A new method of bearing fault diagnostics in complex rotating machines using multi-sensor mixtured hidden Markov models

Z. S. Chen[1], Y. M. Yang[1], Z. Hu[1], Z. X. Ge[1]

[1] *Institute of Mechatronic Engineering, College of Mechatronic Engineering and Automation, National University of Defense Technology, Changsha, Hunan, P. R. China, 410073*
czs_study@sina.com
yangyongmin@yahoo.com

ABSTRACT

Vibration signals from complex rotating machines are often non-Gaussian and non-stationary, so it is difficult to accurately detect faults of a bearing inside using a single sensor. This paper introduces a new bearing fault diagnostics scheme in complex rotating machines using multi-sensor mixtured hidden Markov model (MSMHMM) of vibration signals. Vibration signals of each sensor will be considered as the mixture of non-Gaussian sources, which can depict non-Gaussian observation sequences well. Then its parameter learning procedure is given in detail based on EM algorithm. In the end the new method was tested with experimental data collected from a helicopter gearbox and the results are very exciting.

1. INTRODUCTION

Today's industry uses increasingly complex rotating machines, some with extremely demanding performance criteria. Machine failures are significantly contributed to both safety incidents and maintenance costs. The root cause of faults in complex rotating machines is often faulty bearings. A bearing condition monitoring system is therefore necessary to prevent major breakdowns due to progression of undetected faults. Over the past tens years, much research has been focused on vibration-based fault diagnostics techniques (Paul and Darryll, 2005). For complex rotating machines, however, it is still difficult to achieve a high degree of accuracy in classifying faults of a bearing inside due to the complexity of vibration signals.

Hidden Markov Model (HMM) has been a dominant method in speech recognition since 1960s and becomes very popular in the late 1980s and 1990s (Rabiner, 1989). The structure of HMM is useful for modeling a sequence that has a hidden stochastic process. It has become popular in various areas like signal analysis and pattern recognition, such as speech processing and medical diagnostics. Recently, HMMs have been introduced into mechanical diagnostic areas and many HMMs were proposed and extended successfully for mechanical systems monitoring and diagnostics (Baruah and Chinnam, 2005; Leea, et al., 2004; Bunks, et al., 2000). In practice, it is an important issue how to select an appropriate HMM model. Most existing HMM-based fault diagnostic methods mainly assume that each state generates observations according to a Gaussian or Gaussian mixture model (Baruah and Chinnam, 2005; Leea, et al., 2004; Bunks, et al., 2000; Wang, et al., 2009). Also these methods often use a single sensor system to perform condition monitoring and diagnostics. Whereas vibration signals of complex rotating machines are often known to be highly non-Gaussian and non-stationary (Bouillaut and Sidahmed, 2001), such as a helicopter gearbox. Thus classical HMMs with Gaussian or Gaussian-mixtured observations have serious limitations for bearing fault diagnostics in complex rotating machines.

Obviously, a multi-sensor fault diagnostic system can overcome the limitations of a single sensor system and has improved performance. So our motivation is to build a novel HMM with non-Gaussian observations based on multi-sensor signals and then use it for bearing fault diagnostics in complex rotating machines. Vibration signals from a sensor on complex rotating machines can be looked as emanating from a number of sources caused by these components within it. This naturally fits an independent component analysis (ICA) process (Lee, et al., 2000). By this way, this paper will present a multi-sensor mixtured hidden Markov model (MSMHMM) for bearing fault diagnostics, which is improved on classical HMMs with mixtured non-Gaussian observation models.

2. DEFINITION OF MSMHMM

For a Gaussian observation model, the observation O_t at time t is assumed to be generated from a Gaussian process, which is a scalar value corresponding to a single sensor. While for a multi-sensor system with N sensors, the observation $\mathbf{O}_t$ at time t will be a vector, i.e. $\mathbf{O}_t = [x_{t1}, x_{t2}, \ldots, x_{tN}]^T$. As mentioned before, signals from each sensor on a helicopter gearbox can be considered to be mixed by M sources caused by its inner components. In this paper a linear mixing process is considered. Denoting $\mathbf{W}_k$ as the mixing matrix at state k and $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M]^T$ as M sources, the observation vector at time t for state k can be calculated according to an independent component analysis process as follows,

$$\mathbf{O}_t^k = \mathbf{W}_k \mathbf{S}_t \tag{1}$$

Where $\mathbf{W}_k$ is the $N \times M$ mixing matrix, $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_M$ are statistically independent. For the sake of simplicity, we only consider $N = M$ in this paper and $\mathbf{W}_k$ is a square matrix. Then we have

$$\mathbf{S}_t = \mathbf{V}_k \mathbf{O}_t^k \tag{2}$$

Where $\mathbf{V}_k$ is called as the unmixing matrix and $\mathbf{V}_k = (\mathbf{W}_k)^{-1}$.

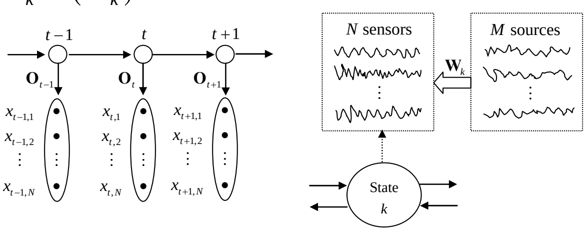


Figure 1: A graphical MSMHMM

A standard MSMHMM is shown as a graphical model in Fig. 1. Then based on the maximum likelihood framework of an independent component analysis process, the multivariate probability of the multi-sensor observation vector can be calculated from the source densities as follows (W. D. Penny, 1998),

$$P(\mathbf{O}_t^k) = \frac{P(\mathbf{S}_t)}{|J_k|} \tag{3}$$

Where $|J_k| = \det(\mathbf{W}_k) = 1/\det(\mathbf{V}_k)$, $P(\mathbf{S}_t) = \prod_{i=1}^{N} P(s_{it}^k)$ and $s_{it}^k = \sum_{j=1}^{N} \mathbf{V}_{kij} x_{tj}$.

Then Eq.(3) can be transformed as

$$\log P(\mathbf{O}_t^k) = \log \frac{P(\mathbf{S}_t)}{|J_k|} = -\log|\det(\mathbf{W}_k)| + \sum_{i=1}^{N} \log P(s_{it}^k) \tag{4}$$

It can be easily seen from Eq.(4) that the probability density of each observation sequence is determined by the probability density of source components. Thus in practice, we should choose proper non-Gaussian source density models to represent non-Gaussian observation sequence, such as vibration signals of helicopter gearboxes. Assuming that non-Gaussian source density model at state k is depicted by the parameter set $\{\theta_k\}$, a multi-sensor mixtured hidden Markov model can be built by the complete parameter set as follows,

$$\lambda_{\text{MSMHMM}} = (\boldsymbol{\pi}, \mathbf{A}, \mathbf{W}, \boldsymbol{\theta}) \tag{5}$$

Next we need to train the MSMHMM before using it, which refers to the estimation of parameters: $\boldsymbol{\pi}$, $\mathbf{A}$, $\mathbf{W}$ and $\boldsymbol{\theta}$.

3. MSMHMM PARAMETERS LEARNING BASED ON EM ALGORITHM

Actually a MSMHMM is improved on a standard HMM, so its parameters learning frame is similar to that of a standard HMM. Thus expectation maximization (EM) algorithm can also be used for MSMHMM parameters learning. That is to say, it needs to maximize, $E(\lambda_{\text{MSMHMM}}, \hat{\lambda}_{\text{MSMHMM}})$, the expectation of the joint log likelihood of an observation sequence $\mathbf{O} = [\mathbf{O}_1, \mathbf{O}_2, \ldots, \mathbf{O}_T]$ and hidden state sequence $\mathbf{Q}$. Here (W. D. Penny, 1998),

$$\begin{aligned} E(\lambda_{\text{MSMHMM}}, \hat{\lambda}_{\text{MSMHMM}}) = & \sum_{\mathbf{Q}} P_{\lambda_{\text{MSMHMM}}}(\mathbf{Q}|\mathbf{O}) \log P_{\hat{\lambda}_{\text{MSMHMM}}}(q_1) \\ & + \sum_{\mathbf{Q}} P_{\lambda_{\text{MSMHMM}}}(\mathbf{Q}|\mathbf{O}) \sum_{t=2}^{T} \log P_{\hat{\lambda}_{\text{MSMHMM}}}(q_t | q_{t-1}) \\ & + \sum_{\mathbf{Q}} P_{\lambda_{\text{MSMHMM}}}(\mathbf{Q}|\mathbf{O}) \sum_{t=1}^{T} \log P_{\hat{\lambda}_{\text{MSMHMM}}}(\mathbf{O}_t | q_t) \end{aligned} \tag{6}$$

Obviously, Eq.(6) composes of three terms which can be used to train MSMHMM model parameters respectively: the first term for the initial state probabilities ($\boldsymbol{\pi}$), the second term for the state transition probabilities ($\mathbf{A}$) and the third one for the observation model parameters, i.e. the mixing matrix ($\mathbf{W}$) and source density parameters ($\boldsymbol{\theta}$).

3.1 Initial state probabilities learning ($\boldsymbol{\pi}$)

The initial state probabilities π_i can be updated by maximizing the first term, $\sum_{\mathbf{Q}} P_{\lambda_{\text{MSMHMM}}}(\mathbf{Q}|\mathbf{O}) \log P_{\hat{\lambda}_{\text{MSMHMM}}}(q_1)$. Furthermore we have,

$$\begin{aligned} \sum_{\mathbf{Q}} P_{\lambda_{\text{MSMHMM}}}(\mathbf{Q}|\mathbf{O}) \log P_{\hat{\lambda}_{\text{MSMHMM}}}(q_1) &= \sum_{i=1}^{K} P_{\lambda_{\text{MSMHMM}}}(q_1 = i|\mathbf{O}) \log P_{\hat{\lambda}_{\text{MSMHMM}}}(i) \\ &= \sum_{i=1}^{K} \gamma_1(i) \log \hat{\pi}_i \end{aligned} \tag{7}$$

Where the constraints are as follows: $\sum_{i=1}^{K}\hat{\pi}_i = 1, \sum_{i=1}^{K}\gamma_1(i) = 1$.

By maximizing Eq.(7), we can get the final update formula as

$$\hat{\pi}_i = \gamma_1(i) \quad (8)$$

Where $\gamma_1(i)$ can be calculated using the forward-backward algorithm.

3.2 State transition probabilities learning ($\mathbf{A}$)

The state transition probabilities $\mathbf{A}$ can be updated by maximizing the second term, $\sum_{\mathbf{Q}} P_{\lambda_{\mathrm{MSMHMM}}}(\mathbf{Q}\,|\,\mathbf{O})\sum_{t=2}^{T}\log P_{\hat{\lambda}_{\mathbf{MSMHMM}}}(q_t\,|\,q_{t-1})$. Furthermore we have,

$$\begin{aligned}
&\sum_{\mathbf{Q}} P_{\lambda_{\mathrm{MSMHM}}}(\mathbf{Q}\,|\,\mathbf{O})\sum_{t=2}^{T}\log P_{\hat{\lambda}_{\mathbf{MSMHM}}}(q_t\,|\,q_{t-1})\\
&=\sum_{i=1}^{K}\sum_{j=1}^{K}\sum_{t=2}^{T} P_{\lambda_{\mathrm{MSMHM}}}(q_t=j, q_{t-1}=i\,|\,\mathbf{O})\log P_{\hat{\lambda}_{\mathbf{MSMHM}}}(q_t\,|\,q_{t-1})\\
&=\sum_{i=1}^{K}\sum_{j=1}^{K}(\sum_{t=1}^{T-1}\xi_t(i,j))\log(a_{ij})
\end{aligned} \quad (9)$$

where $\xi_t(i,j)=\dfrac{\alpha_t(i)a_{ij}b_j(\mathbf{O}_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{K}\sum_{j=1}^{K}\alpha_t(i)a_{ij}b_j(\mathbf{O}_{t+1})\beta_{t+1}(j)}$.

By maximizing Eq.(9), we can get the final update formula as

$$\hat{a}_{ij}=\frac{\sum_{t=1}^{T}\xi_t(i,j)}{\sum_{t=1}^{T}\gamma_t(i)} \quad (10)$$

3.3 Mixing matrix ($\mathbf{W}$) and source density parameters ($\boldsymbol{\theta}$) learning

The observation model parameters, i.e. the mixing matrix ($\mathbf{W}$) and source density parameters ($\boldsymbol{\theta}$), can be updated by maximizing the third term, $\sum_{\mathbf{Q}} P_{\lambda_{\mathrm{MSMHMM}}}(\mathbf{Q}\,|\,\mathbf{O})\sum_{t=1}^{T}\log P_{\hat{\lambda}_{\mathbf{MSMHMM}}}(\mathbf{O}_t\,|\,q_t)$. However, the update process is determined by the observation model. In this paper in order to represent non-Gaussian vibration signals of a helicopter gearbox, we need to choose proper non-Gaussian source models in MSMHMM. (S. J. Roberts, 1998) has pointed out that a signal consisting of multiple sinusoids has a multimodal probability density function (PDF) and generalized autoregressive (GAR) source models can provide better unmixing than generalized exponential (GE) source models for multimodal PDFs sources. On the other hand, as we all know that a rotating machine works under periodic motions and its vibration source are often multi-frequencies sinusoids, so GAR source models will be used in this paper.

A GAR source model is shown as follows,

$$s_{it}^k=-\sum_{d=1}^{p}c_i^k[d]s_{i(t-d)}^k+e_{it}^k, 1\le i\le N, 1\le t\le T, 1\le k\le K \quad (11)$$

Where $c_i^k[\cdot]$ are the GAR coefficients for the ith source at state k and denoted as $\mathbf{c}_i^k$, e_{it}^k is a non-Gaussian additive noise and p is the model order. In practice, e_{it}^k denotes the GAR prediction error and can be calculated as,

$$\begin{aligned}
e_{it}^k &= s_{it}^k-\hat{s}_{it}^k\\
\hat{s}_{it}^k &= -\sum_{d=1}^{p}c_i^k[d]\hat{s}_{i(t-d)}^k
\end{aligned} \quad (12)$$

Then each GAR source density at state k is (S. J. Roberts, 1998)

$$P(s_{it}^k)=\frac{R_i^k\left(\beta_i^k\right)^{1/R_i^k}}{2\Gamma(1/R_i^k)}\exp\left(-\beta_i^k\,|\,e_{it}^k\,|\right)^{R_i^k} \quad (13)$$

Where $\Gamma(\cdot)$ is the gammar function, R_i^k , β_i^k are the two density parameter for ith source at state k.

So in this paper source density parameters ($\boldsymbol{\theta}$) composes of {p, $\mathbf{c}_i^k$, R_i^k , β_i^k }. Next we will train these parameters according to the third term in Eq.(6). Furthermore we have,

$$\begin{aligned}
&\sum_{\mathbf{Q}} P_{\lambda_{\mathrm{MSMHMM}}}(\mathbf{Q}\,|\,\mathbf{O})\sum_{t=1}^{T}\log P_{\hat{\lambda}_{\mathbf{MSMHMM}}}(\mathbf{O}_t\,|\,q_t)\\
&=\sum_{k=1}^{K}\sum_{t=1}^{T}P_{\lambda_{\mathrm{MSMHMM}}}(q_t=k\,|\,\mathbf{O})\log P_{\hat{\lambda}_{\mathbf{MSMHMM}}}(\mathbf{O}_t\,|\,k)\\
&=\sum_{k=1}^{K}\sum_{t=1}^{T}\gamma_t[k]\log P_{\hat{\lambda}_{\mathbf{MSMHMM}}}(\mathbf{O}_t\,|\,k)
\end{aligned} \quad (14)$$

Where $\log P_{\hat{\lambda}_{\mathbf{MSMHMM}}}(\mathbf{O}_t\,|\,k)$ can be calculated by Eq.(4). That is,

$$\begin{aligned}
&\log P_{\hat{\lambda}_{\mathbf{MSMHMM}}}(\mathbf{O}_t\,|\,k)=\log P_{\hat{\lambda}_{\mathbf{MSMHMM}}}(\mathbf{O}_t^k)\\
&=-\log|\det(\mathbf{W}_k)|+\sum_{i=1}^{N}\log P(s_{it}^k)
\end{aligned} \quad (15)$$

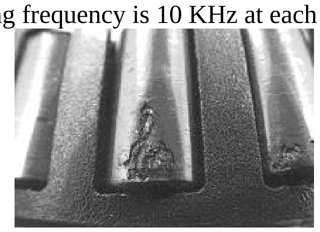
By substituting Eq.(12), (13), (15) into Eq.(14), updating of { $\mathbf{W}$, $\mathbf{c}_i^k$, R_i^k , β_i^k } can be derived by differentiating Eq.(14) on W_{ij}^k , $\mathbf{c}_i^k$, R_i^k , β_i^k respectively.

Besides of $\boldsymbol{\pi}$, $\mathbf{A}$, $\mathbf{W}$ and $\boldsymbol{\theta}$, there are some other parameters needed to be determined, including the number of sources, N, the number of states, K, and the order of GAR, p. How to select these parameters is a problem to be solved, which will not be discussed deeply in this paper.

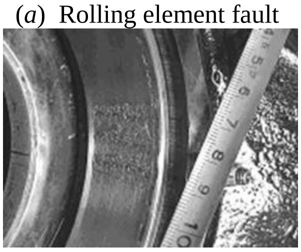
By now, the algorithm of MSMHMM parameters learning can be implemented by Matlab software.

4. A CASE OF BEARING FAULT DIAGNOSTICS IN A HELICOPTER GEARBOX

In the experiment, a bearing in a helicopter gearbox is selected and two classical faults are seeded on it, i.e. rolling element fault and outer race fault, shown in Fig. 2. Then vibration signals are collected from five sensors under normal and faulty conditions respectively. The sampling frequency is 10 KHz at each channel.



(*a*) Rolling element fault



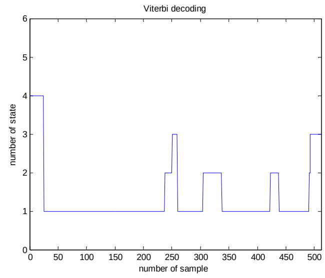
(*b*) Outer race fault

Figure 2: Two kinds of faults on the bearing

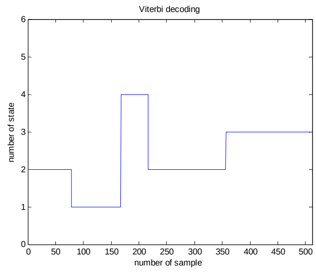
4.1 MSMHMMs training

In the scheme of MSMHMM-based fault diagnostics, firstly it needs to determine the number of sources, the number of states and the order of GAR. Because the gearbox consists of five main components in this paper, the number of sources is selected as N=5 here. Then five vibration sensors for observations are mounted on the gearbox. The number of states is selected as K=4 and the order of GAR is selected as p=6 artificially. The length of observation sequence is selected as T=512.

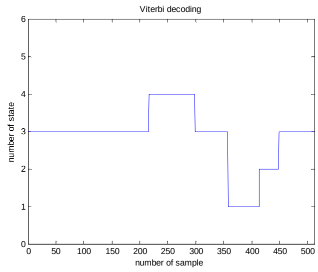
By initializing initial probabilities, $\boldsymbol{\pi}_{K\times 1}$, transition matrix, $\mathbf{A}_{K\times K}$, mixing matrix, $\mathbf{W}_{K\times N\times N}$, source density parameters, $\mathbf{c}_{K\times N\times p}$, $R_{K\times N}$, $\beta_{K\times N}$, different MSMHMMs under three conditions are trained based on 10 training samples respectively. After training, we can get three MSMHMMs (MSMHMM1 for normal, MSMHMM2 for rolling element fault and MSMHMM3 for outer race fault) and the corresponding state sequences are shown in Fig. 3.



(*a*) Normal



(*b*) Rolling element fault



(*c*) Outer race fault

Figure 3: State sequences for different bearing conditions after training

4.2 MSMHMMs-based bearing faults identification

After three MSMHMMs has been built and trained, we can use them to isolate different conditions using testing samples. The number of testing samples under each condition is selected as 15. Then each MSMHMM is used to analyze normal, rolling element fault and outer race fault samples to test its classification ability respectively, and then the corresponding results are shown as Fig. 4~Fig. 6. In Fig. 4, MSMHMM_1 is used and the maximum log-likelihood corresponds to normal condition, so MSMHMM_1 identify health condition of the bearing accurately. Similar results can be obtained in Fig. 5 and Fig. 6. Thus it demonstrates that MSMHMMs can identify faults in the helicopter gearbox accurately.

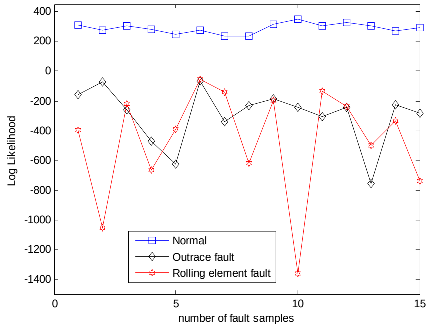


Figure 4: Identified results based on $MSMHMM_1$

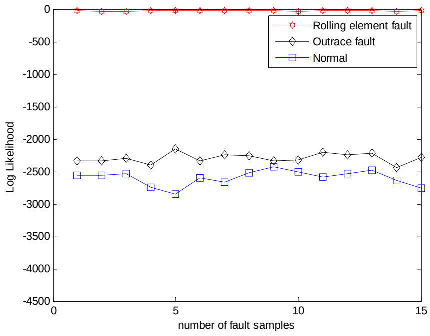


Figure 5: Identified results based on $MSMHMM_2$

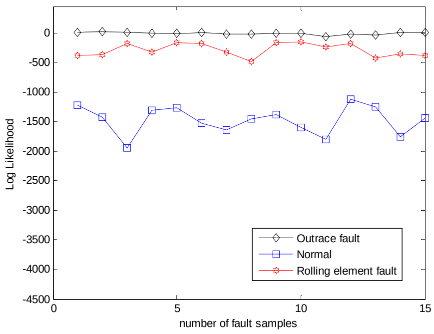


Figure 6: Identified results based on $MSMHMM_3$

In order to testify that Gaussian observation HMM (GHMM) may not fit for bearing fault diagnostics in the helicopter gearbox, we will use the above training samples to build and train three GHMMs ($GHMM_1$ for normal, $GHMM_2$ for rolling element fault and $GHMM_3$ for outer race fault)), where the number of states is also selected as K=4. Then three GHMMs are used to analyze normal, rolling element fault and outer race fault testing samples, and then the corresponding results are shown as Fig. 7~Fig. 9 respectively. It can be seen that GHMMs cannot identify the anticipated condition and provide mistaken results. The reason may be that observation sequences from the helicopter gearbox are truly non-Gaussian and non-stationary. Also we can find the log-likelihood values in Fig. 7~ Fig. 9 fluctuate more than those in Fig. 4~Fig. 6. The reason may be that the observation sequences are non-stationary. Thus it testifies that the proposed MSMHMM is a better tool than traditional GHMM for bearing fault diagnostics.

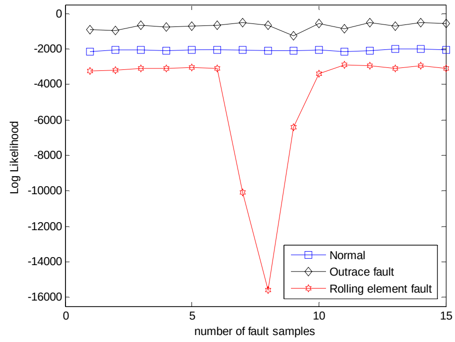


Figure 7: Identified results based on $GHMM_1$

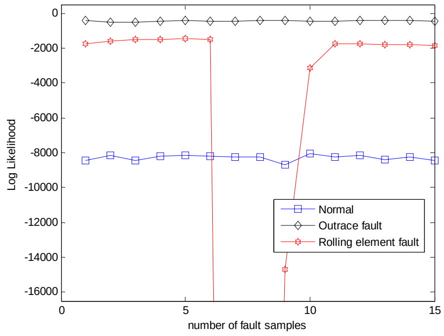


Figure 8: Identified results based on $GHMM_2$

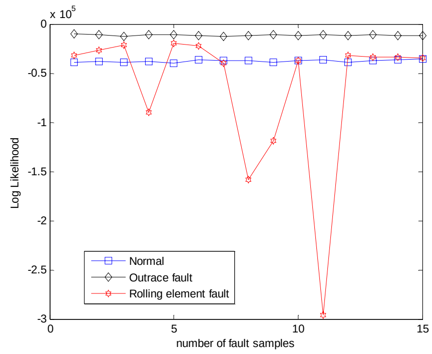


Figure 9: Identified results based on $GHMM_3$

5. CONCLUSION

This paper has presented a MSMHMM-based bearing fault diagnostics method for complex rotating machines using multi-sensor observation signals. Each sensor signals was considered as the mixture of non-Gaussian sources, which can depict non-Gaussian observation sequences well. Then its parameter learning algorithm was proposed based on EM algorithm. In the end through the experimental study on a bearing in a helicopter gearbox, we have testified that MSMHMMs can identify bearing faults more accurately than traditional GHMMs. Furthermore, the proposed MSMHMMs can be extended for fault diagnostics of other complex rotating machines.

Future work will include how to determine the number of states and the order of GAR models in MSMHMMs theoretically, which may be solved by understanding particular mechanical systems and their working processes.

ACKNOWLEDGMENT

The authors sincerely appreciate the funding provided to this research by the National Nature Science Foundation of China under grant number 50805142. Also the authors would like to thank Dr. W. D. Penny for his help and contribution to this work.

REFERENCES

D. S. Paul, J. P. Darryll (2005). A review of vibration-based techniques for helicopter transmission diagnostics, *Journal of Sound and Vibration*, vol. 282, pp. 475–508.

L.R. Rabiner (1989). A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of IEEE*, vol. 77, pp. 257–285.

P. Baruah, R. B. Chinnam (2005). HMMs for diagnostics and prognostics in machining processes, *International Journal of Production Research*, vol. 43, pp. 1275–1293.

J. M. Leea, S. J. Kima, Y. Hwang, et al.(2004). Diagnosis of mechanical fault signals using continuous hidden Markov model, *Journal of Sound and Vibration*, vol. 276, pp.1065–1080.

C. Bunks, D. McCathy, T. Al-Ani (2000). Condition-based maintenance of machines using hidden Markov models, *Mechanical Systems and Signal Processing*, vol. 14, pp. 597–612.

F. G. Wang, Y. B. Li , Z. G. Luo (2009). Fault classification of rolling bearing based on reconstructed phase space and Gaussian mixture model, *Journal of Sound and Vibration*, vol. 323, pp. 1077–1089.

L. Bouillaut, M. Sidahmed (2001). Helicopter gearbox vibrations: cyclostationary analysis or bilinear approach, *International symposium on signal processing and its application*, pp. 13–16.

T-W. Lee, M. Girolami, A.J. Bell, et al.(2000). A unifying information-theoretic framework for independent component analysis, *Computers & Mathematics with Applications*, vol. 31, pp.1–21.

W. D. Penny, S. J. Roberts (1998). Hidden Markov models with extended observation densities, *Technical Report TR-98-15* .

R. Everson, S. J. Roberts (1998). Independent Component Analysis: A flexible non-linearity and decorrelating manifold approach. *Proceedings of IEEE conference on neural network and signal processing*, pp. 33–42.

Z. S. Chen was born in Anhui Province, China, on Aug. 13, 1977. He received the B.E. and Ph.D. degrees in Mechatronic Engineering from the National University of Defense Technology, P. R. China in 1999 and 2004, respectively. From 2008, he worked as an associate professor in Institute of Mechatronic Engineering, National University of Defense Technology. Current research interests include condition monitoring and fault diagnosis, mechanical signal processing and data fusion, vibration energy harvesting, etc.