Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Journal of Sound and Vibration 291 (2006) 60–71
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Fault recognition method for speed-up and speed-down process of rotating machinery based on independent component analysis and Factorial Hidden Markov Model

Zhinong Li[a,b,*], Yongyong He[b], Fulei Chu[b], Jie Han[a], Wei Hao[a]

[a] *Vibration Engineering Research Institute, Zhengzhou University, Zhengzhou 450002, PR China*
[b] *Department of Precision Instruments and Mechanology, Tsinghua University, Beijing 100084, PR China*

Received 14 July 2004; received in revised form 29 April 2005; accepted 23 May 2005
Available online 19 August 2005

Abstract

The behavior characteristics of the speed-up and speed-down process of rotating machinery possess the distinct diagnostic value. The abundant information, non-stationarity, poor repeatability and reproducibility in this process lead to the necessity to find the corresponding method of feature extraction and fault recognition. In this paper, combining independent component analysis (ICA) and Factorial Hidden Markov Model (FHMM), a new method of the fault recognition named ICA–FHMM is proposed. In the proposed method, ICA is used for the redundancy reduction and feature extraction of the multi-channel detection, and FHMM as a classifier to recognize the faults of the speed-up and speed-down process in rotating machinery. This method is compared with another recognition method named ICA–HMM, in which ICA is similarly used for the feature extraction, however Hidden Markov Model (HMM) as a

Abbreviations: FHMM, Factorial Hidden Markov Model; HMM, Hidden Markov Model; ICA, Independent Component Analysis; ICA–FHMM, a fault recognition method, in which ICA is used for a feature extraction, and FHMM, as a classifier; ICA–HMM, a fault recognition method in which ICA is similarly used for a feature extraction however HMM as a classifier; PCA, principal component analysis; MI, mutual information; JADE, joint approximate diagonalization of eigen-matrices; SVA, structured variational approximation; EM, expectation maximization; p.d.f., probability density function; rpm, revolutions per minute

*Corresponding author. Department of Precision Instruments and Mechanology, Tsinghua University, Beijing 100084, PR China.

E-mail addresses: lizhinong@tsinghua.org.cn (Zhinong Li), heyy@pim.tsinghua.edu.cn (Yongyong He), chufl@pim.tsinghua.edu.cn (Fulei Chu).

0022-460X/$ - see front matter
doi:10.1016/j.jsv.2005.05.020

classifier. Experimental results show that the proposed method is very effective, and the ICA–FHMM recognition method is superior to the ICA–HMM recognition method.

1. Introduction

For rotating machinery, the speed-up and speed-down process contains abundant running information of the set and some useful fault symptoms, which cannot be found in the stationary process and may be revealed in this process. Therefore, the behavior characteristics of this process have the distinct diagnostic value, and the fault diagnosis of this process has owned its definite place in the fault diagnosis of rotating machinery.

However, the vibration signal of this process is a non-stationary signal, whose frequencies and amplitudes both vary with time. When the set speeds up or speeds down through the first or second resonance speed, the variation of the vibration signals is very drastic. Because this process is affected by various random factors from the set, the features of its vibration signals have poor repeatability and reproducibility. Considering the vibration features in this process, it is very necessary to find a corresponding method of feature extraction and fault recognition.

Recently, the diagnosis in the speed-up and speed-down process has been developed [1–5]. However, the distinct diagnostic value of the behavior characteristics of the set in this process need to be further explored thoroughly.

Factorial Hidden Markov Model (FHMM), which is a recognition tool of the dynamic model, is suitable for modeling the dynamic time series, and has a strong capability of pattern classification, especially for the signal with the characteristics of abundant information quantity, non-stationarity, poor repeatability and reproducibility. Theoretically, FHMM can process the random long sequences. Contrasting the features of vibration signals in the speed-up and speed-down process, we can find that FHMM is a very suitable modeling and pattern recognition tool for the speed-up and speed-down process of rotating machinery. Li [6–9] is probably the first to introduce FHMM to the machine fault diagnosis successfully.

However, Pattern recognition is based on the feature extraction. The feature extraction is the key of machine fault diagnosis. In the condition monitoring and fault diagnosis of rotating machinery, fault-related machine vibration is usually corrupted by the vibration from the structured machine itself and noise from interfering machinery. Moreover, the vibration signal from the sensors mounted upon the machine is only a mixture of the underlying vibration source signals because of spatial redundancy and diversity, many fault-related peaks in the spectrum of the signals from several sensors will be all visible at the same time, this redundancy makes it more difficult to extract the fault features effectively [10,11]. So it is necessary to find an effective method of redundancy reduction.

Recently, independent component analysis (ICA) has been widely applied to the feature extraction of speech, image and biomedicine, etc. [10–23]. In fact, ICA is a powerful tool for analyzing non-Gaussian data. As the extension of standard principal component analysis (PCA) to higher-order statistics, ICA imposes statistical independence on the extracted components and has no orthogonality constraint [11,13,21]. Similarly with the feature extraction of speech, image and biomedicine based on ICA, ICA can be also used for the feature extraction of mechanical fault. ICA is potentially a useful tool in the feature extraction of mechanical faults.

In this paper, based on ICA and FHMM, a new method of fault recognition for the speed-up and speed-down, i.e. the ICA–FHMM recognition method, is proposed. The proposed method is compared with another recognition method named ICA–HMM. The experiment is carried out to verify the proposed method.

The organization of the paper is as follows: A method of the feature extraction based on ICA is described in Section 2. The basic concept and algorithm of FHMM is introduced in Section 3. A new fault recognition method named ICA–FHMM is proposed in Section 4. Experimental tests in the speed-up and speed-down process for rotating machinery are stated in Section 5. Finally, a conclusion is summarized in Section 6.

2. ICA-based feature extraction

The use of ICA for feature extraction is motivated by the theory of redundancy reduction [24]. One aspect of redundancy reduction is that the input data are represented using components (features) that are independent from each other as much as possible. Such a representation seems to be very useful for later pattern recognition.

For instantaneous mixtures and if a noise-free case is assumed, the general model is as follows:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{1}$$

where $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_m(t)]^{\mathrm{T}}$ is an m-dimensional observed vector, $\mathbf{s}(t)$ is an unknown n-dimensional source vector containing the source signals $s_1(t), s_2(t), \ldots, s_n(t)$, which are assumed to be statistically independent. Usually, the number of the sources is assumed to be smaller than or equal to the number of the observations, i.e. $n \leqslant m$. $\mathbf{A}$ is an unknown $m \times n$ mixing matrix, which is written as

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = [\, a_1| \quad a_2| \quad \cdots \quad a_n|\,], \tag{2}$$

where $a_i|$, $i = 1, 2, \ldots, n$ denotes the ith column of $\mathbf{A}$. So Eq. (1) can be rewritten as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \to \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_m(t) \end{bmatrix} = [\, a_1| \quad a_2| \quad \cdots \quad a_n|\,] \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_n(t) \end{bmatrix}. \tag{3}$$

Obviously from Eq. (3), the columns of $\mathbf{A}$ represent features, and $s_i(t)$ is the coefficient of the ith feature in an observed data vector $\mathbf{x}(t)$. The aim of ICA is to find a linear transform denoted by

W, which can be treated as an approximate pseudo-inverse of **A**, such that

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t), \tag{4}$$

where $\mathbf{y}(t) = [y_1(t), y_2(t), \ldots, y_n(t)]^{\mathrm{T}}$ is the approximate estimate of the source signals, **W** is also called the ICA estimated basis, i.e. separating matrix. So an m-dimensional observed vector is transferred to an n-dimensional vector by the ICA transform. Owing to $m \geqslant n$, this implicates that the ICA transform primarily compresses the dimensions of the observation to some extent. On the basis of primary compression, information redundancy can be further compressed deeply by the mutual information.

The mutual information is an important factor that evaluates the independence between the components. It is always non-negative, and zero if and only if the components are statistically independent. The less the mutual information, the better the independence between components. However, in practice, the statistical independence of the underlying vibration sources is never satisfied strictly. So the components estimated by the ICA method have different non-zero residual mutual information, which can be used as the feature vector of the fault recognition.

The symbol **MI** is assumed to be the mutual information between the components $y_1(t), y_2(t), \ldots, y_n(t)$. **MI** is written as follows:

$$\mathbf{MI} = \begin{bmatrix} \mathrm{MI}_{11} & \mathrm{MI}_{12} & \cdots & \mathrm{MI}_{1n} \\ \mathrm{MI}_{21} & \mathrm{MI}_{22} & \cdots & \mathrm{MI}_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{MI}_{n1} & \mathrm{MI}_{n2} & \cdots & \mathrm{MI}_{nn} \end{bmatrix}. \tag{5}$$

The mutual information in non-diagonal of the matrix **MI** is used as a measurement of the independence between various components, and the diagonal as a measurement of the independence between the same components. In this paper, the average of all the columns in the **MI** matrix is selected as the subsequent feature extraction.

The mutual information is estimated by a simple and reliable histogram-based method, which is presented by Moddemeijer [25].

At present, there exit many ICA algorithms. Here, a joint approximate diagonalization of eigen-matrices (JADE) method, which has been presented by Cardoso [26], is used to extract the independent components from the multi-channel observations. The steps of JADE are described in Ref. [26].

3. Description of factorial Markov models [27–30]

FHMM is first described by Ghahramani and Jordan [27]. In his work, Ghahramani presented FHMM and introduced several methods to efficiently learn their parameters. Logan and Moreno [28] studied the applicability of FHMMs to speech modeling. Their goal is to study FHMMs as a viable replacement for HMMs. Our focus, however, is on exploring the application of FHMM to the machine fault diagnosis.

HMM is a probabilistic model describing a sequence of the observation vector $\mathbf{Y} = \{Y_t : t = 1, 2, \ldots, T\}$. It is characterized by a hidden state sequence and an output probability which depends on the current state in all the layers. Given a sequence of the observations $\mathbf{Y} = \{Y_t : t = 1, 2, \ldots, T\}$ and the hidden states $\mathbf{S} = \{S_t : t = 1, 2, \ldots, T\}$, then the joint probability for the sequence of the states and the observations can be factored as [27]

$$P(\{S_t, Y_t\}) = \pi(S_1)P(Y_1/S_1)\prod_{t=2}^{T} P(S_t/S_{t-1})P(Y_t/S_t), \tag{6}$$

where $P(S_t/S_{t-1})$ is the state transition probability from the state S_{t-1} at time $t-1$ to the state S_t at time t, $\pi(S_1)$ is the prior probability of being in states S_1 at time $t = 1$, and $P(Y_t/S_t)$ is the probability density function (p.d.f.) of the observation vector Y_t given the state S_t. For a continuous observation vector, its p.d.f. $P(Y_t/S_t)$ can be modeled in many different forms, such as a Gaussian, a mixture of Gaussians, or even a neural network. We assume that the model has K states.

A HMM can be represented by a dynamic belief network, as shown in Fig. 1. This representation shows the evolution of the state sequence in time. Each node represents the state at each time slices. This context switch to dynamic belief networks allows many new modeling possibilities such as FHMM.

FHMM arises by forming a dynamic belief network composed of several layers. This is shown in Fig. 2. Each layer has the independent dynamics, and the observation vector depends upon the current state in each of the layers. This is achieved by letting the state variable in Eq. (6) to be represented by a collection of the state variables. That is to say, we now have a state variable S_t

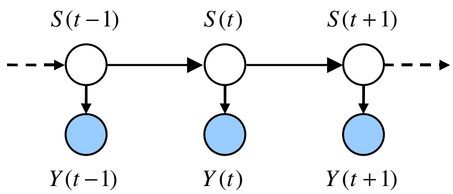

Fig. 1. Dynamic belief network representation of an HMM.

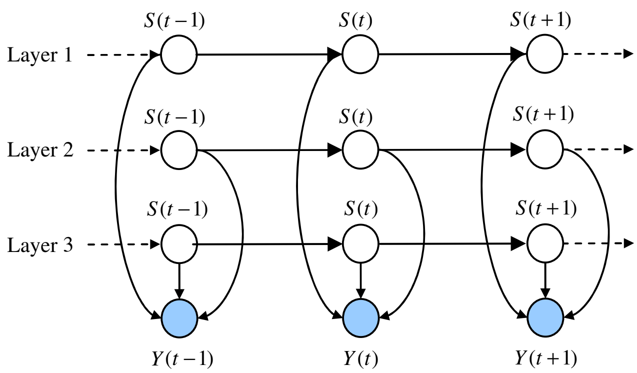

Fig. 2. Dynamic belief network representation of an FHMM.

which is represented by M states as follows

$$S_t = S_t^{(1)}, S_t^{(2)}, \ldots, S_t^{(M)}. \tag{7}$$

Here the superscript is the layer index with M being the number of layers. The layer nature of the model arises by only allowing transitions between the states in the same layer. If we were to allow unrestricted transition between states, we would have a regular HMM with a $K^M \times K^M$ transition matrix. Intermediate architectures in which some limited transition between states in different layers is allowed have also been presented in [29].

By dividing the states into layers, we form a system that models several processes with loosely couple dynamics. Each layer has similar dynamics to a basic hidden Markov model. But the probability of an observation at each time depends upon the current state in all of the layers. For simplicity, the number of possible states in each layer is K. Thus we have a system that requires $MK \times K$ transition matrices. Such a FHMM system can be represented as a traditional HMM with a $K^M \times K^M$ transition matrix. For example, considering a two-layer FHMM system with four states per layer, i.e. $M = 2$, $K = 4$, the transition matrix for the equivalent basic HMM system with $K^M = 16$ transition matrix. We can see that an explosion of the computation occurs with the increment of the number of the states. For this reason, as noted in Refs. [27,28], it is preferable to use the $MK \times K$ transition matrices over the equivalent $K^M \times K^M$ representation simply on computational grounds.

We now consider the probability of the observation given the state S_t. As mentioned above, the probability density of the observation vector Y_t depends on the current states in all layers. In Refs. [27,28], this probability was modeled by a Gaussian p.d.f., which is given as follows

$$P(Y_t/S_t) \propto \exp\left\{-\frac{1}{2}\left(Y_t - \sum_{m=1}^{M} \mu^{(m/S_t)}\right)^{\mathrm{T}} C^{-1}\left(Y_t - \sum_{m=1}^{M} \mu^{(m/S_t)}\right)\right\}. \tag{8}$$

Here $\mu^{(m/S_t)}$ is the mean of the layer m given the state S_t, C is the covariance. Other symbols are the same as those previously defined. This probability is the basis of pattern classification according to the models.

The parameters in the FHMM are estimated by the structured variational approximation (SVA) algorithm, which is the approximate expectation maximization (EM) algorithm, the detail of this algorithm can be found in Refs. [27,28,30].

4. ICA–FHMM recognition method

The basic idea of the ICA–FHMM recognition method is described as follows: The feature vectors are extracted from the multi-channel observations by the ICA transform, then this set of the feature vectors is input into FHMM of each fault mode for training, and the output probability of each FHMM is obtained, where the FHMM with the maximum probability determines the condition of the set. The relation between ICA and FHMM is that between the feature extraction and the pattern classification. The scheme of the whole system is shown in Fig. 3.

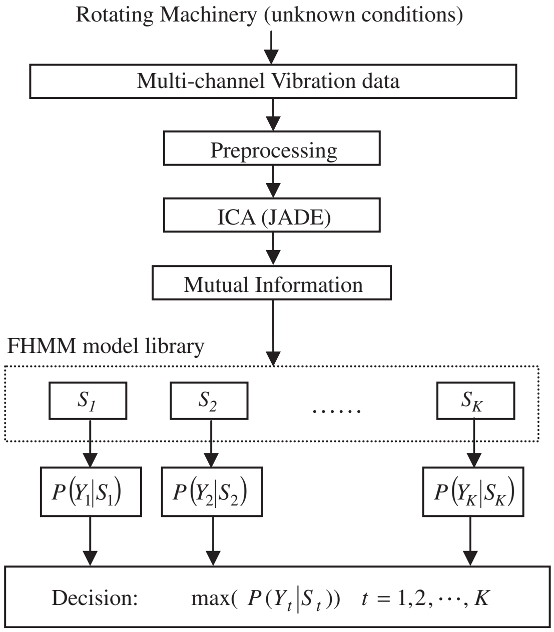

Fig. 3. FHMM recognition method based on ICA feature extraction.

5. Experimental research

Here we verify the proposed method by the experiment. Four running conditions, which are unbalance, rubbing, oil whirl and pedestal looseness, are performed on Bently rotor kit. In the speed-up and speed-down process, two sensors mounted on the bearing block horizontally and vertically respectively are used to acquire the vibration signals. The number of the sampling points is 64 in per rotating period. The system samples 8 rotating periods continuously with the same sampling rate. So the length of one group of the sampling data is 512 points.

This fault detection and diagnosis system tracks the rotating speed of the rotor from 500 revolutions per minute (rpm) to 8000 rpm. Each 50 rpm is used as a rotating speed segment. So, there are all together $(8000 - 500)/50 = 150$ segments, i.e. 500–550 rpm, 550–600 rpm, $600-650$ rpm, ..., $7850-7900$ rpm, 7900–7950 rpm, 7950–8000 rpm. For the fault recognition, one group of the data, the length of which is 512, is randomly extracted from each rotating speed segment, respectively. Thus, 150 groups of the data are used to train FHMM. The feature vectors are extracted by the above-described method and input into each FHMM for the fault classification. In each FHMM, the number of the layers is $M = 3$, the number of the states in each layer is $K = 2$, the maximum iterative step is set to 100, and the convergence error of the algorithm is set to 0.0001. The iterative step of each FHMM and its corresponding log likelihood probability in each step is shown in Fig. 4. The testing results are shown in Table 1, where the maximum log likelihood probability, which is obtained from the above four states, is marked in bold.

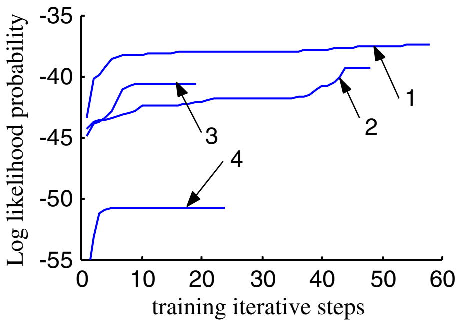

Fig. 4. ICA–FHMM training curves: 1. oil whirl, 2. pedestal looseness, 3. unbalance and 4. rubbing.

Table 1
Test results of ICA–FHMM

	Models			
States	Unbalance	Rubbing	Oil whirl	Pedestal looseness
Unbalance	**−40.57**	−47.69	−47.23	−42.81
Rubbing	−113.80	**−50.70**	−174.06	−82.84
Oil whirl	−40.07	−46.00	**−37.42**	−40.17
Pedestal looseness	−44.56	−46.02	−50.38	**−39.27**

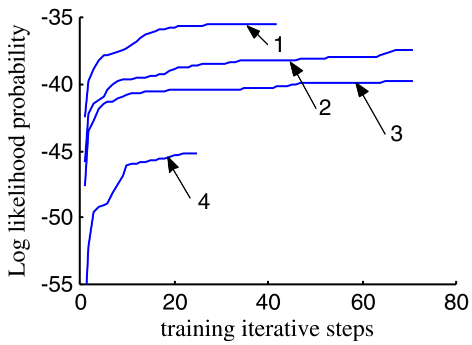

Fig. 5. ICA–HMM training curves: : 1. oil whirl, 2. unbalance, 3. pedestal looseness and 4. rubbing.

For comparison, we study another fault recognition method named ICA–HMM, in which ICA is similarly used for a feature extraction, however HMM as a classifier. In order to compare ICA–FHMM with ICA–HMM in the same state–spaces and the same convergence error, here in each HMM, the number of the states is 8, the maximum iterative step is set to 100, and the convergence error of the algorithm is also set to 0.0001. The iterative step of each HMM and its corresponding log likelihood probability in each step is also shown in Fig. 5. The testing results are shown in Table 2, where the maximum log likelihood probability, which is obtained from the above four states, is also marked in bold.

Table 2
Test results of ICA–HMM

States	Models			
	Unbalance	Rubbing	Oil whirl	Pedestal looseness
Unbalance	**−37.43**	−44.02	−56.35	−43.46
Rubbing	−138.14	**−45.24**	−261.97	−80.32
Oil whirl	−42.49	−42.28	**−35.55**	−41.75
Pedestal looseness	−45.50	−43.26	−60.41	**−39.81**

Now we compare the two recognition methods in two aspects: recognition effect and train speed.

5.1. Comparison of recognition effect

From Fig. 4 and Table 1, Fig. 5 and Table 2, whether the ICA–FHMM recognition method or the ICA–HMM recognition method, the classification is satisfactory. In the two recognition methods, the feature vector estimated by the ICA transform is input into each state's FHMM or HMM for training, the obtained log likelihood probability greatly approximates the maximum output probability of the model which belongs to this state, and greatly deviates from the maximum output probability of the model which does not belong to this state. Here we illustrate the fault recognition effect using an example of rubbing fault recognition.

In the ICA–FHMM recognition method, the rubbing data are input into this state's FHMM, its output probability is −50.70, which greatly approximates the maximum output probability of this state. However, the rubbing data are input into the other FHMMs, i.e. unbalance, oil whirl and pedestal looseness, their output probabilities are −113.80, −174.06, −82.84, respectively, which are all smaller than the maximum output probability of the rubbing FHMM. Their difference is −63.10, −123.24, −32.14, respectively. The classification is obvious.

Similarly, In the ICA–HMM recognition method, the rubbing data are input into this state's HMM, and its output probability is −45.24, which greatly approximates the maximum output probability of this state. However, the rubbing data are input into the other HMMs, i.e. unbalance, oil whirl and pedestal looseness, their output probabilities are −138.14, −261.97, −80.32, respectively, which are all smaller than the maximum output probability of the rubbing HMM. Their difference is −92.90, −216.73, −35.07, respectively. The classification is also obvious.

The other faults, i.e. unbalance, oil whirl and pedestal looseness, can be analyzed similarly. The experiment shows that the two recognition methods are very effective.

5.2. Comparison of training speed

In the two recognition methods, i.e. ICA–FHMM, and ICA–HMM, the feature vector is the same; however, the classifier is different, one is an FHMM, another is an HMM. Although the test results show that the two recognition methods are very effective, however their training speed has

Table 3
Comparison of training time (unit: s)

State-space size (K^M)	Maximum iterative steps	Convergence error	Recognition method	
			ICA–FHMM	ICA–HMM
4(2^2)			4.26	4.56
8(2^3)			8.75	9.25
16(2^4)	100	0.0001	15.56	16.15
32(2^5)			20.32	26.25
64(2^6)			72.71	88.10
128(2^7)			161.43	218.70

obvious difference in the two recognition methods. In order to compare their convergence speed, here we test it using the same feature vectors. The maximum iterative step is set to 100. The convergence error of algorithms is set to 0.0001. The training results are given in Table 3 when the state space size is 4, 8, 16, 32, 64 and 128.

From Table 3, when state–space size is small, such as 4, 8 and 16, the training time consumption of the two recognition methods has a little difference. However, the training speed has obvious difference with the increment of state–space size. For example, state–space size is 32, time consumption of the ICA–HMM recognition method is 26.25 s, and that of the ICA–FHMM recognition method 20.32 s, their difference is 5.93 s. When state–space size is 64, time consumption of the ICA–HMM and ICA–FHMM recognition method is 88.10 and 72.71 s, respectively, their difference is 15.39 s. However, when state–space size is 128, the difference in training time is very obvious, the training speed of the ICA–FHMM recognition method is obviously faster than that of the ICA–HMM recognition method. This comparison analysis confirms the standard HMM is extremely slow for the model with large state–space, which is also proved in Ref. [27]. A traditional HMM with a $K^M \times K^M$ transition matrix can be represented as an FHMM with M $K \times K$ transition matrices, so the calculation is greatly speeded up.

6. Conclusions

Considering the features in the speed-up and speed-down process of rotating machinery, we propose a new recognition method named ICA–FHMM, and verify the proposed method by the experiment. The experiment result shows that the proposed method is very effective, i.e. that fault data (such as rubbing) are input into this state's FHMM, and its output probability greatly approximates the maximum output probability of this state. However, when these fault data are input into other FHMMs, their output probabilities are all smaller than the maximum output probability of this state's FHMM, the classification is obvious. The proposed method is compared with another recognition method named ICA–HMM. The simulation shows that the training speed of the former is faster than that of the latter for the same state–space size. Especially the larger the state–space size, the more obvious the superiority of the former. The research also indicates that the proposed method has the very potentiality to solve a wide range of the fault recognition of other equipment or machine with similar non-stationary process.

Acknowledgements

The research described in this paper was supported by Natural Science Foundation of China (50105007), the Postdoctoral Foundation (20040350061) and the Trans-Century Training Programme Foundation for the Talents by the Ministry of Education, China.

References

[1] M. Ingleby, G.P.L. Ronval, Classification of vibration signatures of turbogenerators, in: *The Fourth International Conference on Electrical Machines and Drives*, London, UK, 13–15 September 1989, pp. 172–176.

[2] M. Ingleby, G.P.L. Ronval, Automatic taxonomy and modal analysis for vibration diagnosis, in: *Fifth International Conference on Electrical Machines and Drives*, (Conference Publication No. 341), London, UK, 11–13 September 1991, pp. 237–240.

[3] T.S. Davies, C.M. Jefferson, R.M. Mayer, Automated synchronization, in: *IEE Colloquium on Microcomputer Instrumentation and Control Systems in Power Electronics*, London, UK, 25 April 1988, pp. 3/1–3/3.

[4] M.Q. Xu, W.F. Huang, J.Z. Zhang, Application of Haar wavelet on analysis of vibration signal of rotating machinery in fast run-up state, *Journal of Vibration Engineering* 13 (2) (2000) 216–221.

[5] X. Liu, Z.Y. Zhao, L.L. Qu, *Rotor Monitoring and Diagnosis System*, Xi'an Jiatong University Presss, Xi'an, 1991.

[6] Z.N. Li, Q.Q. Ding, Z.T. Wu, C.J. Feng, Study on bispectrum-FHMM recognition method in speed-up and speed-down process of rotating machinery, *Journal of Vibration Engineering* 16 (2) (2003) 171–174.

[7] Z.N. Li, Z.T. Wu, Q.Q. Ding, Y.Y. He, F.L. Chu, Application of wavelet transform and FHMM in fault diagnosis of speed-up and speed-down process of rotating machinery, *Control Engineering of China* 10 (2) (2003) 299–301.

[8] Z.N. Li, Q.Q. Ding, C.J. Feng, Z.T. Wu, Study on BSI–FHMM recognition method in speed-up & speed-down process of a rotating machinery, *China Mechanical Engineering* 13 (20) (2002) 1730–1733.

[9] Q.Q. Ding, Z.N. Li, Z.T. Wu, S.X. Zhen, Research on fault diagnosis methods for rotating machines based on factorial hidden Markov model, *Power Engineering* 23 (4) (2003) 2560–2563.

[10] A. Ypma, A. Leshem, R.P.W. Duin, Blind separation of rotating machine sources: bilinear forms and convolutive mixtures, *Neurocomputing—Special Issue on ICA/BSS* 49 (1–4) (2003) 349–368.

[11] W.D. Jiao, S.X. Yang, Z.T. Wu, Extracting invariable fault features of rotating machines with multi-ICA networks, *Journal of Zhejiang University Science* 4 (5) (2003) 595–601.

[12] A. Ypma, Learning Methods for Machine Vibration Analysis and Health Monitoring, Ph.D. Dissertation, Pattern Recognition Group, Department of Applied Physics, Delft University of Technology, November 2001, 224pp. (ISBN 90-9015310-1).

[13] W.D. Jiao, S.X. Yang, ICA based neural networks for pattern recognition of mechanical faults, *Computing Technology and Automation* 22 (2) (2003) 63–67.

[14] A.J. Bell, T.J. Sejnowski, Learning higher-order structure of a natural sound, *Network* 7 (1996) 261–266.

[15] S. Makeig, A.J. Bell, T.-P. Jung, T.J. Sejnowski, Independent component analysis of electroencephalographic data, in: M. Mozer, et al. (Eds.), *Advances in Neural Information Processing Systems*, Vol. 8, MIT press, Cambridge, MA, 1996, pp. 145–151.

[16] A. Hyvärinen, P.O. Hoyer, Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces, *Neural Computation* 12 (7) (2000) 1705–1720.

[17] J. Karhunen, A. Hyvärinen, R. Vigario, J. Hurri, E. Oja, Applications of neural blind separation to signal and image processing, in: *Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing*, Munich, Germany, 1997, pp. 131–134.

[18] P. Sanna, J. Pedro, H. Heikki, Independent component analysis of vibrations for fault diagnosis of an induction motor, http://www.control.hut.fi/Research/vaivi/ICA_final2.pdf.

[19] W.D. Jiao, Research on Methods for Fault Diagnosis of Rotating Machinery Based on Independent Component Analysis, Ph.D. Thesis, Zhejiang University, 2003.

[20] A. Ypma, D.M.J. Tax, R.P.W. Duin, Robust machine fault detection with Independent Component Analysis and Support Vector Data Description, *Proceedings of 1999 IEEE International Workshop on Neural Networks for Signal Processing*, Madison, Wisconsin (USA), 23–25 August 1999, pp. 67–76.

[21] S. Chitroub, A. Houacine, B. Sansal, Compound PCA–ICA neural network model for enhancement and feature extraction of multi-frequency polarimetric SAR imagery, in: *Proceedings of the 2000 International Conference on Image Processing*, (ICIP 2000), Vancouver, BC, Canada, 10–13 September 2000.

[22] M. Kotani, Y. Shirata, S. Maekawa, S. Ozawa, K. Akazawa, Application of independent component analysis to feature extraction of speech, *International Joint Conference on Neural Networks*, Vol. 5, 10–16 July 1999, Washington, DC, USA, pp. 2981–2984.

[23] J.-H. Lee, H.-Y. Jung, T.-W. Lee, S.-Y. Lee, Speech feature extraction using independent component analysis, *Proceedings*, *2000 IEEE International Conference on Acoustics*, *Speech*, *and Signal Processing* (*ICASSP'00*), Vol. 3, 5–9 June 2000, Istanbul, Turkey, pp. 1631–1634.

[24] A. Hyvärinen, E. Oja, Survey on independent component analysis, *Neural Computing Surveys* 2 (1999) 94–128.

[25] R. Moddemeijer, On estimation of entropy and mutual information of continuous distributions, *Signal Processing* 16 (3) (1989) 233–246.

[26] J.F. Cardoso, A. Souloumiac, Blind beamforming for non-Gaussian signals, *IEE Proceedings-F* 140 (6) (1993) 362–370.

[27] Z. Ghahramani, M.I. Jordan, Factorial hidden Markov models, *Machine Learning* 29 (1997) 245–273.

[28] B. Logan, P. Moreno, Factorial HMMS for acoustic modeling, *Proceedings of the 1998 IEEE International Conference on Acoustics*, *Speech*, *and Signal Processing* (*ICASSP '98*), Vol. 2, WA, USA, 12–15 May 1998, pp. 813–816.

[29] M. Brand, Coupled hidden Markov models for modeling interacting processes, Technical Report 405, MIT Media Lab Perceptual Computing/Learning and Common Sense, June 1997.

[30] B. Logan, P.J. Moreno, Factorial hidden Markov models for speech recognition: preliminary experiments, Cambridge Research Laboratories Technical Report Series, CRL 97/7, September, 1997.