2006 IEEE International Conference on
Systems, Man, and Cybernetics
October 8-11, 2006, Taipei, Taiwan

Faults Detection Using Gaussian Mixture Models, Mel-Frequency Cepstral Coefficients and Kurtosis

Fulufhelo V. Nelwamondo and Tshilidzi Marwala

***Abstract*—Most machines failures can be associated with mechanical failures on bearing failures. This paper proposes a novel approach to detect and classify three types of common faults in rolling element bearings. The approach proposed here makes use Gaussian Mixture model to classify. Mel-frequency Cepstral Coefficients (MFCC) and Kurtosis are extracted from the bearing vibration signal and are used as features. A classification rate of 95% is obtained when using the MFCC features only while a classification rate improves to 99% when Kurtosis features are added to the MFCC..**

I. Introduction

ROLLING bearings are the most widely used components in industrial rotating machinery and are one of the major causes of catastrophic breakdowns in rotating machinery. Failures of these bearings often result in lengthy industrial downtimes that have economic consequences. Proper condition monitoring is therefore required to prevent unexpected bearing failures. An automated condition monitoring of bearings is necessary as manual checks may take an unacceptably long duration resulting in money losses. Vibration-based monitoring is the most popular approach and has been used extensively in various bearing condition-monitoring techniques [1][2]. Vibrations can also be used to detect existence of faults such as mass imbalance, shaft misalignment and gear failures by simply comparing the vibration signals of a machine operating with and without faults. There are several causes of bearing failure such as incorrect design, acid corrosion, poor lubrication and plastic deformation [3]. Damage in bearings is typically on the rolling element, inner race or outer race [1]. The difficulty of fault detection in bearings lies in the fact that the signature of a defective bearing is spread across a wide frequency band and can be masked by noise.

There is a great deal of information on fault detection and diagnosis on rolling bearings, and these include Lou and Loparo [1][4] who have used wavelet transform and neuro-fuzzy interference to detect and classify different faults in bearings. Nikolaou and Antoniadis [5] also used the wavelet for fault diagnosis while Rojas and Nandi [6] have successfully used support vector machines with both spectral and statistical features in classification of bearing faults. Peng et al. [7] as well as Wang and Kootsookos [8] compared Hilbert-Huang transform with wavelet transform in bearing fault diagnosis in favor of the former. Junsheng et al. [9] proposed a feature extraction method based on empirical mode decomposition method and autoregressive model for the roller bearings. In their autoregressive modeling, Baillie and Mathew [10] have classified bearing faults with a high accuracy. Their autoregressive model does not only classify, but also provides a one-step-ahead prediction of the vibration signal using the previous outputs.

Yang et al [11] applied the basis pursuit and obtained better results than wavelet transforms. The inspiration from the success of hidden Markov models in speech led to [3] successfully applying hidden Markov models for the diagnosis of multi-faults in bearings. Like speech signals, vibration signals have a large variability that makes direct comparisons almost impossible. Due to the success of Gaussian Mixture Model (GMM) over Neural networks in speech recognition, this paper proposes for the first time the use of (GMM) classifier in conjunction with the Mel-frequency Cepstral Coefficients (MFCC) and Kurtosis features for bearing fault diagnosis.

II. Theoretical Background

A. Vibration signals from bearing faults

Changes in vibrations signals of mechanical systems may be used to detect faults in the system [8]. A bearing working under normal and perfect conditions is expected to produce less vibration than when it has defects. Some of the causes of these changes in vibration of rotating systems include mass unbalance around the centre of rotation and this mass unbalance produces sinusoidal force given by [8]

$$F = A_1 \cos(\ 2\pi f_s + \theta_1) \tag{1}$$

where A_1 is the amplitude determined by the size of the unbalance, f_s is the running speed and θ_1 is the appropriate phase angle. A bearing defect generates an impulsive force given by [8][12]

$$F = \sum_{k=1}^{\infty} A_3^k \delta\left(t - \frac{k}{f_c}\right) \approx f_s \sum_{m=1}^{N_2} A_3^m \cos\left(2\pi m f_c + \theta_2^m\right) \tag{2}$$

where A_3^k is the appropriate amplitude of the m^{th} harmonic and is determined by the defect size, f_c is the characteristic frequency, θ_2^m is the initial phase of the m^{th} harmonic and N_2 is the total number of harmonics in the impulse train. Vibrations from the inner raceway, outer raceway and

Manuscript received November 29, 2005.

T. Marwala is a professor of Electrical Engineering at the school of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa. (phone: +27 11 717 7217; fax: +27 11 403 1929; e-mail: t.marwala@ee.wits.ac.za).

F. V. Nelwamondo is with the school of Electrical and Information Engineering at the University of the Witwatersrand, Johannesburg, South Africa. (e-mail: f.nelwamondo@ee.wits.ac.za).

rolling element faults generate spectra with different frequency components whose magnitudes can be used to determine the condition of the bearing.

B. Gaussian Mixture Models

Gaussian mixture models have been a powerful classification tools in many applications of pattern recognition, particularly in speech and face recognition. GMM have proven to perform better than Hidden Markov Models in text independent speaker recognition [13]. The success of GMM in classification of dynamic signals has also been demonstrated by many researchers such as Cardinaux and Marcel [14] who compared GMM and MLP in face recognition and found that the GMM approach easily outperforms the MLP approach for high resolution faces and is significantly more robust in imperfectly located faces. Their experiments also showed that the computational requirements of the GMM approach could be significantly smaller than the MLP approach. Gaussian mixture models are therefore very suitable for classification of bearing faults since vibration data are dynamic due to various factors such as the size of the load, severity of the fault and others. Moreover, there is no prior knowledge on how the vibration signal will look like when a fault exists. Other advantages of using a GMM are that it is computationally inexpensive and is based on well-understood statistical models [13].

A Gaussian Mixture Model (GMM) is a statistical model used to model features from a signal. The GMM works by creating a model of each fault, which is written as [13]:

$$\lambda = (w, \mu, \Sigma) \tag{3}$$

where λ is the model, w represent the weights assigned to the Gaussian means, μ is the diagonal variance of the features used to model the fault and Σ is the covariance matrix. GMM contains a probability density function (pdf) of the observation consisting of a sum of normal observations [15]. A weighted sum of Gaussians normally provides an accurate model of the data. Each Gaussian comprises of a mean and a covariance; hence, it is referred to as a mixture of components. The Gaussian probability density function is given by [15]:

$$p(x) = \frac{1}{\sigma\sqrt{2x}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{4}$$

where μ is the mean and σ is the standard deviation of the distribution of a variable x. For a case where x is a vector of features, (4) becomes [13][15]

$$p(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}\left[(x-\mu)'\Sigma^{-1}(x-\mu)\right]} \tag{5}$$

where n is the size of x. The log-likelihood is then computed as [13][16][17]

$$\hat{s} = \arg \max_{1 \le f \le F} \sum_{k=1}^{K} \log \ p(x_k \mid \lambda_f) \tag{6}$$

where f represents the index of the type of fault, while **F** is the total number of known fault condition and $x = \{x_1, x_2, \ldots, x_K\}$ is the unknown fault vibration segment. $p(x_k|\lambda_f)$ represents the mixture density function. An arbitrary probability density of a sample vector x can be approximated by a mixture of Gaussian densities [16][17] as

$$p(\mathrm{x} \mid \lambda) = \sum_{i=1}^{M} w_i \, p(\mathrm{x}) \tag{7}$$

where all mixtures weights, w_i, are adjusted to satisfy the constrains [18], $\sum_{i=1}^{M} w_i = 1$ and $0 \le w_i \le 1$

Training of the GMM is a fast and straightforward process, which estimates the mean and covariance parameters from the training data [15][18]. The training procedure estimates the model parameters from a set of observations. Normally the Expectation Maximization (EM) algorithm is preferred [13]. The EM algorithm tries to increase the expected log-likelihood of the complete data x given the partially observed data [18] and finds the optimum model parameters by iteratively refining GMM parameters for a given bearing fault feature vector.

In the investigation in this paper, only the diagonal covariance matrices are used. The reason for this is that the diagonal covariance matrix was found to be more computationally efficient and outperforms the full covariance matrix.

C. Mel-frequency cepstral coefficients

In this paper a feature extraction technique that extracts both linear and non-linear features is required and here we implement the Mel-frequency Cepstral Coefficients (MFCC). The MFCC is a type of wavelet in which frequency scales are placed on a linear scale for frequencies less than 1 kHz and on a log scale for frequencies above 1 kHz [19]. The complex cepstral coefficients are called the MFCC [19]. The MFCC contain both time and frequency information of the signal and this makes them more useful for feature extraction. MFCC have widely been used in the field of speech recognition and have managed to handle the dynamic features as they extract both linear and non-linear properties of the signal. MFCC can be a useful tool of feature extraction in vibration signals as vibrations contain both linear and non-linear features. The following steps are involved in MFCC computations.

a) Take the Fast Fourier Transform (FFT) of the signal, *x(n)*, using [20]:

$$Y(m) = \frac{1}{F}\sum_{n=0}^{F-1} x(n)w(n)e^{-j\frac{2\pi}{F}nm} \tag{8}$$

where F is the number of frames, $0 \le m \le F-1$ and *w(n)* is the hamming window function given by:

$$w(n) = \beta\left(0.5 - 0.5\cos\frac{2\pi n}{F-1}\right) \tag{9}$$

where $0 \le n \le F-1$ and β is the normalization factor defined such that the root mean square of the window is unity [20].

b) Mel-frequency wrapping is performed by changing the frequency to the Mel using the following equation [3][20].

$$mel = 2595 \times \log_{10}(1 + \frac{f(Hz)}{700}) \qquad (10)$$

Mel-frequency warping uses a filter bank, spaced uniformly on the Mel scale. The filter bank has a triangular band pass frequency response, whose spacing and magnitude are determined by a constant Mel-frequency interval.

c) The final step converts the logarithmic Mel spectrum back to the time domain. The result of this step is what is called the Mel-frequency Cepstral Coefficients. This conversion is achieved by taking the Discrete Cosine Transform of the spectrum as:

$$C_m^i = \sum_{n=0}^{F-1} \cos\left(m \frac{\pi}{F}(n + 0.5) \right) \log_{10}(H_n) \qquad (11)$$

where $0 \leq m \leq L-1$ and L is the number of MFCC extracted form the i^{th} frame of the signal. H_n is the transfer function of the n^{th} filter on the filter bank. These MFCC are then used as a representation of the signal.

D. *Kurtosis*

In this paper there is a need to deal with the occasional spiking of vibration data, which is caused by some types of faults and to achieve this task Kurtosis is used. Wang *et al* [20] have used the Kurtosis features together with neural networks to detect grinding burn from acoustic emission. Kurtosis features of vibration data have also been successfully used in tool condition monitoring by El-Wardany [21]. The success of Kurtosis in vibration signals is based on the fact that vibration signals of a system under stress or having defects differ from those of a normal system. The sharpness or spiking of the vibration signal changes when there are defects in the system.

Kurtosis is a measure of the sharpness of the peak and is defined as the normalized fourth-order central moment of the signal [3][20]. The Kurtosis value is useful in identifying transients and spontaneous events within vibration signals [22][23] and is one of the accepted criteria in fault detection. The calculated Kurtosis value is typically normalized by the square of the second moment, as shown in (12) [3]. A high value of Kurtosis implies a sharp distribution peak.

$$K = \frac{1}{N} \sum_{i=1}^{N} \frac{(x_i - \overline{x})^4}{\sigma^4} \qquad (12)$$

E. *Principal Component Analysis*

In the work of this paper, there is also a need to reduce the dimension of the input space and to achieve this task the Principal Component Analysis (PCA), which is a non-parametric method that extracts relevant information from any complex data, is used. The PCA reduces the dimensionality of the data while retaining important information [24] by eliminating variables with high correlation. Consequently, PCA uses the least number of uncorrelated variables called principal components. The necessity of the PCA in this application is that it simplifies the data and makes the training of GMM computationally efficient and increases the identification accuracy. Another advantage is that the loss of information is minimized when PCA is used [16]. The dimensionality of a vector of data can be reduced to principal components using (13)

$$Z = \gamma^{-1}\overline{x} \qquad (13)$$

where Z represents the principal components and γ is an orthogonal matrix of eigenvectors of the correlation matrix of the variables in data $\overline{x}$ [24]. The process of reducing dimensionality is conducted in this paper using these steps:

a) Calculate the covariance matrix of the input data.
b) Calculate the eigenvalues and eigenvectors of the covariance matrix.
c) Retain the largest eigenvalues and corresponding eigenvectors which contain at least 80% of the variance of the data
d) Project the original data into the retained eigenvectors to form new input data with reduced space.

F. *Feature extraction and model generation*

The techniques explained above are feature extraction methods. These features are the MFCC and Kurtosis and are reduced to lower dimension using the PCA. Classical fault detection techniques extract features only in one domain, either being time or frequency. Because of the non-linear factors such as friction and stiffness, that can affect the vibration signals, it is not easy to accurately evaluate the condition of roller bearing only in time domain or in frequency domain [22].

The vibration signal for a given fault type was first broken into segments, each being five revolutions long. On extracting the MFCC features, each segment of the signal was further broken into 14 frames of equal duration The MFCC were then extracted from each frame of the signal. In this paper, extracted MFCC formed a matrix of size 14x13 where 13 is the number of MFCC features extracted and 14 is the number of frames within a particular segment. In addition to the extracted MFCC, the Kurtosis features were used.

Model generation on the other hand involves building a reference model based on the extracted features. Detection of a defect in bearings is a process of classifying any of the obtained vibration signals into classes that have been defined during the model generation stage. Fig. 1 shows the model generation and testing stages in the classification of bearing faults. The features extracted were used to train a GMM. In order to ease the computational process, all rows of the feature matrix were then concatenated to form a long single rowed vector.

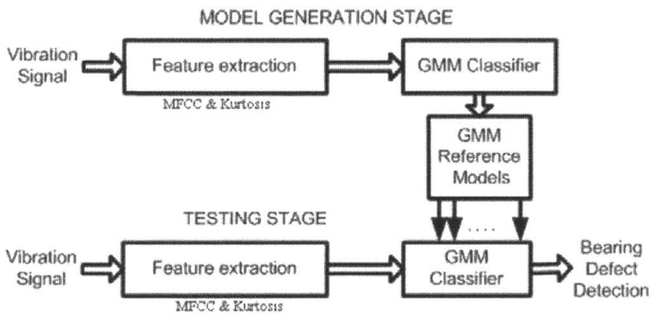

Fig.1. A schematic presentation of the fault diagnosis and classification

The entire vector is an input to the GMM. However, the dimension of the vector had to be reduced using PCA while retaining intrinsic information. In this paper the dimension of the input is reduced from 182 to 11 and the dimension of the features that serve as inputs to the GMM is 12.

III. EXPERIMENTATION

The investigation in this paper is entirely based on the data obtained in the experimentation done by [4] at the Rockwell Science Centre. Their experimental setup comprised a Reliance Electric 2HP IQPreAlert connected to a dynamometer. Faults of size 0.007, 0.014, 0.021 and 0.028 inches were introduced into the drive-end bearing of a motor using the EDM method. An impulsive force was applied to the motor shaft and the resulting vibration was measured using two accelerometers, one mounted on the motor housing and the other on the outer race of the drive-end bearing. All signals were recorded at a sampling frequency of 12 kHz.

Practically, most faults start small and build up to more severe faults. It is therefore very important to detect faults at an early stage. However, it is more difficult to detect a small damage than a big one. Detecting a 0.007-inch scratch on a bearing should be more difficult than detecting a 0.028 inch one. The most challenging aspect of this problem is detecting faults when the severity is still small and for this reason, only the 0.007 inch data will be used to detect faults.

IV. RESULTS AND DISCUSSION

We first demonstrate how the vibration signals differ under different fault conditions. Fig. 2 shows signals from the bearings during the normal condition and during three different fault conditions under the same loading. We present in Fig. 3, the MFCC features extracted in an arbitrary frame of the vibration signal. The horizontal axis shows all the 13 MFCC extracted in one frame for all conditions under investigation. Here it should be noted that the input to the GMM is not per frame but for the whole segment, which has 14 frames.

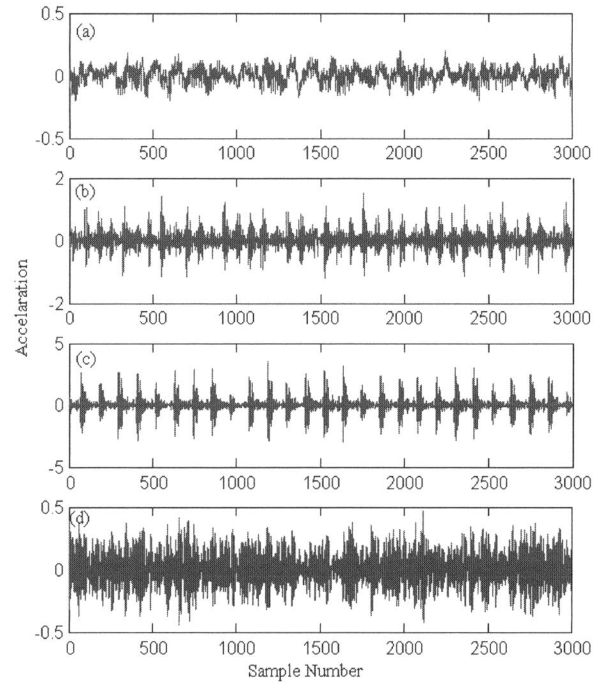

Fig. 2. Vibration signals on the bearing under (a) normal condition, (b) Inner race fault, (c) outer race fault and (d) ball fault conditions

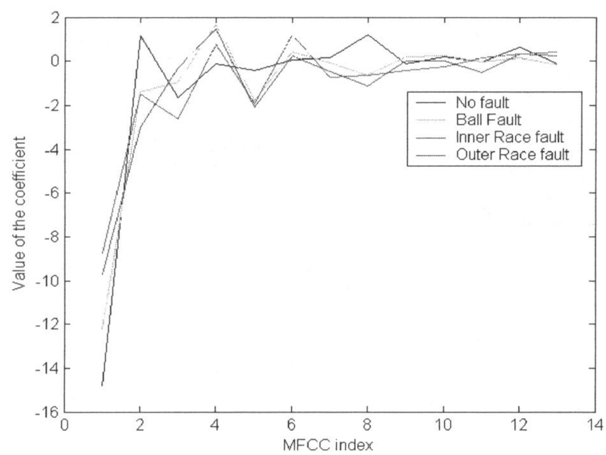

Fig. 3. MFCC values corresponding to different fault conditions

The Kurtosis was computed in time for each segment of the signal. Fig. 4 shows how Kurtosis of the vibration signals varies with different fault conditions for 60 segments of the vibration data. A high value of Kurtosis demonstrates that the distribution has a sharp peak and in this investigation, also tells us that there has been a change in the conditions of the bearing.

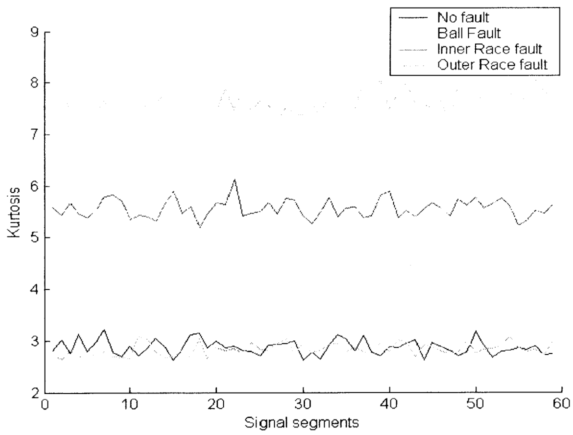

Fig.4. Kurtosis values corresponding to different fault conditions

Table 1 summarizes the result of the classification of the bearing faults using the GMM classifier. Results obtained with MFCC features are compared to those obtained when MFCC are used together with Kurtosis features. The computational times are also compared and it can be seen that there is no major difference in computational time even when Kurtosis features are added. It can also be observed that the Kurtosis features lead to an improvement of about 5% to the results obtained with MFCC only. Table 2 shows the normalized confusion matrix for fault classification using the MFCC features with Kurtosis.

Table 1: Comparison of classification rate using MFCC only and MFCC with Kurtosis

	GMM Classification rate	Computation Time (s)
MFCC	94%	0.898
MFCC + Kurtosis	99%	0.903

Table 2: Confusion matrix for MFCC features in percentages

		Obtained			
		Normal	Inner	Outer	Ball
True fault	Normal	100	0	0	0
	Inner	0	100	0	0
	Outer	0	0	100	0
	Ball	1.8	0	0	98.2

A good classification rate obtained in this investigation can be attributed to feature extraction methodologies. Features seem to be more unique for different fault conditions. However, there seems to be a very small misclassification between ball faults and the normal condition. The reason for this is that the signals seem to have a very similar spiking as shown by the Kurtosis values presented in Fig. 4.

Trying to improve the classification performance, we further investigate the effect of changing the number of MFCC extracted from each frame and compare the classification results. The results of this are shown in Fig. 5.

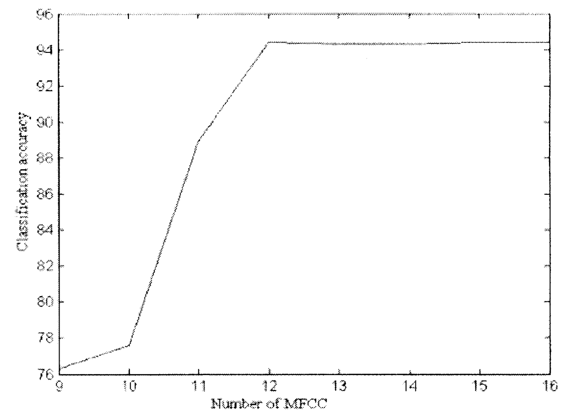

Fig. 5. Demonstration of how the classification accuracy varies with the number of MFCC

It is observed that increasing the number of MFCC above 13 does not improve the classification results. The possible explanation for this behavior is that the values of MFCC look very similar from the 9th MFCC onwards as can be noted from Fig. 3. The duration of the training process of the GMM with 13 MFCC is approximately 0.903 seconds with a Pentium IV having a processor speed of 2.4GHz. Using the same machine, the feature extraction process takes 3.069 seconds on average which was computed after 12 runs.

V. Conclusions

A procedure for diagnosis and classification of bearing condition using a Gaussian Mixture model has been presented. The feature extraction is the most important step in roller bearing fault diagnosis. Both MFCC and Kurtosis features are extracted and used. It is found that combining the MFCC with Kurtosis improves the classification process by approximately 5% to 99% when compared to using the MFCC in isolation.

References

[1] X. Lou, K. A. Loparo, F. M. Discenzo, J. Yoo, A. Twarowski, A model-based technique for rolling element bearing fault detection, International Conference on Acoustics, Noise and Vibration , Canada, 8-12 August 2000.

[2] R. B. Randall, Detection and diagnosis of incipient bearing failure in helicopter gearboxes, Engineering Failure Analysis, vol. 11, 2004 pp. 177–190.

[3] V. Purushothama, S. Narayanana, A. N. S. Prasad, Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition, NDT&E International, vol. 38, 2005, pp. 654–664.

[4] X. Lou, K. A. Loparo, Bearing fault diagnosis based on wavelet transform and fuzzy inference, Mechanical Systems and Signal Processing, vol. 18, 2004, pp. 1077–1095
[5] N. G. Nikolaou, L. A. Antoniadis, Rolling element bearing fault diagnosis using wavelet packets, NDT&E International, vol. 35, 2002, pp. 197-205
[6] A. Rojas, A. K. Nandi, Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines, Mechanical Systems and Signal Processing, 2005, pp. 1-13
[7] Z. K. Peng, P. W. Tse, F.L. Chu, A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing, Mechanical Systems and Signal Processing, vol.19, 2005, pp. 974–988
[8] Y. F. Wang, P. J. Kootsookos, Modelling of low shaft speed bearing faults for condition monitoring, Mechanical Systems and Signal Processing, vol.12, no. 3, 1998, pp. 415-426.
[9] C. Junsheng, Y. Dejie, Y. Yu, A fault diagnosis approach for roller bearings based on EMD method and AR model, Mechanical Systems and Signal Processing. vol. 20, 2006, pp.350–362.
[10] D. C. Baillie, J. Mathew, A comparison of autoregressive modelling techniques for fault diagnosis of rolling element bearings, Mechanical Systems and Signal Processing, 2006, (to appear)
[11] H. Yang, J. Mathew, L. Ma, Fault diagnosis of rolling element bearings using basis pursuit, Mechanical Systems and Signal Processing, vol. 19, 2005, pp. 341–356
[12] C. Hsieh, E. Lai and Y. Wang, Robust Speaker Identification System Based on Wavelet Transform and Gaussian Mixture Model, Journal of Information Science and Engineering, 19 (2003) 267-282
[13] D. A. Reynolds, T. F. Quatieri, R. B. Dunn, Speaker verification using adapted Gaussian Mixture Models, Digital Signal Processing, vol. 10, 2000, pp 19–41.
[14] F. Cardinaux, C. Cardinaux' S. Marcel, Comparison of MLP and GMM Classifiers for Face Verification on XM2VTS, Springer-Verlag GmbH. 2688, 2003, pp. 911-920.
[15] J. Wang, J. Wang, Y. Weng, Chip design of MFCC extraction for speech recognition, Integration, The VLSI Journal, vol. 32, 2002, pp. 111–131.
[16] S. W. Choi, J. H. Park, I. Lee, Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis, Computers and Chemical Engineering, vol. 28, 2004, pp. 1377–1387.
[17] C. M. Bishop, Neural networks for pattern recognition, New York: *Oxford University press*, 1995.
[18] L. Meng, M. G. Frei, I. Osorio, G. Strang, T. Q. Nguyen, Gaussian mixture models of ECoG signal features for improved detection of epileptic seizures, Medical Engineering & Physics, vol. 26, 2004, pp. 379–393.
[19] H. Ocak, K. A. Loparo, Estimation of the running speed and bearing defect frequencies of an induction motor from vibration data, Mechanical Systems and Signal Processing, 18 (2004) 515–533.
[20] Z. Wang, P Willett, P. R. DeAguiar, J Webster, Neural network detection of grinding burn from acoustic emission, International Journal of Machine Tools & Manufacture, vol. 41, 2001, pp. 283-309.
[21] T. I. El-Ward any, D. GAO, M. A. Elbestawi, Tool condition monitoring in drilling using vibration signature analysis, International Journal of Machine Tools & Manufacture, vol. 36, 996, pp. 687-711.
[22] C. J. Li, S. M. Wu, On-line detection of localized defects in bearings by pattern recognition analysis, ASME Journal of Engineering and Industry, vol. 111, 1989, pp. 331-336.
[23] J. Antoni, The spectral Kurtosis: a useful tool for characterising non-stationary signals, Mechanical Systems and Signal Processing, vol. 20, 2006, pp. 282–307
[24] I. T. Jolliffe, Principal component analysis, Springer-Verlag, New York, 1986