Bearing Performance Degradation Assessment Using Linear Discriminant Analysis and Coupled HMM

T Liu\textsuperscript{1}, J Chen\textsuperscript{1}, X N Zhou\textsuperscript{1} and W B Xiao\textsuperscript{2}

\textsuperscript{1} State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, P R China
\textsuperscript{2} Shanghai Institute of Satellite Engineering, Shanghai 200240, P R China

E-mail address: Liutao_sh@sjtu.edu.cn (T Liu)

Abstract. Bearing is one of the most important units in rotary machinery, its performance may vary significantly under different working stages. Thus it is critical to choose the most effective features for bearing performance degradation prediction. Linear Discriminant Analysis (LDA) is a useful method in finding few feature’s dimensions that best discriminate a set of features extracted from original vibration signals. Another challenge in bearing performance degradation is how to build a model to recognize the different conditions with the data coming from different monitoring channels. In this paper, coupled hidden Markov models (CHMM) is presented to model interacting processes which can overcome the defections of the HMM. Because the input data in CHMM are collected by several sensors, and the interacting information can be fused by coupled modalities, it is more effective than HMM which used only one state chain. The model can be used in estimating the bearing performance degradation states according to several observation data. When becoming degradation pattern recognition, the new observation features should be input into the pre-trained CHMM and calculate the performance index (PI) of the outputs, the changing of PI could be used to describe the different degradation level of the bearings. The results show that PI will decline with the increase of the bearing degradation. Assessment results of the whole life time experimental bearing signals validate the feasibility and effectiveness of this method.

1. Introduction

Equipment maintenance is usually carried out when failures are serious or even have been occurred in rotating machine, which causes high maintenance expenses. Condition based maintenance (CBM) has already becoming the major way in products, however, the CBM is significantly influenced by the performance of equipment components [1]. Bearing is one of the most components in rotating machinery, it performance influences the whole machine’s [2]. Condition based monitoring and diagnosis in bearing performance assessment have received considerable attentions and lots of methods have been applied in this field [3, 4]. The problem in performance assessment is that it is hard to convert the monitoring data into the performance index to guide the maintenance [5]. Therefore, features extraction of monitoring data and building the model to recognize the performance degradation of the bearing is valuable in equipment maintenance.
The monitoring data relevant to system health should be processed in order to obtain the information for better understanding and interpretation of the components’ condition. Large number of techniques have been applied in data feature extraction and analysis, which can be classified into three domains: time domain, frequency domain and time-frequency domain [6]. Besides the traditional features analysis, some new characteristic features such as the spectrum entropy [7], the second generation wavelet packet [5] and lifting wavelet packet decomposition [2] have been introduced into bearing performance assessment.

In machine diagnosis and prognosis models, it is difficult to choose which feature can best interpret the performance because their sensitivity and accuracy are various in different conditions. Furthermore, there are many factors such as locations of the sensors, signal-to-noise ratio of the original data etc. can affect the effectiveness of the features [6]. Thus, in order to ensure the useful information of the acquisition data is sufficient to depict the bearing performance, we have to choose as many features as possible. And it will cause the dimension catastrophe in building the bearing performance degradation model. Therefore, it is crucial to reduce the feature dimensions and extract the most useful information hidden in the vibration signals of bearing [8].

LDA is one of the most well-known linear dimensionality reduction algorithms. The aim of LDA is to project the high dimensional feature set onto the optimal discriminant vector space. The reduction vectors extract the useful classified information and compress the features’ dimension from the given data. This is accomplished by minimizing the intra class distance and maximizing the inter class distance of the original features, that is to say, the new features in the pattern space can be best separated. In fact, LDA can be viewed as a redundant data dimension reduction technique which can compress the multi-dimensional predictors into a one dimensional line [9]. In condition monitoring, the values of bearing features are changing in different status, which the status can be regarded as distinct classes. So the dimension of original feature set from time, frequency, time-frequency domain can be reduced by LDA, and the new projective feature vectors reserve the classified information of the bearing status. The ability and suitability for LDA in extracting useful features as inputs of the bearing performance assessment models will be investigated in this paper.

Another challenge in bearing performance assessment is how to effectively evaluate the performance degradation based on the extracted features [1]. Considerable intelligent algorithms have been proposed on fault diagnosis to realize the prediction and maintenance of machine. Wahyu Caesarendra et. al. [10] discussed the application of the relevance vector machine and logistic regression for machine degradation assessment. Based on fixed cycle features test, Liao [5] introduced the principal component analysis and Gaussian mixture model into the machine performance estimation. Yuna Pan [2] have presented a bearing performance degradation assessment base on fuzzy c-means algorithm. Huang [4] investigate the prediction of bearing’s remaining useful life based on self-organizing map and back propagation neural network methods. Zhao et. al. [11] suggested the application of pattern recognition in machine performance degradation.

Hidden Markov models (HMM) is a time series statistical models with a doubly embedded stochastic process which an underlying stochastic process is not observable but can only be observed through another set of stochastic processes that produce the sequence of observations [12]. As a powerful statistical analysis model, HMM has been widely applied in speech recognition, face recognition and fault diagnosis [13, 14]. In bearing fault diagnosis and performance assessment, the features of vibration signal can be viewed as the observation sequences. On the other hand, the bearing’s condition performance is the hidden process which should be calculated by HMM. Ocak [15] and Jong Min Lee et al[16] applied the HMM in mechanical fault diagnosis. However, many interesting system are composed of multiple interacting processes, for instance, multi-sensors monitoring in rotating machine. With a single state variable, one chain Markov models are unbefitting in these cases. In this paper, coupled hidden Markov models (CHMM) is introduced to capture these interactions [17]. It is more suitable in expressing the multi-channel statistic features.

This paper is organized as follows. In section 2, the proposed method is presented after introducing the LDA, and CHMM. In section 3, a method based on CHMM to assess the bearing’ performance
degradation is carried out. Then an experiment is proposed to validate the feasibility and effectiveness of this method. Our conclusion is in section 4.

2. Technical background

2.1. LDA

LDA creates a linear combination of a number of given independent features relative to the origin data. The goal of LDA is to find a dimension reducing transformation that minimizes the scatter within the class and maximizes the scatter between classes in a reduced dimensional space [18]. The within class scatter can be defined as a matrix $S_w$ and the between classes scatter matrix is $S_B$. The process using LDA to reduce the features’ dimension is as follows:

1. Given a $p$-dimensional feature set $X$ with class information, then the within class scatter matrix $S_w$ can be calculated according to the following:

$$S_w = \sum_{j=1}^{c} \sum_{i=1}^{N_j} (x_i^j - \mu_j)(x_i^j - \mu_j)^T$$  \hspace{1cm} (1)

where $x_i^j$ is the feature samples belong to class $c$, and $N_j$ is the sample points in class $j$. The number of samples categories is $c$, and $\mu_j$ is the mean of samples in class $j$ $(1 \leq j \leq c)$.

2. Calculate the between classes scatter matrix $S_B$

$$S_B = \sum_{j=1}^{c} N_j (\mu_j - \mu)(\mu_j - \mu)^T$$  \hspace{1cm} (2)

where $\mu$ is the mean of the whole samples in all classes.

3. Estimate the projection vector $w$. Because the aim is to maximize the between class measure while minimizing the within class measure. One way is to realize this is to find an optimal vector $w^*$ which can maximize the ratio defined in equation below:

$$w^* = \arg \max_w \frac{w^T S_B w}{w^T S_w w}$$  \hspace{1cm} (3)

Here, Lagrange multiplier method is utilized to estimate the vector $w^*$. So the optimal processing can be converted into solving the eigenvector in equation (4).

$$S_B w_i = \lambda S_w w_i$$  \hspace{1cm} (4)

If $S_w$ is a nonsingular matrix then the eigenvectors of $S_w^{-1} S_B$ is the projection matrix $W$. Furthermore, the maximum eigenvalue corresponding the best projective eigenvector $w^*$. It should be noted that there are at most $c$ non-zero generalized eigenvectors and we require at least $t + c$ samples to guarantee that $S_w$ does not become singular [19].

4. Create the new vector $Y$. It can be accomplished by projecting the original features into the low dimension space.

$$Y = V^* X$$  \hspace{1cm} (5)
where \( V \) is the projective eigenvectors and \( X \) is the raw features, \( Y \) is the new features which the number of dimension is decided by \( V \), ie. the dimensions of \( Y \) is corresponding to the number of eigenvectors \( V \).

### 2.2. Hidden Markov models

HMM is a powerful probabilistic framework for modelling non-stationary process which become popular in various areas in last decades. Generally, the first order Markov chain is utilized in practice and the models is shown in figure 1a. The hidden stochastic state sequence \( S = \{S_1, S_2, ..., S_N\} \) in HMM can be estimated by an observation sequence \( O = \{o_1, o_2, ..., o_N\} \). And their joint probability can be evaluated by

\[
P(O, S | \lambda) = \prod_{t=1}^{T} P(S_t | S_{t-1}) P(O_t | S_t)
\]

where \( \lambda \) expresses the model, \( q_t \) is the state in time \( t \). \( P(q_t | q_{t-1}) \) is the state transition probability and \( P(o_t | q_t) \) is the observation probability. Therefore, a HMM can be indicated by the following parameter set [12]:

\[
\lambda = (A, B, \pi)
\]

In equation (6), \( A = \{a_{ij}\} \) which is the transition probability distribution and is given by

\[
a_{ij} = P(q_{i+1} = S_j | q_i = S_i), 1 \leq i, j \leq N
\]

where the state at time \( t \) are supposed to be independent of the previous, namely,

\[
P(q_t | q_1, q_2, ..., q_{t-1}) = P(q_t | q_{t-1})
\]

Parameter \( B \) is the observation symbol probability distribution in state \( j \). \( B = \{b_j(k)\} \), where

\[
b_j(k) = P(O_k = v_{jk} | q_t = S_j), 1 \leq j \leq N, 1 \leq k \leq M
\]

Here, the observation at any time is assumed to be independent of other states given at that time. And initial state distribution \( \pi = \{\pi_i\} \) is defined as

\[
\pi_i = P(q_1 = S_i), 1 \leq i \leq N
\]

Because the vibration signals is continuous, the model probability density function (pdf) is adopted to ensure that the pdf can be re-estimated in a continuous way[16]. The formulation of the pdf is as follows:

\[
b_j(O) = \sum_{m=1}^{M} c_{jm} N(o_j, \mu_{jm}, U_{jm}), 1 \leq j \leq N
\]

where \( O \) is modeled vector, \( c_{jm} \) is the mixture coefficient of the \( m \)th mixture in state \( j \). \( N \) is a Gaussian distribution with mean vector \( \mu_{jm} \) and covariance matrix \( U_{jm} \) for the \( m \)th mixture component in state \( j \). Generally, a Continuous HMM can be presented as follows:

\[
\lambda = (A, C, \mu, U, \pi)
\]

where \( C = \{c_{jm}\}, \mu = \{\mu_{jm}\}, U = \{U_{jm}\} \).
2.3. Coupled hidden Markov models

A CHMM can be regarded as a collection of several HMM chains which are coupled through introducing conditional probabilities between their hidden states [20]. It should be noted that the observations of chain \( c \) in time \( t \) only rely on the state of the chain \( c \) at the same time, i.e. the probabilities of observations can be expressed as \( P(o^c_t | q^c_t) \). However, the states in CHMM not only dependent on its previous states in the same chain but also reckon on the previous states in the coupled chains, namely, the state transition probability of chain \( c \) in time \( t \) is \( P(q^c_{t+1} | q^c_t, \ldots, q^c_0) \). This property is suitable in applying the CHMM into the multi channels modeling. Figure 1b illustrate a two chains CHMM, where each chain is associated with an observation sequence \( \{o^c_1, o^c_2, \ldots, o^c_T\} \) and the hidden state sequence \( \{q^c_1, q^c_2, \ldots, q^c_T\} \).

**Figure 1.** (a) One chain hidden Markov models and (b) Coupled HMM with two chains which the state sequences coupled through cross time and cross chain conditional probabilities.

Correspondingly, the CHMM can be represented similarly to the HMM with equation (7), but the parameters \( \pi, A, B \) are distinct [20, 21]:

\[
\pi^c = \prod_{c} \pi^c_j = \prod_{c} P(q^c_t = S^c_j)
\]

(14)

\[
a_{i,j}^c = \prod_{c} a_{i,j}^c = \prod_{c} P(q^c_t = S^c_j | q^c_{t-1} = S^c_i)
\]

(15)

\[
b_j(o^c_t) = \prod_{c} b_j^c(o^c_t) = \prod_{c} P(o^c_t | q^c_t = S^c_j)
\]

(16)

where \( c \) indexes the number of the chains, for instance, in figure 1b, \( c \in (1, 2) \). Here, \( q^c_t \) is the state of the two chains in time \( t \) and \( q^c_t = (q^1_t, q^2_t) \) for a two chains CHMM.

Furthermore, the probability distribution of the continuous observations also can be expressed as pdf according to the Gaussian mixed models (GMM).

\[
b_j^c(o^c_t) = \sum_{m=1}^{M^c} w_{j,m}^c N(o^c_t, \mu_{j,m}^c, \Sigma_{j,m}^c)
\]

(17)
where $M_j^c$ is the number of Gaussian mixtures of chain $c$ in state $S_j^c$, $w_j^c$ is weight and $\mu_{j,m}^c$, $\sum_{j,m}^c$ are the mean vector and covariance matrix respectively for the $m$th Gaussian mixture in state $S_j^c$ of chain $c$. The probabilities inference and parameters estimation based on CHMM can be seen in the references [21], in this study, we will pay more attention to the applying of the CHMM on the bearing performance assessment.

2.4. Performance degradation assessment based on LDA and CHMM

Because LDA is effective in reduce the feature dimension of the bearing vibration signal and the CHMM is convenient in modeling the multi-channels signal, the algorithm have been applied in speech recognition and image recognition etc. In this paper, LDA is utilized to reduce the feature set which consisted with RMS, peak to peak, skewness factor, kurtosis factor, peak factor, clearance factor, pulse factor, shape factor, spectral overall value and amplitude spectral entropy. Thenceforth, the new reduction features is input into the CHMM to train the model. Finally, the test data is employed to estimate the performance degree of the bearing. The main steps of this method are as follows:

1. Feature extraction. The ten features of the bearing vibration are calculated and then the dimension of the features is reduced according to the LDA. Here, we select three dimension new features to replace the original feature set. It required that all channels’ original data will be extracted and reduced respectively.

2. Build the CHMM model. The features data in the normal condition is adopted as the train data to calculate the parameters of CHMM, and the object of modeling is to adjust the parameters to maximize the probabilities of the observations $O$ generated by $\lambda$, ie., find the $\lambda^*$ where

$$\lambda^* = \arg \max_{\lambda} P(O|\lambda)$$

This can be solved by expectation maximization (EM) algorithm and the results will be depicted with the maximum likelihood [12, 22].

3. Assess the performance of the bearing. The testing samples are computed in training CHMM database and get the probabilities of the testing samples. And the bearing performance degradation can be presented by the probability.

The process discussed above is shown in figure 2.

![Figure 2. Performance degradation assessment scheme.](image)

In order to weaken the influences about feature sequence and the number of the channels, we will process the log-likelihood of the CHMM, and the new equipment performance index is given by

$$LL = \frac{\log P(O|\lambda)}{TC}$$ (19)
where $T$ is the length of the feature sequence and $C$ is the number of channels. Furthermore, the exponential weighted moving average (EWMA) is introduced to improve the sensitivity and reliability of $LL$ and new index is as follows:

$$PI_t = \alpha LL_t + (1 + \alpha)PI_{t-1}$$

(20)

where $LL_t$ is the index calculated with (19) at time $t$, and $\alpha$ is the smoothing factor which $\alpha \in [0,1]$

### 3. Experimental validation

The effectiveness and facility of the method will be demonstrated by a full life time accelerator experiment. The accelerated bearing life test has been performed to collect the data, which can simulate the whole bearing performance degradation in less time.

#### 3.1. Experiment system

The experiment system consists of the test rig and the data acquisition system. The accelerated bearing life tester (ABLT-1A) is provided by Hangzhou Bearing Test & Research Center, and the locations of the acceleration transducers are shown in figure 3. There are three channels in data acquiring, hereby, it is propitious to model by CHMM. Another important part is the data acquisition system which constituted of three acceleration sensors, a SCXI-1531 signal conditional model and two PCI_6023E data acquisition card, furthermore, the data acquisition soft is programmed by NI LabVIEW. The sketch of the experiment system is illustrated in figure 4. The testing bearing type is 6037 which parameters and operation conditions are depicted in table 1. The vibration signal sampling frequency is 25.6 kHz and the data length is 20,480 points (1 group) per minute, where the vibration data of three channels are synchronous acquisition.

![Figure 3. Testing objects. (a) Locations of the acceleration sensors. (b) Force and sensors installation.](image)

![Figure 4. The sketch of the experiment system](image)
Table 1. Bearing’s parameters and operation condition.

| Type | Ball diameter(d) | Pitch diameter(D) | Number of balls | Contact angle(\(\alpha\)) | Motor speed(rpm) | Load (kN) |
|------|------------------|-------------------|-----------------|---------------------------|-----------------|-----------|
| 6307 | 13.494mm         | 57.5mm            | 8               | 0                         | 3000            | 12.744    |

3.2. Results and analysis

Feature extraction (FE) in fault diagnosis is mainly focused on detecting and identifying various faults. While the performance degradation assessment has some differences which need to estimate the equipment conditions and performance degradation degree. In this study, there are three channels in the test bearing, and we use channel three as example to explain how to calculate the new features by LDA. The full life time data is used to validate the effectiveness of the methods. The data is divided into 2469 groups in sequence (one group is 20,480 points and collected in one minute), and each group is split into 10 segments, so we can separate the time wave into a data matrix as \(2048 \times 10 \times 2469\), that is, every segment data can represent the feature of this group data. Figure 5 display the bearing features mentioned in 2.4. Because the traditional features is not convenient in representing the condition of the running bearing which only sensitive to the bearing’s different states respectively. In this paper, we utilize the LDA to reduce features dimensions. First of all, the features in full life time is selected to estimate the LDA projective space, here, we utilize the normal features (1 to 200 minutes, \(11 \times 10 \times 200\)), degradation features (1501 to 1600 minutes, \(11 \times 10 \times 100\)) and failure features (2351 to 2450 minutes, \(11 \times 10 \times 50\)) to calculate the projection vectors according to the method discussed in 2.1. Secondly, the reduced dimensional features will be obtained by equation (5). The reduction result is shown in figure 6a, and figure 6b is the enlargement of the first dimensional feature after 2200 minutes, it is clearly that the trend and consistency of the new feature is more facilitate than the original features.

Furthermore, the new features has better classification ability, and this is can be d from figure 7a which the condition in original features is hard to identify the normal condition and degradation condition of the bearing. In contrast, figure 7b displays that three conditions can be explicitly divided and each class is better clustering than before. However, figure 7c shows that the new LDA feature three is difficult in classifying the failure condition with other two, so at last, we selected the first two new features ( ) as input to train the CHMM.
After three channels features are reduced according to the method mentioned above, then, the multi-channels features will input into the CHMM to estimate the bearing performance degradation. First, the normal data (1 to 600min) is utilized to calculate the CHMM parameters according the algorithm in 2.3. Then the full life time data will be tested in the model $\lambda$ to estimate the degradation
of the bearing. Furthermore, the bearing performance index is calculated with equation (20). In this study, a three chains CHMM with 3 states \((N = 3)\) and \(M = 2\) mixtures is selected. Compared with the single channel HMM index shown in figure 8, the performance indicator curve of CHMM is shown in figure 9. It is obviously that, the bearing condition in CHMM can be classified more distinctly. Before 1295 minutes, the bearing is in normal condition, and between 1295 to 2304 minutes, the bearing performance is in deteriorating, and after 2304 the bearing fall into failure rapidly.

![Figure 8](image1.png)

**Figure 8.** The performance indicator calculated with single channel CHHM. (a) Performance indicator curve of full life time with channel 1 data, and (b) is the curve getting from channel 2 data.

![Figure 9](image2.png)

**Figure 9.** The performance indicator calculated with three channels CHHM, where (a) is the performance indicator curve of full life time and (b) is the enlargement after 2200minutes.

4. Conclusion

In this paper, CHMM is applied in bearing performance degradation assessment. In this new scheme, LDA is used to reduce the dimensions of vibration features, and the normal data of new features are utilized to estimate the parameters of CHMM. Finally, the full life time data is tested in the CHMM and the performance degradation indicator validate the reasonable and reliable the method. The decline curve of the performance index is consistent to the trend of bearing damage. So the method is intuitive and effective in bearing performance degradation assessment, and feasible in guiding the CBM of rotating equipment.
Acknowledgements
This work is supported by the National Natural Science Foundations of China (Grant Nos. 51035007, 51175329 and 51105243). We appreciate the support of Hangzhou Bearing Test & Research Center with our experiment.

References
[1] Jardine, A.K.S., Lin, D. and Banjevic, D. A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mechanical Systems and Signal Processing, 2006, 20(7), 1483-1510.
[2] Pan, Y., Chen, J. and Li, X. Bearing performance degradation assessment based on lifting wavelet packet decomposition and fuzzy c-means. Mechanical Systems and Signal Processing, 2010, 24(2), 559-566.
[3] Widodo, A. and Yang, B.-S. Application of relevance vector machine and survival probability to machine degradation assessment. Expert Systems with Applications, 2011, 38(3), 2592-2599.
[4] Huang, R., Xi, L., Li, X., Richard Liu, C., Qiu, H. and Lee, J. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. Mechanical Systems and Signal Processing, 2007, 21(1), 193-207.
[5] Liao, L. and Lee, J. A novel method for machine performance degradation assessment based on fixed cycle features test. Journal of Sound and Vibration, 2009, 326(3-5), 894-908.
[6] Yu, J. Bearing performance degradation assessment using locality preserving projections and Gaussian mixture models. Mechanical Systems and Signal Processing, 2011, 25(7), 2573-2588.
[7] Pan, Y., Chen, J. and Li, X. Spectral Entropy-A complementary index for Rolling Element Bearing Performance Degradation Assessment. Journal of Mechanical Engineering System, 2009, 223, 1223-1231.
[8] Sun, W., Chen, J. and Li, J. Decision tree and PCA-based fault diagnosis of rotating machinery. Mechanical Systems and Signal Processing, 2007, 21(3), 1300-1317.
[9] Subasi, A. and Gursoy, M.I. EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Systems with Applications, 2010, 37(12), 8659-8666.
[10] Caesarendra, W., Widodo, A. and Yang, B.S. Application of relevance vector machine and logistic regression for machine degradation assessment. Mechanical Systems and Signal Processing, 2010, 24(4), 1161-1171.
[11] Zhao, D. and Yan, J. Performance prediction methodology based on pattern recognition. Signal Processing, 2011, 91(9), 2194-2203.
[12] Rabiner, L.R. A Tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 1989, 77(2), 257-286.
[13] Flynn, R. and Jones, E. Robust distributed speech recognition in noise and packet loss conditions. Digital Signal Processing, 2010, 20(6), 1559-1571.
[14] Nair, N.U. and Sreenu, T.V. Multi-Pattern Viterbi Algorithm for joint decoding of multiple speech patterns. Signal Processing, 2010, 90(12), 3278-3283.
[15] Ocak, H. and Loparo, K.A. HMM-based fault detection and diagnosis scheme for rolling element bearings. Journal of Vibration and Acoustics, 2005, 127(4), 299-306.
[16] Lee, J.M., Kim, S.J., Hwang, Y. and Song, C.S. Diagnosis of mechanical fault signals using continuous hidden Markov model. Journal of Sound and Vibration, 2004, 276(3-5), 1065-1080.
[17] Brand, M. Coupled hidden Markov models for modeling interacting processes. Learning and Common Sense Technical Report 405 (MIT Media Lab Perceptual Computing, 1996).
[18] Kim, H., Drake, B.L. and Park, H. Multiclass classifiers based on dimension reduction with generalized LDA. Pattern Recognition, 2007, 40(11), 2939-2945.
[19] Martinez, A.M. and Kak, A.C. PCA versus LDA. IEEE Transactions on Pattern Analysis and
Machine Intelligence, 2001, 23(2), 228-233.

[20] Brand, M., Oliver, N. and Pentland, A. Coupled hidden Markov models for complex action recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 994-999 San Juan, PR, USA, 1997.

[21] Xie, L. and Liu, Z.-Q. A coupled HMM approach to video-realistic speech animation. Pattern Recognition, 2007, 40(8), 2325-2340.

[22] Li, H.-Z., Liu, Z.-Q. and Zhu, X.-H. Hidden Markov models with factored Gaussian mixtures densities. Pattern Recognition, 2005, 38(11), 2022-2031.