frequency features	97.14	94.89
Wavelet features	100	99, 59

capture the local features of the signal. The extracted wavelet feature vectors are composed from wavelet observations Fig. 8.

In order to validate the proposed method, the HMM was trained based on the data from different fault states. The vibration data was divided into two separate groups, one for training and the other for testing. Thus, data used for testing was not used in training. After training procedure each class of faults is represented by an HMM model λ = (A, B,π). Consequently fives model are estimated Fig. 3.

In decision making the problem to be solved is reduced to the estimation of the conditional probability between the observations and each HMM model. Therefore, each observation is assigned to the class corresponding to maximum conditional probability.

Results show that, using observations from temporal vibration signal as features for the HMM classifier gives a recognition rate of 96,21% for the training set and 94,47% for the test set Table 1. The introduction of the frequency features, improve slightly the reconnaissance rate. However, the use of wavelet features permits to improve considerably the recognition ratio and gives best results. Consequently, all the elements of the training set are correctly classified while a rate of 99,59% of good classification in the testing set is obtained.

The error of 0,41% is due to the confusion between two elements from classes 3 and 4, which represent respectively broken cage with one loose element and damaged cage with four loose elements. These two classes are similar in the damage of the bearing cage and differ only in the number of loosing elements. However, the remaining classes 1, 2 and 5 are completely separated.

CONCLUSION

In this study, we develop a method for vibration signal analysis and fault diagnosis. It permits the recognition of four faults affecting the ball bearings and the normal state. These faults are detected by analysing the vibration signals recorded under different operating conditions. Wavelet transform allows separating signals high frequency components corresponding to bearing defects from low frequency. Faults diagnosis based on Hidden Markov models is realised in two phases. In the first phase the feature vectors are extracted to train the HMM. In the second phase feature vectors are assigned to the class that gives the maximum conditional probability of observation to the corresponding model. The obtained results show that the proposed method based on wavelet features gives a satisfactory recognition rate, compared to the methods based on the temporal and frequency features.

REFERENCES

Alguindigue I., A. Loskiewics-Buczak and R. Uhrig, 1993. Monitoring and diagnosis of rolling element bearings using artificial neural networks, IEEE. Trans. Indus. Elec., 40: 209-216.

Daubechies, I., 1992. Ten Lectures on Wavelets. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Duda R., Hart P. and Stork D., 2001. Pattern Classification, (2nd Edn.), Wiley-Interscience.

Ericsson, S., N. Grip, E. Johansson, L. Persson, R. Sjo¨berg and J. Stro¨mberg, 2005. Towards automatic detection of local bearing defects in rotating machines, Mech. Sys. Signal Processing, 19: 509-535.

Harris T., 1991. Rolling bearing analysis, (3rd Ed.), New York: Wiley.

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

Li, Y., S. Billington, C. Zhang, T. Kurfess, S. Danyluk and S. Liang 1996, Adaptive prognostics for rolling element bearing condition, Mech. Sys. Signal Process, 10: 1-17.

Lou X. and K.A. Loparo 2004. Bearing fault diagnosis based on wavelet transform and fuzzy inference, Mech. Sys. Signal Process., 18: 1077-1095.

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

Miao, Q. and V. Makis, 2007. Condition monitoring and classification of rotating machinery using wavelets and hidden Markov models, Mech. Sys. Signal Processing, 21: 840-855.

Purushotham, V., S., Narayanan and S.A.N. Prasad, 2005. Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition, NDT and E International, 38: 654-664.

Rabiner LR., 1989. A tutorial on hidden Markov models and selected applications in speech recognition, Proc. IEEE. Ultrasonic Symp, 77: 257-86.