Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Sound and Vibration 276 (2004) 1065–1080
JOURNAL OF SOUND AND VIBRATION
www.elsevier.com/locate/jsvi

Diagnosis of mechanical fault signals using continuous hidden Markov model

Jong Min Lee[a], Seung-Jong Kim[a], Yoha Hwang[a,*], Chang-Seop Song[b]

[a] *Tribology Research Center, Korea Institute of Science and Technology, Songbuk Hawolgokdong 39-1, Seoul 136-791, South Korea*

[b] *School of Mechanical Engineering, Hanyang University, 17 Haengdang, Sengdong, Seoul 133-791, South Korea*

Received 11 April 2003; accepted 15 August 2003

Abstract

Hidden Markov Model (HMM) has been actively studied in speech recognition since 1960s and increasingly used in many other fields. However, its application to mechanical engineering has been very limited. HMM is not only very accurate and robust in analyzing signals but also can be a very powerful method of predicting target system's condition change. In this paper, continuous HMM (CHMM) has been tuned to be used in mechanical signal analysis and applied to diagnose of various mechanical signals including rotor fault signals. The results show HMM's big potential as an intelligent condition monitoring tool based on its accuracy, robustness, and forecasting ability.

1. Introduction

Hidden Markov Model (HMM) has been a dominant method in speech recognition since 1960s and become popular in various areas in last decade. The increasing popularity of HMM is based on two facts; rich mathematical structure and proven accuracy on critical applications [1]. It has a doubly embedded stochastic process with an underlying stochastic process that can be observed through another set of stochastic processes. This structure of HMM is useful for modeling a sequence that does not look like a stochastic process but has a hidden stochastic process. Although it has become popular in various areas like signal analysis and pattern recognition, its use in mechanical engineering field has been very limited.

In this paper, the potential of the HMM, especially continuous HMM (CHMM) which is the HMM of a vector sequence, for intelligent machine condition monitoring has been studied. In the

*Corresponding author. Tel.: +82-2-958-5656; fax: +82-2-958-5659.
E-mail address: yoha@kist.re.kr (Y. Hwang).

0022-460X/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2003.08.021

intelligent machine condition monitoring, expert system, neural network, and fuzzy method are generally used as advanced techniques. However, their accuracies are often unreliable and they also have their own problems to be widely used. In an expert system, complex algorithm has to be updated by an experienced engineer for every target system. Neural network method needs training patterns of all failure modes of the target system and has to be learned again to add other failure modes. Fuzzy system generally has difficulties in selection of suitable membership function for target system. HMM also needs patterns for every symptom of the system, however, new symptom can be simply trained and added to existing pattern library. It makes a decision through a statistical process, which is basically stable and robust. So once symptoms are well trained, it makes a very accurate decision. HMM also has a very good trend forecasting ability, which is essential for early warning.

Among applications of HMM to mechanical signals, Smyth has used CHMM for antenna pointing system's fault detection and showed that CHMM is more accurate than neural network and Gaussian classifier, however, his method is more like a Markov chain than a CHMM method [2]. Bunks suggested that HMM could be a very good tool to monitor helicopter gearbox failures, but did not apply it to real model [3]. Ertunc claimed that he had used CHMM for drill wear monitoring and achieved a good result, however he did not specify his CHMM method [4]. Kwon applied discrete HMM (DHMM) which is the HMM of a symbol sequence to identify accident patterns in nuclear power plant and showed its robustness [5]. The authors have successfully applied CHMM method using AR for feature vector for chatter signal analysis and forecasting [6].

According to the authors' own research, characteristic vectors of mechanical failure signals have very small dispersion and this makes it very difficult to get statistical characteristics of some groups which have no or very small number of vectors [6]. This often makes the calculation blows up. It has turned out that this problem is common in mechanical signal analysis so authors have made several modifications to conventional CHMM algorithm. Scaled forward/backward variables and initialization of CHMM parameters using maximum distance clustering method were used. Low threshold and option of setting off-diagonal element's value to zero in covariance matrix were also introduced. The authors also have derived all equations used for CHMM training to use multiple observation vector sequences. CHMM method with AR coefficients was used to detect and predict chatter of a lathe and CHMM method with spectrum and filter bank was used to diagnose various rotor failures. CHMM method's accuracy and early trend detection ability have been proved with experimental data.

2. Procedure of CHMM training and fault diagnosis

Fig. 1 shows the general procedure of using CHMM in diagnosis. First, feature vector extraction method has to be selected. This is a very important step because the feature vector should suitably represent the information hidden in the raw data. For example, if AR model is selected, then all model parameters like model order have to be carefully selected to correctly represent the signal characteristics. The next step is training on HMM. In this step, model structure has to be selected. The model structure includes the numbers of state and mixture and the structure of transition matrix, like ergodic model or left-to-right model [1]. Training patterns

are also needed. So, signals for each symptom have to be sampled and converted into feature vectors. These feature vectors are training patterns. After that, the best CHMM for each symptom is trained. Using this step, all CHMMs for every symptom are calculated. The last step is fault diagnosis. In this step, the log-likelihood of each trained CHMM on the sampled data of the

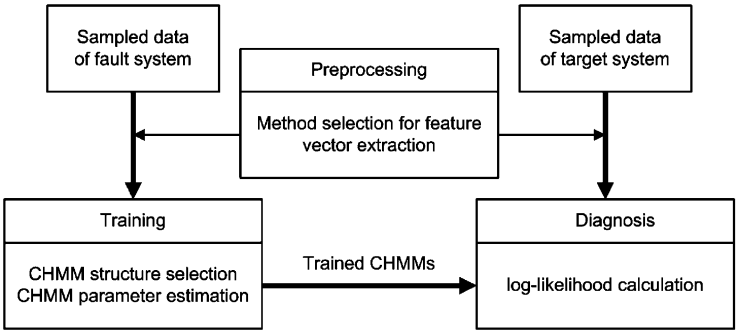

Fig. 1. General procedure of using CHMM in diagnosis.

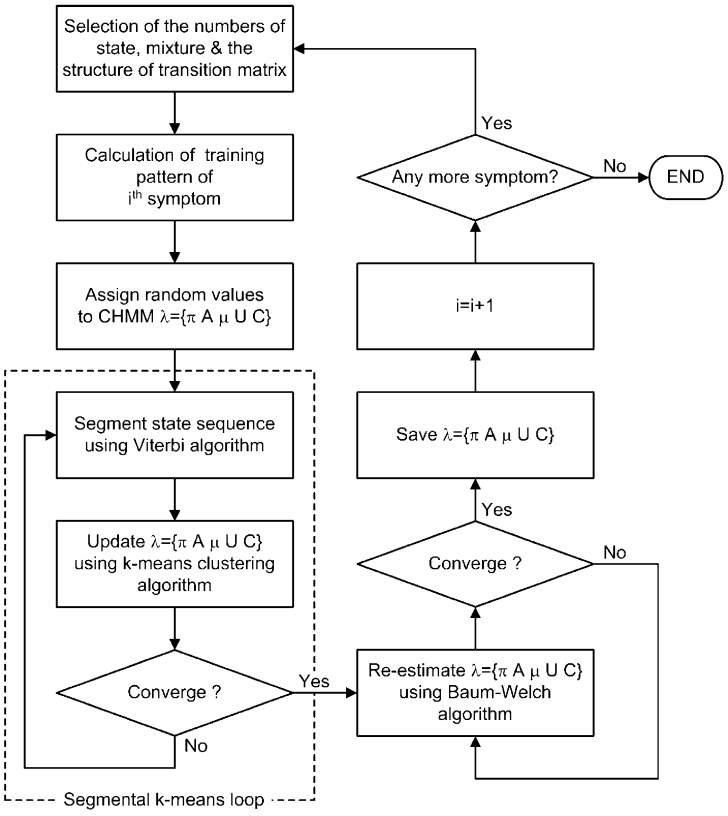

Fig. 2. Training procedure of CHMM.

target machine is calculated and the model which has the highest log-likelihood is selected as a fault symptom.

2.1. Training of CHMM

Fig. 2 shows training procedure of CHMM. Parameters of CHMM have to be decided first and training patterns of interested symptoms are calculated. Then, random values are given to the CHMM parameters and the initial parameters are calculated using segmental k-means loop. Next, CHMM parameters are re-estimated until Baum–Welch algorithm converges. Then model parameters of ith symptom are fixed. This procedure is repeated for every symptom.

2.2. Diagnosis using CHMM

Diagnosis procedure is shown in Fig. 3. From the target machine that is to be monitored, signals are sampled and feature vectors are extracted to get vector sequence. This vector sequence is current pattern from the target machine. In the next step, log-likelihood of each trained CHMM on the pattern is calculated. The model that has the highest log-likelihood is selected and it shows the current condition of the target machine.

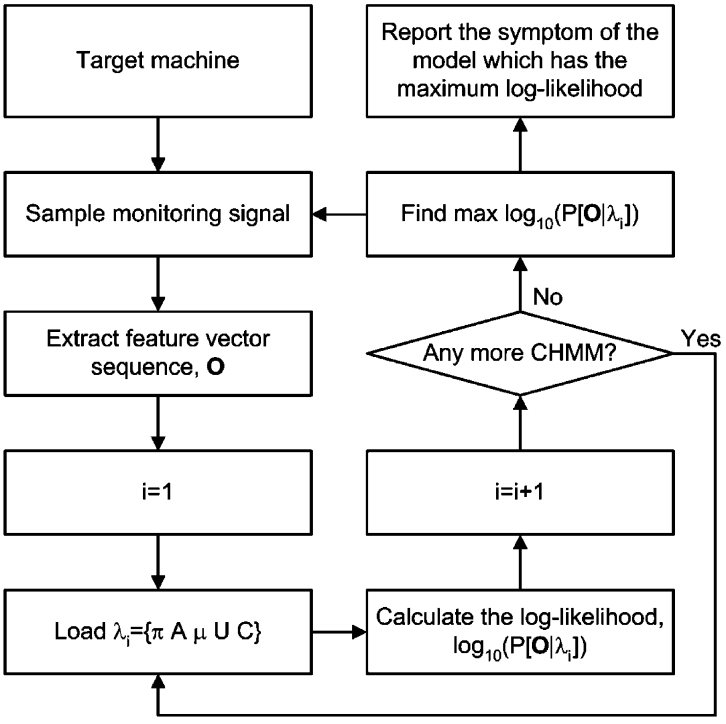

Fig. 3. Diagnosis procedure of CHMM.

3. CHMM algorithm for machine condition monitoring

To correctly model a robust CHMM, it is needed to use many observation vector sequences of a fault simultaneously. However, previous CHMM researches have explained equations to handle only a single observation vector sequence as in Refs. [1,7]. In this paper, all equations needed to handle multiple observation vector sequences have been derived. CHMM algorithm also has been tuned to handle mechanical fault signals and the maximum distance clustering method has been developed to reduce iteration times of k-means clustering algorithm. This section explains all the extensions and modifications in detail.

3.1. Definition of CHMM

Let us assume that there is a set of K observation vector sequences as shown below

$$\mathbf{O} = \{O^{(1)}, O^{(2)}, \ldots, O^{(K)}\}, \tag{1}$$

where $O^{(k)} = \{\vec{o}_1^{(k)}, \vec{o}_2^{(k)}, \ldots, \vec{o}_{T_k}^{(k)}\}$ is the kth observation vector sequence and $\vec{o}_t^{(k)}$ is the D-dimensional tth observation vector of $O^{(k)}$. To get CHMM which models statistical characteristics of such vector sequences, N states and M mixtures are selected as

$$S = \{s_1, s_2, \ldots, s_N\}, \tag{2}$$

$$X = \begin{Bmatrix} x_{11} & x_{12} & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & x_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NM} \end{Bmatrix} \tag{3}$$

where x_{nm} is mth mixture of state s_n. Number of state N and number of mixture M are selected by rule of thumb. Then, CHMM can be expressed as

$$\lambda = (\pi, A, C, \mu, U). \tag{4}$$

If q_t is a state at time t and $P[\cdot]$ is probability, parameters of CHMM, initial state probability distribution π, state transition probability distribution A, mixture gain C, mean μ, and covariance U are defined as

$$\pi = [\pi_1 \ \ \pi_2 \ \ \cdots \ \ \pi_N]^{\mathrm{T}}, \quad \pi_n = P[q_1 = s_n] \geqslant 0, \quad \sum_{n=1}^{N} \pi_n = 1, \tag{5}$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix}, \quad a_{ij} = P[q_{t+1} = s_j | \, q_t = s_i] \geqslant 0, \quad \sum_{j=1}^{N} a_{ij} = 1 \quad \forall i, \tag{6}$$

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1M} \\ c_{21} & c_{22} & \cdots & c_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ c_{N1} & c_{N2} & \cdots & c_{NM} \end{bmatrix}, \quad c_{nm} \geqslant 0, \quad \sum_{m=1}^{M} c_{nm} = 1 \quad \forall n, \tag{7}$$

$$\mu = \begin{bmatrix} \vec{\mu}_{11} & \vec{\mu}_{12} & \cdots & \vec{\mu}_{1M} \\ \vec{\mu}_{21} & \vec{\mu}_{22} & \cdots & \vec{\mu}_{1M} \\ \vdots & \vdots & \ddots & \vdots \\ \vec{\mu}_{N1} & \vec{\mu}_{N2} & \cdots & \vec{\mu}_{NM} \end{bmatrix}, \tag{8}$$

$$U = \begin{bmatrix} U_{11} & U_{12} & \cdots & U_{1M} \\ U_{21} & U_{22} & \cdots & U_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ U_{N1} & U_{N2} & \cdots & U_{NM} \end{bmatrix}, \tag{9}$$

where $\vec{\mu}_{nm}$ is a D-dimensional mean vector and U_{nm} is a $D \times D$ dimensional covariance matrix at mixture x_{nm}. If the probability density function (pdf) $b_{t,n}^{(k)}$ of observation vector $\vec{o}_t^{(k)}$ in state s_n is the Gaussian density function, the pdf is given as

$$b_{t,n}^{(k)} = \sum_{m=1}^{M} c_{nm} \times b_{t,nm}^{(k)} = \sum_{m=1}^{M} c_{nm} \times \frac{\exp(-0.5(\vec{o}_t^{(k)} - \vec{\mu}_{nm})^{\mathrm{T}} \cdot U_{nm}^{-1}(\vec{o}_t^{(k)} - \vec{\mu}_{nm}))}{\sqrt{(2 \times 3.141592)^2 \times |U_{nm}|}}, \tag{10}$$

where $b_{t,nm}^{(k)}$ is the pdf of observation vector $\vec{o}_t^{(k)}$ at mixture x_{nm}.

3.2. Scaled forward/backward and posteriori variables

For the calculation of likelihood $P[O^{(k)}|\lambda]$ to observe observation vector sequence $O^{(k)}$ from CHMM λ and re-estimation of CHMM parameters, iteration methods are generally used to reduce calculation times. In this paper, forward/backward variables [1,6,7] and three different posteriori variables [1,7] are used for this. To calculate forward/backward variable it is needed to multiply probability repetitively, however, it often results in an underflow problem. Mechanical signals usually have relatively very small dispersion of statistical property so this problem happens very often. To prevent this problem, scaled forward/backward variables are used [1,7]. Forward variables $\alpha_{t,n}^{(k)}$ are defined as in Eq. (11) and scaled forward variables $\hat{\alpha}_{t,n}^{(k)}$ are calculated inductively as in Eqs. (12) and (13) using scale coefficients $g_t^{(k)}$:

$$\alpha_{t,n}^{(k)} = P[\vec{o}_1^{(k)} \ \vec{o}_2^{(k)} \ \cdots \ \vec{o}_t^{(k)}, \ q_t = s_n|\lambda], \tag{11}$$

$$\begin{aligned} \alpha_{1,n}^{(k)} &= \pi_n b_{1,n}^{(k)}, \\ g_1^{(k)} &= \frac{1}{\sum_{\ell=1}^{N} \alpha_{1,\ell}^{(k)}} \quad (n = 1, 2, \ldots, N), \\ \hat{\alpha}_{1,n}^{(k)} &= g_1^{(k)} \alpha_{1,n}^{(k)}, \end{aligned} \tag{12}$$

$$\widehat{\alpha}_{t,n}^{(k)} = \left(\sum_{\ell=1}^{N} \hat{\alpha}_{t-1,\ell}^{(k)} \alpha_{\ell,n}\right) b_{t,n}^{(k)},$$

$$g_t^{(k)} = \frac{1}{\sum_{\ell=1}^{N} \widehat{\alpha}_{t,\ell}^{(k)}} \quad \begin{pmatrix} n = 1, 2, \ldots, N \\ t = 2, 3, \ldots, T_k \end{pmatrix},$$

$$\hat{\alpha}_{t,n}^{(k)} = g_t^{(k)} \widehat{\alpha}_{t,n}^{(k)}. \tag{13}$$

Then, the log observation likelihoods $P^{(k)}$ of each observation vector sequence $O^{(k)}$ and the log observation likelihood of the set of them are expressed as

$$\log(P[\mathbf{O}|\lambda]) = \log\left(\prod_{k=1}^{K} P[O^{(k)}|\lambda]\right) = \sum_{k=1}^{K} \log(P^{(k)}) = -\sum_{k=1}^{K} \sum_{t=1}^{T_k} \log g_t^{(k)}. \tag{14}$$

If we define backward variables $\beta_{t,n}^{(k)}$ as in Eq. (15), scaled backward variables $\hat{\beta}_{t,n}^{(k)}$ are calculated inductively as Eqs. (16) and (17) using scale coefficients $g_t^{(k)}$ of scaled forward variables.

$$\beta_{t,n}^{(k)} = P[\vec{o}_{t+1}^{(k)} \ \vec{o}_{t+2}^{(k)} \ \cdots \ \vec{o}_{T_k}^{(k)} \,|\, q_t = s_n, \lambda], \tag{15}$$

$$\begin{aligned} \beta_{T_k,n}^{(k)} &= 1 \\ \hat{\beta}_{T_k,n}^{(k)} &= g_{T_k}^{(k)} \beta_{T_k,n}^{(k)} \end{aligned} \quad (n = 1, 2, \ldots, N), \tag{16}$$

$$\begin{aligned} \widehat{\beta}_{t,n}^{(k)} &= \textstyle\sum_{\ell=1}^{N} a_{n,\ell} b_{t+1,\ell}^{(k)} \hat{\beta}_{t+1,\ell}^{(k)} \\ \hat{\beta}_{t,n}^{(k)} &= g_t^{(k)} \widehat{\beta}_{t,n}^{(k)} \end{aligned} \quad \begin{pmatrix} n = 1, 2, \ldots, N \\ t = T_k - 1, T_k - 2, \ldots, 1 \end{pmatrix}. \tag{17}$$

Three kinds of posteriori variables are introduced to make it simple to re-estimate CHMM parameters [1,7]. The first posteriori variable is likelihood $\gamma_{t,n}^{(k)}$ in which q_t is s_n when CHMM λ and observation vector sequence $O^{(k)}$ are given. The second posteriori variable is likelihood $\xi_{t,ij}^{(k)}$ in which q_t is s_i and q_{t+1} is s_j. The third posteriori variable is likelihood $\gamma_{t,nm}^{(k)}$ in which mixture r_t at time t is x_{nm} under same condition. These three posteriori variables are expressed as

$$\gamma_{t,n}^{(k)} = P[q_t = s_n|O^{(k)}, \lambda] = \frac{\hat{\alpha}_{t,n}^{(k)} \hat{\beta}_{t,n}^{(k)}}{\sum_{\ell=1}^{N} \hat{\alpha}_{t,\ell}^{(k)} \hat{\beta}_{t,\ell}^{(k)}} = \sum_{\ell=1}^{N} \xi_{t,n\ell}^{(k)} = \sum_{m=1}^{M} \gamma_{t,nm}^{(k)}, \tag{18}$$

$$\xi_{t,ij}^{(k)} = P[q_t = s_i, q_{t+1} = s_j|O^{(k)}, \lambda] = \frac{\hat{\alpha}_{t,i}^{(k)} a_{ij} b_{t+1,j}^{(k)} \hat{\beta}_{t,n}^{(k)}}{\sum_{\ell=1}^{N} \hat{\alpha}_{t,\ell}^{(k)} \hat{\beta}_{t,\ell}^{(k)}}, \tag{19}$$

$$\gamma_{t,nm}^{(k)} = P[q_t = s_n, r_t = x_{nm}|O^{(k)}, \lambda] = \frac{\hat{\alpha}_{t,n}^{(k)} \hat{\beta}_{t,n}^{(k)}}{\sum_{\ell=1}^{N} \hat{\alpha}_{t,\ell}^{(k)} \hat{\beta}_{t,\ell}^{(k)}} \times \frac{c_{nm} b_{t,nm}^{(k)}}{\sum_{\ell=1}^{M} c_{n\ell} b_{t,n\ell}^{(k)}}. \tag{20}$$

3.3. Estimation of CHMM parameters

If multiple observation vector sequences for training are given and CHMM is defined as above, the parameters of CHMM can be estimated as follows.

Initial estimation: To make the re-estimation converge to the global minimum rather than a local minimum, the parameters of CHMM are initialized as follows [1,6,7]:

(1) Initialize arbitrarily the initial state probability distribution and the state transition probability distribution subjected to the Eqs. (5) and (6). N observation vectors having the largest distance value among all observation vectors $\vec{o}_t^{(k)}$ are selected as centers of each N state. Then, all observation vectors are grouped into N groups by their distances to the selected N centers. Next, for each group, M observation vectors in which each vectors having the largest distance from their centers are selected as M mixture centers. Then observation vectors in each group are re-grouped into M clusters based on their distance to M centers. The total feature vectors are clustered into $N \times M$ and initial value of mean, covariance and mixture gain are calculated using the following:

$$\begin{aligned}
&\vec{\mu}_{nm} = \text{sample mean vector of the observation vectors classified in } x_{nm},\\
&U_{nm} = \text{sample covariance matrix of the observation vectors classified in } x_{nm},\\
&c_{nm} = \frac{\text{number of observation vector in } x_{nm}}{\text{number of observation vectors in } s_n}.
\end{aligned} \tag{21}$$

(2) Once $O^{(k)}$ is observed from CHMM λ estimated by log-Viterbi algorithm, the best state sequence $q_t^{(k)*}$ is calculated as follows. First, $b_{t,n}^{(k)}$ is calculated by Eq. (10) and CHMM parameters are converted to log values as

$$\begin{aligned}
&\tilde{\pi}_n = \log(\pi_n) \qquad (n = 1, 2, \ldots, N),\\
&\tilde{b}_{t,n}^{(k)} = \log(b_{t,n}^{(k)}) \quad (n = 1, 2, \ldots, N,\ \ t = 1, 2, \ldots, T_k,\ \ k = 1, 2, \ldots, K),\\
&\tilde{a}_{ij} = \log(a_{ij}) \qquad (i = 1, 2, \ldots, N,\ \ j = 1, 2, \ldots, N).
\end{aligned} \tag{22}$$

Next, log values of the best score variable $\tilde{\delta}_{t,n}^{(k)}$ and the best score argument array $\psi_{t,n}^{(k)}$ are calculated recursively for each $O^{(k)}$ by

$$\begin{aligned}
&\tilde{\delta}_{1,n}^{(k)} = \tilde{\pi}_n + \tilde{b}_{1,n}^{(k)}\\
&\psi_{1,n}^{(k)} = 0
\end{aligned} \qquad \begin{pmatrix} n = 1, 2, \ldots, N \\ k = 1, 2, \ldots, K \end{pmatrix}, \tag{23}$$

$$\begin{aligned}
&\tilde{\delta}_{t,n}^{(k)} = \max_{1 \leqslant i \leqslant N} [\tilde{\delta}_{t-1,i}^{(k)} + \tilde{a}_{in}] + \tilde{b}_{t,n}^{(k)}\\
&\psi_{t,n}^{(k)} = \arg\max_{1 \leqslant i \leqslant N} [\tilde{\delta}_{t-1,i}^{(k)} + \tilde{a}_{in}]
\end{aligned} \qquad \begin{pmatrix} n = 1, 2, \ldots, N \\ t = 2, 3, \ldots, T_k \\ k = 1, 2, \ldots, K \end{pmatrix}. \tag{24}$$

Then, the best state sequence $q_t^{(k)*}$ is calculated using Eqs. (25) and (26) which are path backtracking methods:

$$q_T^{(k)*} = s_{n^*}, \quad n^* = \arg \max_{1 \leqslant i \leqslant N} [\tilde{\delta}_{T,i}^{(k)}] \ \ (k = 1, 2, \ldots, K), \tag{25}$$

$$q_t^{(k)*} = s_{n^*}, \quad n^* = \psi_{t+1}^{(k)}(q_{t+1}^{(k)*}) \ \ (t = T_k - 1, T_k - 2, \ldots, 1, \ \ 1 \leqslant k \leqslant K). \tag{26}$$

(3) Estimate the CHMM parameters using segmental k-means clustering. Let us assume $\Delta(y)$ is the Direc delta function as

$$\Delta(y) = \begin{cases} 1 & \text{if } y = 0, \\ 0 & \text{if } y \neq 0. \end{cases} \tag{27}$$

Then, the initial state probability distribution and the state transition probability distribution are estimated as follows:

$$\begin{aligned} \bar{\pi}_n &= \frac{\sum_{k=1}^{K} \Delta(q_1^{(k)*} - s_n)}{K} \quad (n = 1, 2, \ldots, N), \\ \bar{a}_{ij} &= \frac{\sum_{k=1}^{K} \sum_{t=2}^{T_k} \Delta((q_{t-1}^{(k)*} - s_i) + (q_t^{(k)*} - s_j))}{\sum_{k=1}^{K} \sum_{t=2}^{T_k} \Delta(q_{t-1}^{(k)*} - s_i)} \quad (i = 1, 2, \ldots, N,\ j = 1, 2, \ldots, N). \end{aligned} \tag{28}$$

Observation vectors grouped for each state in step 2 are re-grouped into M clusters in each state following procedures in step (1). Then mean vector, covariance matrix and mixture gain for each mixture are calculated as

$$\begin{aligned} \bar{\vec{\mu}}_{nm} &= \frac{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \Delta((q_t^{(k)*} - s_n) + (r_t^{(k)} - x_{nm})) \times \vec{o}_t^{(k)}}{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \Delta(q_t^{(k)*} - s_n)}, \\ \bar{U}_{nm} &= \frac{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \Delta((q_t^{(k)*} - s_n) + (r_t^{(k)} - x_{nm})) \times (\vec{o}_t^{(k)} - \bar{\vec{\mu}}_{nm}) \times (\vec{o}_t^{(k)} - \bar{\vec{\mu}}_{nm})^{\mathrm{T}}}{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \Delta(q_t^{(k)*} - s_n)} \\ &\quad \begin{pmatrix} n = 1, 2, \ldots, N \\ m = 1, 2, \ldots, M \end{pmatrix}, \\ \bar{c}_{nm} &= \frac{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \Delta((q_t^{(k)*} - s_n) + (r_t^{(k)} - x_{nm}))}{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \Delta(q_t^{(k)*} - s_n)}. \end{aligned} \tag{29}$$

(4) Iterate steps (2) and (3) until the distance measurement between the previously estimated CHMM and the newly estimated CHMM becomes less than an error bound.

In this paper, maximum distance clustering method is used in step (1) to select values of mean and covariance to improve convergence speed of segmental k-means clustering.

Re-estimation: With the initial estimation values derived by above method, Baum–Welch algorithm [1,6,7] is used to re-estimate parameters of CHMM.

(1) After calculating the observation vector pdf using Eq. (10), scaled forward and backward variables are calculated from Eqs. (12) and (13) and Eqs. (16) and (17), respectively. The likelihood is calculated from Eq. (14).

(2) Using three posteriori variables from step 1, the parameters of CHMM are re-estimated using the following equations:

$$\bar{\pi}_n = \frac{\sum_{k=1}^{K} \gamma_{1,n}^{(k)}}{K} \quad (n = 1, 2, \ldots, N), \tag{30}$$

$$\bar{a}_{ij} = \frac{\sum_{k=1}^{K} \sum_{t=1}^{T_k - 1} \xi_{t,ij}^{(k)}}{\sum_{k=1}^{K} \sum_{t=1}^{T_k - 1} \gamma_{t,i}^{(k)}} \quad \begin{pmatrix} i = 1, 2, \ldots, N \\ j = 1, 2, \ldots, N \end{pmatrix}, \tag{31}$$

$$\bar{c}_{nm} = \frac{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \gamma_{t,nm}^{(k)}}{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \sum_{\ell=1}^{M} \gamma_{t,n\ell}^{(k)}} \quad \begin{pmatrix} n = 1, 2, \ldots, N \\ m = 1, 2, \ldots, M \end{pmatrix}, \tag{32}$$

$$\bar{\vec{\mu}}_{nm} = \frac{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \gamma_{t,nm}^{(k)} \vec{o}_t^{(k)}}{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \gamma_{t,nm}^{(k)}} \quad \begin{pmatrix} n = 1, 2, \ldots, N \\ m = 1, 2, \ldots, M \end{pmatrix}, \tag{33}$$

$$\bar{U}_{nm} = \frac{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \gamma_{t,nm}^{(k)} (\vec{o}_t^{(k)} - \vec{\mu}_n)^2}{\sum_{k=1}^{K} \sum_{t=1}^{T_k} \gamma_{t,nm}^{(k)}} \quad \begin{pmatrix} n = 1, 2, \ldots, N \\ m = 1, 2, \ldots, M \end{pmatrix}. \tag{34}$$

(3) Iterate above steps until the difference of likelihood of observation vector sequences from re-estimated CHMM and previously estimated CHMM becomes less than an error bound.

3.4. Modified covariance matrix

Unlike speech signals, mechanical failure signals generally have relatively very small dispersions of statistical properties especially when the number of training data set is small. In this case, the determinant of the covariance matrix has very small value and pdf calculation using Eq. (10) often becomes very difficult because of underflow problem. So it is often needed to modify covariance matrix to enhance calculation efficiency at the cost of accuracy. One modified covariance matrix is setting low threshold which replaces each covariance matrix component's value less than the threshold and other method is setting off-diagonal element's value to zero assuming characteristic vector's elements are independent of each other [3,8]. If the degradation of CHMM's accuracy is small, then this modified covariance matrix can be effectively used to get pdf while avoiding numerical problems.

3.5. Application of improved CHMM method to a lathe chatter signal

CHMM method explained above is applied to mechanical signal to demonstrate its accuracy. Acceleration signal at the tool post of a lathe was recorded while the cutting condition went through from normal cutting to chatter as shown in the top graph of Fig. 4. The feature vector to be used as an input to CHMM was derived as follows. Three different cutting conditions were assumed and they are normal, transition, and chatter. Each cutting condition roughly corresponds to the following stage, normal cutting stage (data points from 20,001 to 40,000), transition to chatter stage (from 60,001 to 80,000) and chatter stage (from 98,880 to 118,879). 20,000 data points of each period are blocked into 11 segments of 10,000 data points, with adjacent segments

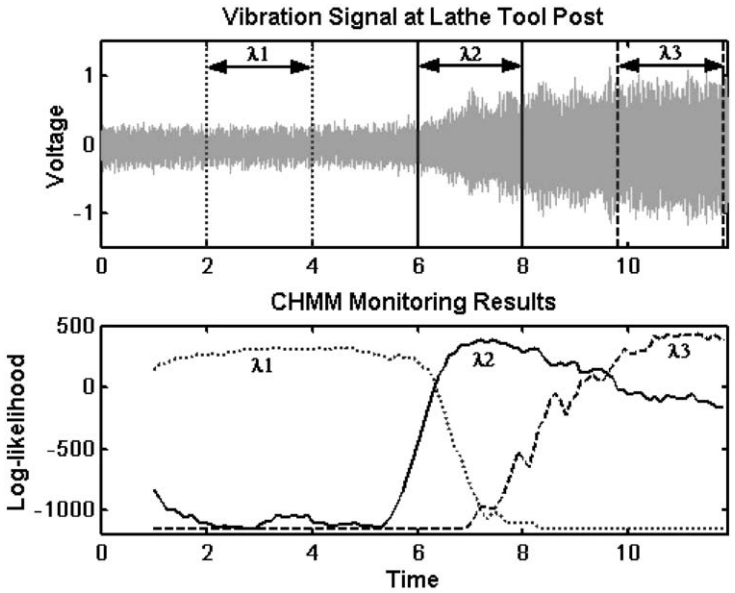

Fig. 4. Diagnosis results of lathe bite using CHMM.

being separated by 1000 data points. 600 data points with 50% overlap over 10,000 points of each segment were used to estimate AR(7) model. These 11 sequences of 32 vectors (7-dimensional) were used as inputs for training three states one mixture CHMM of one category using diagonal covariance matrix. For each category, same procedure was applied and three CHMMs, λ_1–λ_3, were trained.

Three models derived by above training method were used to monitor cutting condition change and the result is shown in Fig. 4. The bottom graph shows the log-likelihood change with time for each model. As expected, the result clearly shows that each model has the highest log-likelihood at each corresponding stage. The result also shows the forecasting ability of CHMM method. For example, at around data point 50,000, the log-likelihood for model 1, λ_1, is still the largest value, however, the log-likelihood of model 2, λ_2, shows smaller value but an increasing trend. So at each overlapping point, the rising and decaying log-likelihoods show not only the current condition estimation but also underlying but significantly increasing condition change thereby proving the powerful forecasting ability of CHMM method. Compared to previous result by authors using old method which needed four models to identify whole signal [6], new improved method only needs three models and it also shows distinctive overlap of conditions for much better prediction.

4. Diagnosis of rotor fault signals

4.1. Experiment using a rotor simulator

A rotor simulator was used to gather rotor fault signals. Fig. 5 shows the experimental setup; a rotor which is driven by a motor and supported by two journal bearings. In addition to normal

operation signal, six other fault signals were measured under artificially controlled fault conditions. They are resonance, stable condition after resonance, bearing housing looseness, misalignment, flexible coupling damage, and unbalance. Each failure condition is explained below.

In resonance experiment, the motor was driven at near resonance frequency. Two disks were used to simulate dynamic loading. Vertical displacement signals at point B near right bearing were measured by a proximeter probe with 1 kHz sampling. Sampled data were converted to averaged auto power spectrum with 50% overlapping and cascade auto power spectrum normalized by the maximum value is shown in Fig. 6. Initially, the rotor was driven at around 1500 r.p.m. and after 2 h, the rotational speed was raised to 1998 r.p.m. wich is close to the resonance frequency at 2040 r.p.m. The system went through pseudo-resonant state and had stabilized after 35 min. This phenomenon was caused by system's characteristic change in which the resonance frequency had been lowered below 1998 rpm. In normal operation, third order ($3\times$) and fifth order ($5\times$) components were higher compared to first order ($1\times$) component and noise level was generally high except at harmonic ($1\times, 2\times, 3\times, 4\times, 5\times, \ldots$) components. But in resonance, both harmonic

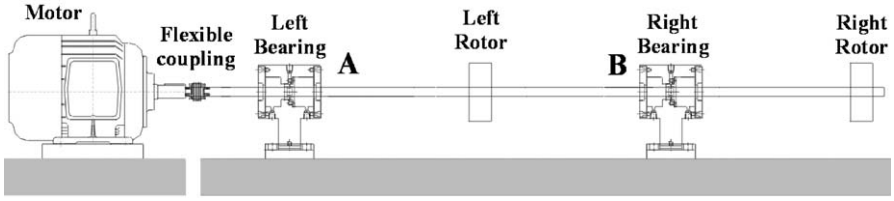

Fig. 5. Simulator setup for resonance and looseness tests.

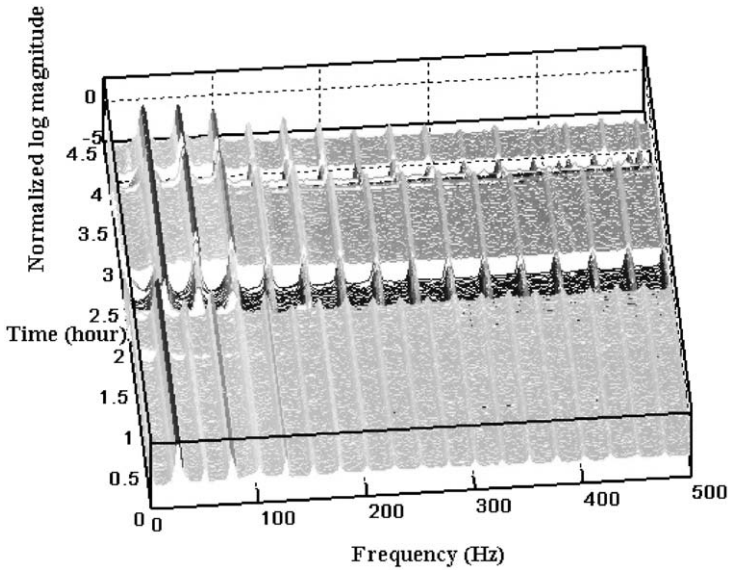

Fig. 6. Cascade plot of resonance test.

components and noise level became very low. When the rotor was stabilized after resonance, second order (2×) and third order (3×) components became bigger and noise level went up, however, fifth order (5×) components had been reduced significantly. So in resonance experiment, three modes were used which are normal, resonance, and stable condition after resonance.

The second experiment is looseness caused by loosening bolts on journal bearing housing. R.p.m. was set to 1800 and same procedure as in resonance experiment was used and the resulting cascade plot is in Fig. 7. The bolt on housing was tightened and system was operated for three and a half hours. Then the bolt on right journal bearing housing was loosened by $45°, 90°, 135°$, and $180°$ for 4 h each. The measured vibration level was not proportional to bolt angle. For $0°$ and $45°$, there are 2× components and level of $3×, 4×$ and noise were relatively low. For $90°, 135°$ and $180°$, levels of 2× and 3× became larger and noise level at low-frequency area became higher. It was decided to use two models, normal and looseness.

The third experiment is misalignment of the shaft. Two disks were installed between two journal bearings which support a motor-driven rotating shaft as in Fig. 8. Vertical displacement signal at point B near right housing was measured. The right housing in Fig. 8 was raised by 4 mm and

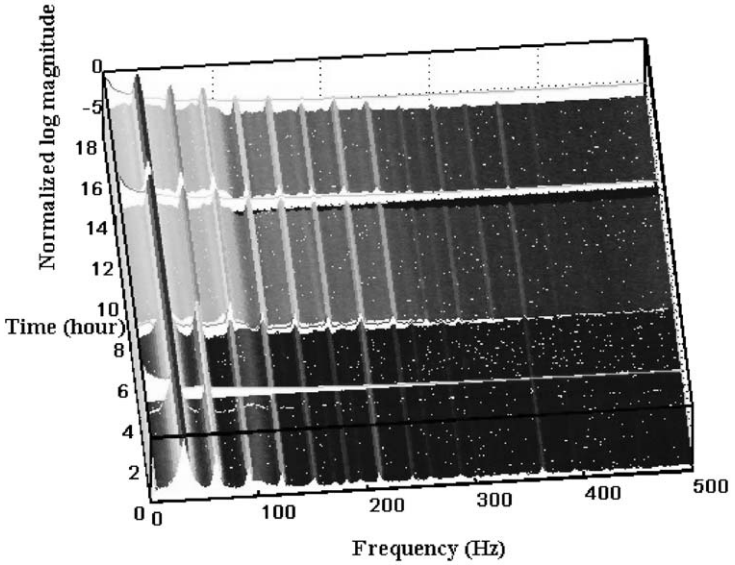

Fig. 7. Cascade plot of looseness test.

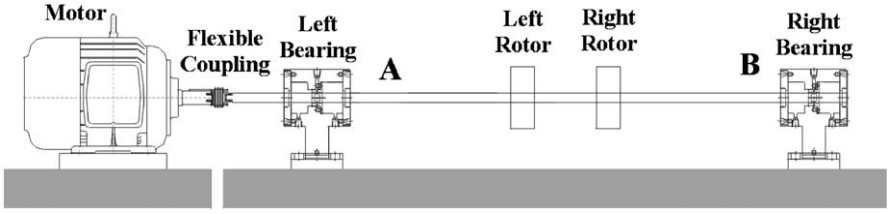

Fig. 8. Simulator setup for misalignment test and unbalance test.

r.p.m. was set to 3500. The result is in Fig. 9. Compared to 1× component, 2×, 3× and 4× components are higher and 3× is the most dominant. The noise level is high. After operation for 18 h, the flexible coupling between motor and shaft was damaged and sub-harmonic components (0.5×, 1.5×, 2.5×, 3.5×, …) emerged and 0.5× and 2.5× components became larger. The coupling was completely damaged after 22 h and 10 min operation. Noise level is similar. Two models, misalignment and coupling failure, were selected.

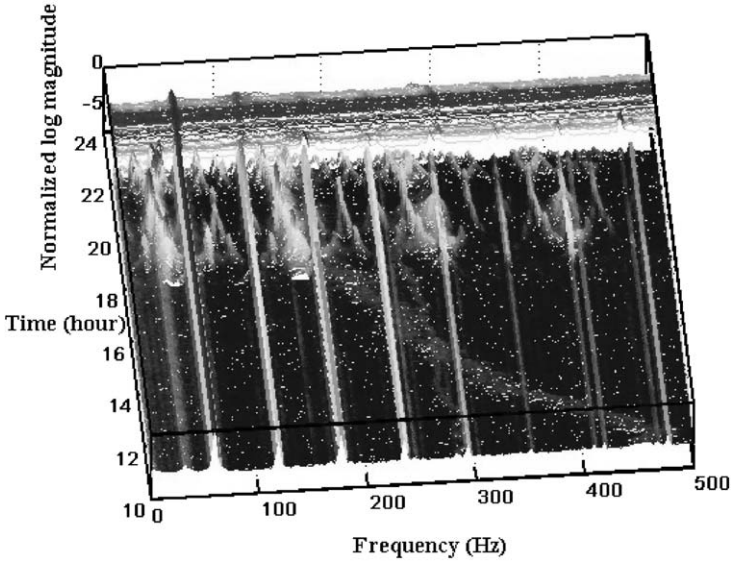

Fig. 9. Cascade plot of misalignment test.

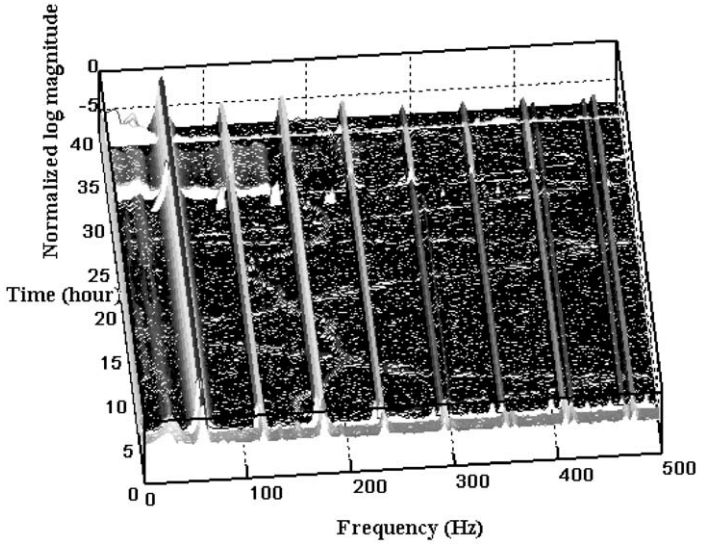

Fig. 10. Cascade plot of unbalance test.

The fourth experiment was for unbalance at 3500 r.p.m. with same experimental setup as in misalignment and the result is in Fig. 10. For the first 1 h, there was no unbalance and compared to 1× component, 3× component was higher and 2× and 5× components were relatively higher. The noise level was also high. Unbalance mass of 10 g was added later to left disk. In this case, 3× component was the highest, however, 2×, 3×, 4× components and noise level were reduced and 5× component was reduced relatively much more. Two models, normal and unbalance, was selected.

4.2. Feature vector extraction and CHMM training

Filter bank was used to extract feature vector from auto power spectrum because use of all spectrum information is computationally inefficient. Selected filter bank was combination of 8 band pass filters. Each band pass filter had band of three lines and 10 orders from 0.5× to 5× by 0.5× increment were selected as center orders. So spectrum was transformed into vector of order 10 by the use of filter bank. All normalized averaged auto power spectrums were transformed into feature vectors of order 10 through filter bank and by grouping consecutive 20 feature vectors, each group became an observation vector sequence $O^{(k)}$ which was used for training and diagnosis for CHMM.

Seven models explained in Section 4.1 were used. The number of observation sequences for each model were 32, 7, 9, 16, 16, 12, and 30 for each and it was used for training CHMM as in Section 2. Diagonal covariance matrix was used for CHMM with one mixture and three states (two for resonance operation).

4.3. Diagnosis results

The total observation vector sequences were diagnosed with seven CHMMs trained as in Section 4.2. The average log-likelihood is shown in Fig. 11. Low threshold value was used to prevent too small values. Each model shows very large likelihood value at its own condition and

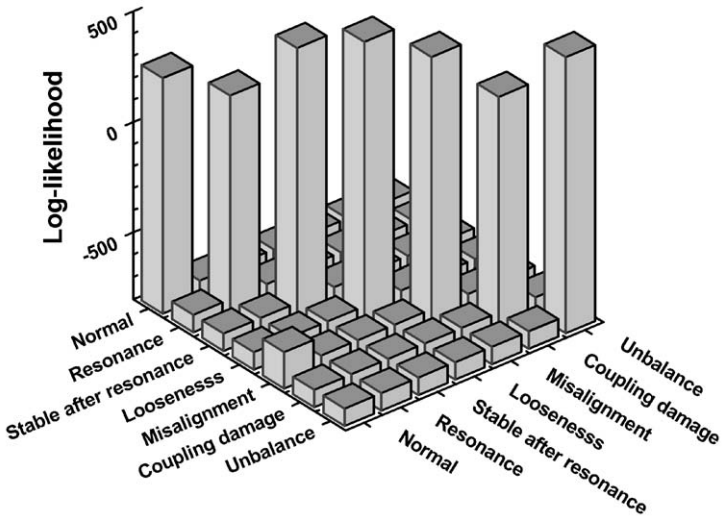

Fig. 11. Diagnosis results (mean of log-likelihood).

very small values at other conditions. Although the result is based on controlled experiment, it is clear CHMM can be a very good tool for mechanical system diagnosis.

5. Conclusion

As a very promising tool for intelligent condition monitoring, continuous HMM (CHMM) was studied and applied to mechanical fault signals. To be effectively used for mechanical signals, which generally have very small dispersion of statistical properties, several modifications have been made to conventional CHMM. It includes initialization using a maximum distance clustering method, use of filter bank, scaled forward/backward variables and diagonal covariance matrix and modification of training equations for multiple observation vector sequences. The improved accuracy of this CHMM algorithm was demonstrated with better result for lathe chatter signal using three models compared to previous result which needed four models. The CHMM's ability of forecasting was also demonstrated.

CHMM was also used for diagnosis of typical rotor failures. Sampled data from simulator were converted into normalized power spectrum and feature vector was estimated using filter bank. Experiment on seven models for four failure conditions has shown very accurate diagnosis result thereby proving CHMM's good potential as a monitoring tool.

References

[1] L.R. Rabiner, A tutorial on hidden Markov models and selected application in speech recognition, *Proceedings of the IEEE* 77 (1989) 257–286.
[2] P. Smyth, Hidden Markov models for fault detection in dynamic systems, *Pattern Recognition* 27 (1994) 149–164.
[3] C. Bunks, D. McCathy, T. Al-Ani, Condition-based maintenance of machines using hidden Markov models, *Mechanical Systems and Signal Processing* 14 (2000) 597–612.
[4] H.M. Ertunc, K.A. Loparo, H. Ocak, Tool wear condition monitoring in drilling operations using hidden Markov models, *International Journal of Machine Tools & Manufacture* 41 (2001) 1348–1363.
[5] K.C. Kwon, J.H. Kim, Accident identification in nuclear power plants using hidden Markov models, *Engineering Applications of Artificial Intelligence* 12 (1999) 491–501.
[6] J.M. Lee, S.-J. Kim, Y. Hwang, Mechanical signal analysis using hidden Markov model, *Ninth International Congress on Sound and Vibration*, Orlando, FL, 2002.
[7] L.R. Rabiner, B.-H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
[8] M.J.G. Gales, Semi-tied covariance matrices for hidden Markov models, *IEEE Transactions on Speech and Audio Processing* 7 (1999) 272–281.