Mechanical Systems and Signal Processing (2000) **14**(4), 597–612
doi:10.1006/mssp.2000.1309, available online at http://www.idealibrary.com on IDEAL®

CONDITION-BASED MAINTENANCE OF MACHINES USING HIDDEN MARKOV MODELS

Carey Bunks and Dan McCarthy

BBN Corporation, 70 Fawcett St 15/2A, Cambridge, MA 02138, U.S.A.

and

Tarik Al-Ani

ESIEE, Cite Descartes, BP 99, 93162 Noisy le Grand Cedex, France

(Received 11 April 2000, accepted 27 March 2000)

1. INTRODUCTION

The term condition-based maintenance (CBM) is used to signify the monitoring of machines for the purpose of diagnostics and prognostics. Diagnostics are used to determine the current health status of a machine's internal components and prognostics are used to predict their remaining useful life. CBM has the potential to greatly reduce costs by helping to avoid catastrophic failures (an extremely important point, for example, for helicopter gearboxes) and by more efficiently determining the intervals required for maintenance schedules. The economic ramifications of CBM are many fold since they affect labour requirements, replacement part costs, and the logistics of scheduling routine maintenance. The reduction of maintenance to that which is strictly necessary can have the effect of prolonging machine life and that of diminishing the defects that can be introduced by the maintenance itself. Finally, the reduction of catastrophic failures which lead to the loss of life and equipment can have an important effect on insurance premiums.

One method for performing CBM is by using vibration measurements. The main objective is the detection of vibrational characteristics which correspond to physical changes in the machine which indicate abnormal operation. Examples of this could be chipping in a roller bearing or spalling on a pinion gear. The primary challenge is to achieve a high degree of precision in classifying a machine's health given that its vibrational characteristics will vary with many factors not all corresponding to defective components. For example, two identical, new machines will generally have different vibrational characteristics due to differences in the manufacturing process. Furthermore, a machine's underlying vibrational character is likely to change over time as a function of the operating conditions, the maintenance schedules it has undergone, and aging of the machine. Finally, the specific vibrational characteristics of the machine will change as torque loads vary.

Clearly, it is important to be able to differentiate between vibrational changes which are due to machine component defects and those due to changing operating conditions. Assuming that torque loads vary slowly with time, recorded vibration data should demonstrate practically stationary statistics over short intervals. Under these conditions the following approach suggests itself. First, all the different operating conditions which give rise to significantly different statistics in vibration data must be identified. Second, the statistics of vibration data for various defects must be determined either experimentally or by modelling. The combined sets of statistics would then serve as a universe of models with

0888–3270/00/070597 + 16 $35.00/0 © 2000 Academic Press

which hypothesis testing could be performed on segments of the data. Continuous processing of sequential segments of the data would result in classification of operating condition and defect type (if any) as a function of time.

Prior knowledge of how a machine were to be operated or information about the relative frequency of the occurrence of different types of defects could also be used to improve the performance of the above classification algorithm. This would be accomplished by constructing a doubly stochastic process which describes probabilistically how the machine is likely to undergo transition from state to state and what the statistics are at each state. This type of process is well suited to a statistical modelling methodology known in the literature as 'Hidden Markov Model' (HMM) [1, 3, 4].

HMMs are well suited to modelling of quasi-stationary signals and thus, as will be discussed in what follows, can be used to perform detection and estimation for machine diagnostics and prognostics. Two features of HMMs are particularly useful in the monitoring of machine health. The first is that computationally efficient methods exist for computing likelihoods using HMMs.† This feature is important since it promises that signal processing tools based on HMMs can be implemented cost effectively. Furthermore, there exist efficient techniques which can be used for system identification with HMMs. This means that HMMs can be used to build data-driven models of machines relieving somewhat the need to identify specific features in data to be used as health indicators.

The next section of this paper compares problems of speech processing, an area where HMMs have been applied extensively, to those of machine health monitoring. This comparison is useful since it will help motivate the use of HMMs for CBM as well as indicate what some of the critical issues will be. Then, the Westland helicopter gearbox data set will be presented. This data set will be used to illustrate some of the characteristics of vibration data under different operating conditions and types of defects. This data set is then used to illustrate the application of HMMs to CBM. Finally, this paper is terminated by a discussion of the results and our conclusions.

2. COMPARING CBM TO SPEECH PROCESSING

This paper describes efforts to apply HMM technology to the problem of CBM. The motivation for using HMMs comes from the success it has had in the area of speech processing. Today, most commercial speech processing software for speech recognition, speaker identification, and speaker verification are based on HMMs. Furthermore, there is a compelling parallel between speech processing problems and CBM which will be developed in what follows.

The major technical challenges in CBM which must be addressed by any processing methodology are summarized in the following points:

1. Machine vibration statistics are quasi-stationary and vary as a function of operating speed, torque, and, in many cases, ambient or atmospheric conditions. CBM techniques must be able to differentiate between these normal changes and those due to defects.
2. Machine character can be quite variable due to differences in machining, part-size variations, fastener tightness, wear variations, replacement-part variations, and aging. Machine health-monitoring techniques must be robust to these differences.
3. Vibration features which are indicative of machine health can be obscured by vibration from other machine parts, a multiplicity of transmission paths, and ambient noise.

†See [6] for an in-depth discussion of the use of likelihood functions for estimation and detection.

Machine health-monitoring techniques must be robust to multipath and must be effective in low signal-to-noise environments.

The first two points in the above list are comparable to the requirements in speech processing. First, the speech signal is also quasi-stationary. The statistics of the speech signal change as the glottal source and vocal tract vary to produce the different elementary sound components (e.g. phonemes) of speech (see [5]). Second, speech processing algorithms must also be robust to speaker variability. Different speakers have different vocal tracts and perhaps more importantly different styles of speaking.

As for the third point in the above list, the speech signal is usually recorded in a relatively quiet, controlled environment. Nevertheless, the speech signal only remains stationary over intervals of about 10 ms. In comparison, most machine vibration of characteristics remain stationary on time scales which can be many seconds or even minutes. Consequently, although the speech signal has a much better signal-to-noise ratio than would most CBM problems, the time interval over which time averaging can be performed in order to improve the signal-to-noise ratio is many orders of magnitude greater for problems of machine vibration analysis.

From the above discussion, we conclude that the signal processing methodologies used in speech processing (i.e. HMMs) have good potential for successful application to CBM. In speech processing HMMs are successfully used in speech recognition, speaker identification, and speaker verification. Furthermore, the processing is robust to speaker variability and over populations of different speakers. These are just the qualities that are required for a successful and highly accurate CBM system. It is the objective of this paper to show how HMMs can be applied to the problem of machine health monitoring, machine diagnostics, and prognostics.

Although the above point-by-point comparison between speech processing and CBM has many similarities there are also some important differences which will represent the critical issues to be resolved for any successful application of HMM technology to the problem of CBM. For example, in speech processing the number of phonemes is a relatively small finite set. Furthermore, words which are constructed as sequences of phonemes also represent a finite (although large) set. These facts allow speech processors to create libraries of sounds and words which can be used to build HMMs.

By comparison, the collection of machine sounds for a single machine may be very large and, of course, each machine will make different sounds. Thus, it may not be practical to experimentally acquire a single database which will be representative of all possible sounds made by any collection of machines. Nevertheless, unlike human vocal tracts, machine elements can be carefully measured. It has been shown that for some cases machine vibration can be predicted using explicit models which take into account the physical characteristics of machine components (see, for example, [2]). A critical issue is whether modelling of vibration caused by machine defects will be sufficiently accurate for robust CBM applications.

3. THE WESTLAND HELICOPTER GEARBOX DATA

In this section, a laboratory data set is presented which is used for illustrating the application of HMMs to the problem of CBM. The data set in question is made available by the Office of Naval Research for the evaluation of machine health monitoring algorithms. This data set consists of vibration measurements from a set of accelerometers placed on the casing of a Westland helicopter gearbox. Full details of the gearbox components and data acquisition are available in a report which can be found at a Penn State University web site.[†] The gearbox was operated at various torque levels and with various seeded defects in a special test rig. This data set is used in examples of the application of HMMs to CBM illustrated in what follows.

This section discusses some of the salient features of the Westland helicopter gearbox data set. The data set consists of 68 distinct operating conditions constituting nine different torque levels and eight different seeded defects (of which one is actually a no-defect case). For each operating condition time samples at a sampling rate of 100 kHz from eight accelerometers are available for analysis.

Several examples are now presented which bring out important features of the Westland helicopter gearbox data. For each of the ensuing examples a plot is presented which shows power spectral estimates of sensor data using Welch's averaged periodogram method. Each estimate was produced by averaging non-overlapping, Hanning windowed, Fourier transforms of 3030 samples. Since the sampling rate of the data is 100 k samples per second this gives a 33 Hz resolution in the power spectral estimates. Inspection of the spectral estimates indicated that when using the above window length 100k samples (i.e. 33 non-overlapping windows of data) were sufficient to obtain power spectral estimates with good confidence intervals.

In the ensuing plots the legend for each trace indicates the sensor number (S) and defect-type number (D) both of which can be decoded in the key at the bottom of the plot. The torque level (T) and defect (L) are also indicated in the trace legend. The torque level can take the values 27, 40, 45, 50, 60, 70, 75, 80, or 100%.

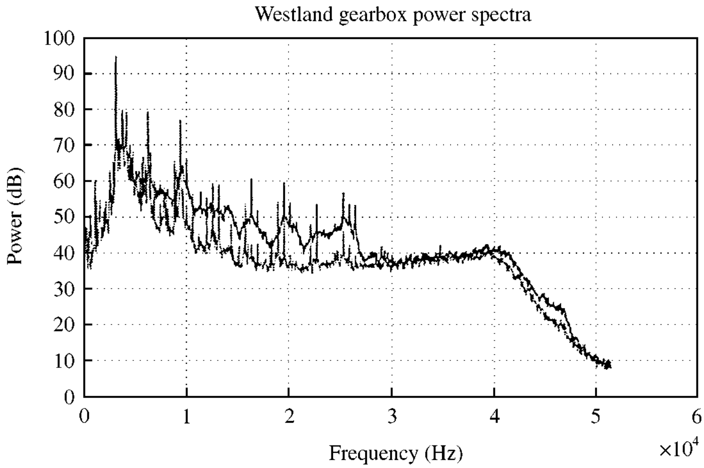
The first example shows that there are significant differences in spectra when using the same sensor but when the gearbox is being operated at different torque levels. Figure 1 shows the estimated spectra for the aft mix box sensor for the no-defect case at 27 and 100% torque levels. Here there is a broadband difference between the two spectra as well as many differences in spectral line locations. An interesting observation is that the overall vibration levels are lower for the 100% torque level (dotted curve) than for the 27% torque level (solid curve). This is true across the board for all sensors and for all defect types. It is presumed that gear pairs were designed to be as close to perfectly involute as possible at the nominal operating torque of 100% and this explains why the gearbox vibration is lowest for this torque level.

An important point brought out by the example in Fig. 1 is that the gearbox data are indeed, like the speech signal, only quasi-stationary. It can be inferred from this figure that different operational conditions related to helicopter maneuvering and perhaps varying wind conditions give rise to significantly different vibration spectra. This example, then, supports the comparison between speech processing and CBM in Section 2.

The next plot illustrates the difference between spectra for different defect types at the same torque level. Here the aft mix box sensor was used to estimate spectra for spalling on the spiral bevel input gear in comparison with the no-defect case. Here it can be seen that the spectra are significantly different but not in an intuitively predictable way (see Fig. 2). The differences in the two spectra are not just differences related to the mesh frequency and its harmonics of the spiral bevel input gear (mesh frequency = 1109 Hz). Rather, they are broad spectral differences, especially in the band from 15 to 40 kHz. Thus, although this figure indicates that CBM should be possible since the features of the two spectra are so different it also indicates that the relationship between the defect and its effects on the spectrum is complex.

Finally, Fig. 3 illustrates the difference in spectra for the same defect type at the same sensor but at different levels of severity. The data in the figure illustrate the spectra from the

[†]http://wisdom.ar1.psu.edu/Westland/report/westcovr.htm.



SENSORS POSITIONS

0 Tachometer
1 Starboard engine input
2 Port engine input
3 Aft mix box
4 Starboard quill shaft
5 Starboard planetary
6 Port planetary
7 Port quill shaft
8 Accessory drive

DEFECT DESCRIPTIONS

1 Pristine no defect
2 Planetary bearing corrosion (DF = 155 Hz)
3 Input pinion bearing corrosion (DF = 784 Hz)
4 Spiral bevel input pinion spalling (DF = 1109 Hz)
5 Helical input pinion chip./freewheel brin. (DF = 9089/1923 Hz)
6 Helical idler gear crack propagation (DF = 9089 Hz)
7 Collector gear crack propagation (DF = 3156 Hz)
8 Quill shaft crack propagation (DF = 43 Hz, 1448 Hz, 1491 Hz)
9 Rebuilt no defect

Figure 1. Different torques produce different vibrational spectra: —— $S=3, D=9, T=27\%, L=1$; - - $S=3, D=9, T=100\%$ $L=1$.

aft mix box sensor for spalling on the spiral bevel input gear operated at 27% torque at two levels of severity. The figure illustrates that there are significant differences in the two spectra in the 15–40 kHz range. The substantively different character of the spectra leads us to conclude that prognostics (i.e. predicting the remaining useful life) should be feasible.

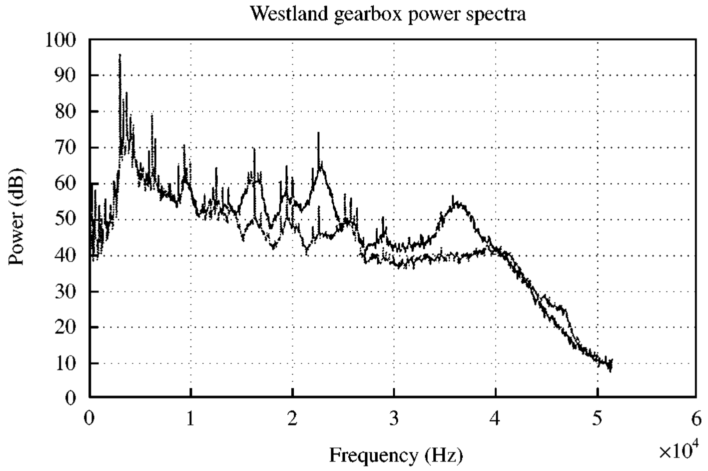
Thus, to summarize the results of this section, the preceding plots illustrate the following points:

- The data are not stationary as a function of operating torque level.
- There are clear, although complex, differences in the spectra as a function of the types of defects.
- There are clear, although complex, differences in the spectra as a function of the severity of defect level.

These observations are important since they validate the comparisons made in Section 2, imply the feasibility of prognostics, and motivate the construction of HMM models for CBM which is the subject of the next section.

4. HMM MODELLING FOR THE WESTLAND HELICOPTER GEARBOX DATA

In this section, the objective is to construct some HMM models which can be used on data from the Westland helicopter gearbox. HMMs are composed of two main



SENSORS POSITIONS	DEFECT DESCRIPTIONS
0 Tachometer	1 Pristine no defect
1 Starboard engine input	2 Planetary bearing corrosion (DF = 155 Hz)
2 Port engine input	3 Input pinion bearing corrosion (DF = 784 Hz)
3 Aft mix box	4 Spiral bevel input pinion spalling (DF = 1109 Hz)
4 Starboard quill shaft	5 Helical input pinion chip./freewheel brin. (DF = 9089/1923 Hz)
5 Starboard planetary	6 Helical idler gear crack propagation (DF = 9089 Hz)
6 Port planetary	7 Collector gear crack propagation (DF = 3156 Hz)
7 Port quill shaft	8 Quill shaft crack propagation (DF = 43 Hz, 1448 Hz, 1491 Hz)
8 Accessory drive	9 Rebuilt no defect

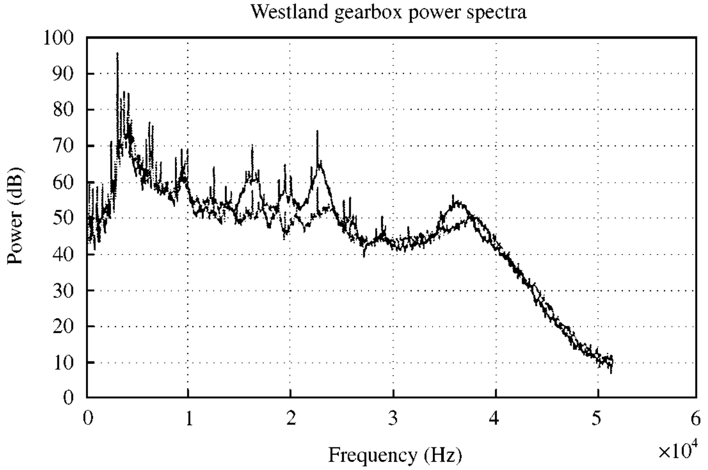
Figure 2. Different defect types produce different vibration spectra: —— $S = 3$, $D = 4$, $T = 27\%$, $L = 1$; -- $S = 3$, $D = 9$, $T = 27\%$ $L = 1$.

components.† There is the state transition matrix which probabilistically describes how the model moves from one stationary type of behaviour to another. There is also the probability distribution associated to each state which describes the statistics of the particular stationary model. In order to effectively model the Westland helicopter gearbox data presented in the previous section it is necessary to construct a reasonable model for both of these components. This is achieved by first specifying the total number of states for the model and then by estimating the parameters of an appropriate probability density for each state. As for the state transition matrix this information can only be obtained by using a prior experimental knowledge of the frequency of occurrence of each defect and the average time spent operating at each of the torque levels. For the purposes of this paper this information is not available; nevertheless, some effort is made in this paper to construct a reasonable state transition model.

4.1. MODELLING STATE PROBABILITY DENSITIES

As noted in Section 3 the characteristics of the power spectra of the Westland data change significantly for different torque levels and for different defect types. For the Westland data set there are a total of 68 different torque-level defect-type pairs available. Thus, the HMM model constructed in this section contains the same number of states.

†For an excellent tutorial on HMMs see [4].



SENSORS POSITIONS
0 Tachometer
1 Starboard engine input
2 Port engine input
3 Aft mix box
4 Starboard quill shaft
5 Starboard planetary
6 Port planetary
7 Port quill shaft
8 Accessory drive

DEFECT DESCRIPTIONS
1 Pristine no defect
2 Planetary bearing corrosion (DF = 155 Hz)
3 Input pinion bearing corrosion (DF = 784 Hz)
4 Spiral bevel input pinion spalling (DF = 1109 Hz)
5 Helical input pinion chip./freewheel brin. (DF = 9089/1923 Hz)
6 Helical idler gear crack propagation (DF = 9089 Hz)
7 Collector gear crack propagation (DF = 3156 Hz)
8 Quill shaft crack propagation (DF = 43 Hz, 1448 Hz, 1491 Hz)
9 Rebuilt no defect

Figure 3. Different levels of defect severity produce different vibration spectra: —— $S = 3$, $D = 4$, $T = 27\%$, $L = 1$; -- $S = 3$, $D = 4$, $T = 27\%$ $L = 2$.

An essential part of applying HMM technology to any problem is to decide what the appropriate observations are. In general, the observations could be the raw data or some function or transformation of the data. A transformation of the data is preferable when the result allows for the diminution of the quantity of data which needs to be processed without losing the ability to effectively monitor the HMM process. This may be important for the Westland data since it is acquired at a rate of 100k samples per second for each of the eight sensors and thus represents a relatively heavy data volume.

Each state of the Markov chain associated to an HMM must have a state process model which implicitly or explicitly specifies a probability density function. In the HMM literature the Gaussian distribution is often used although multi-modal distributions are also used by taking mixtures of Gaussians [1]. Other common choices include mixtures of autoregressive and autoregressive moving-average models [3]. In this paper a simple multi-dimensional Gaussian distribution is used. A different Gaussian is estimated for each of the 68 operating conditions. Each Gaussian is eight-dimensional (8D) (due to the eight sensors) and is estimated using the first $10k$ samples of each of the operating condition runs (i.e. the first 0.1 s of data). The mean vectors, μ_k, and covariance matrices, Λ_k for each of the $k = 1, 2, \ldots, 68$ cases were obtained using the following formulas:

$$\mu_k = \frac{1}{N} \sum_{n=1}^{N} y_k(n) \tag{1}$$

$$\Lambda_k = \frac{1}{N} \sum_{n=1}^{N} [y_k(n) - \mu_k][y_k(n) - \mu_k]^T \tag{2}$$

where $N = 10\,000$ and $y_k(n)$ is the 8-vector of observations at time $n\Delta t$ for data from operating condition k.

We note that, in principle, the vibration data should be zero mean since having a DC level of vibration is meaningless. Nevertheless, it is important to take the sample mean into account to cancel the effects of bias coming from electronic amplification. We further note that modelling the time sample data as an independent and identically distributed 8D Gaussian process does not take advantage of the spectrally banded features of the data that were pointed out in Section 3. The choice does have, however, a physical interpretation. Since the mean does not contain any information the difference in data can only be contained in the second-order statistics, namely the covariance. Looking at the broadband covariance matrix associated to an operating condition yields information about the rms vibrational power on each sensor (the diagonal terms of the covariance matrix) as well as the rms cross-power between sensors. Since the various sensors are distributed equally around the gearbox it can be concluded that this model provides information on proximity (i.e. relative power) and multipath effects (i.e. events that lead to high correlation between sensors). As will be seen in what follows, this model results in an extremely robust classification of the Westland helicopter gearbox data.

To measure the ability of the 68 Gaussian models to discriminate the data they were used in three classification experiments. The experiments were conducted by testing which of the 68 models maximized the likelihood of observed data. This was done with varying sample lengths for the observed data ($10k$, $1k$, and 200 samples). The three tests were performed using the fourth second of data. Note that since the models were created using the first 0.1 s of data that these tests are performed by applying the models to data removed by 3.9 s.

The results of the above three experiments are illustrated in Figs 4–6. Each figure illustrates the classification index (1–68) vs the data set index (also 1–68). Thus, each circle in Fig. 4 yields the result of deciding which is the best model (classification) given the data

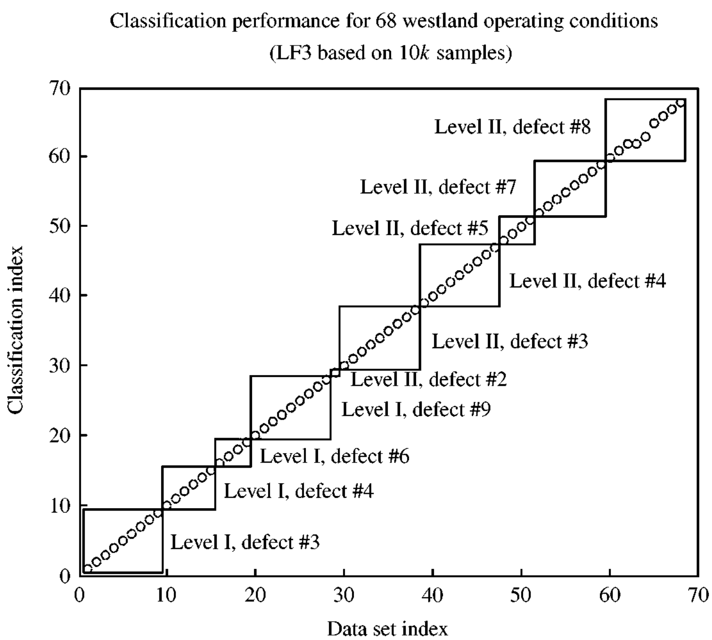


Figure 4. Classification test using fourth second's first 10k samples.

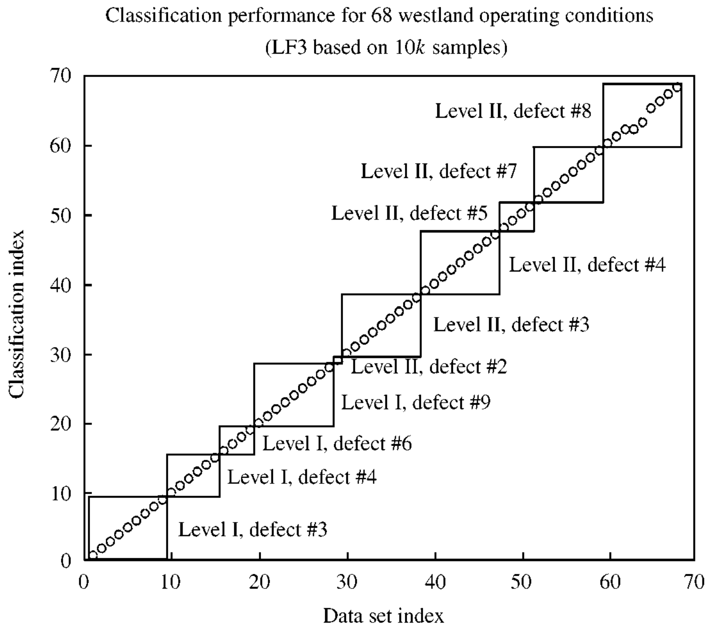


Figure 5. Classification test using fourth second's first 1k samples.

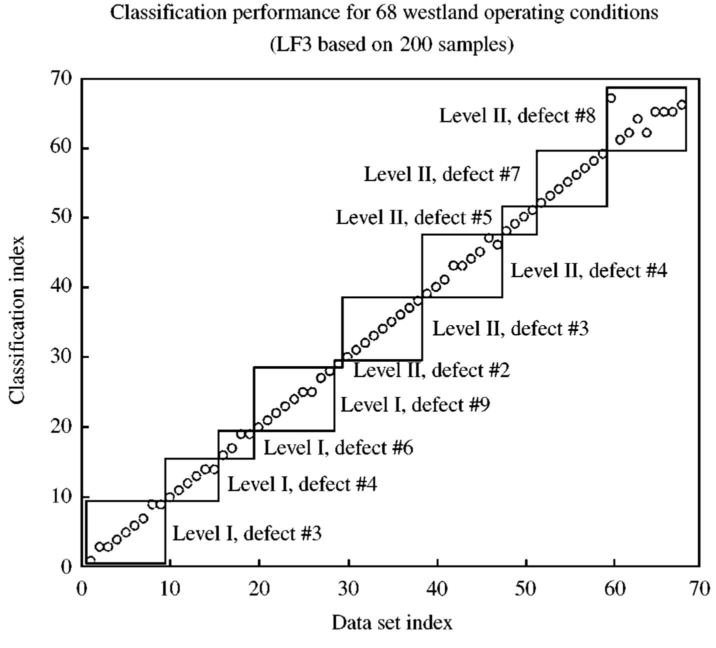


Figure 6. Classification test using fourth second's first 200 samples.

set specified on the independent axis. Perfect classification requires that all the circles lie on the diagonal as they almost do in Fig. 4. The classification indices were chosen to group operating conditions together by defect type. The ensemble of rectangular boxes in Fig. 4 indicates groups of indices whose defect type and level are the same. As shown in the figure

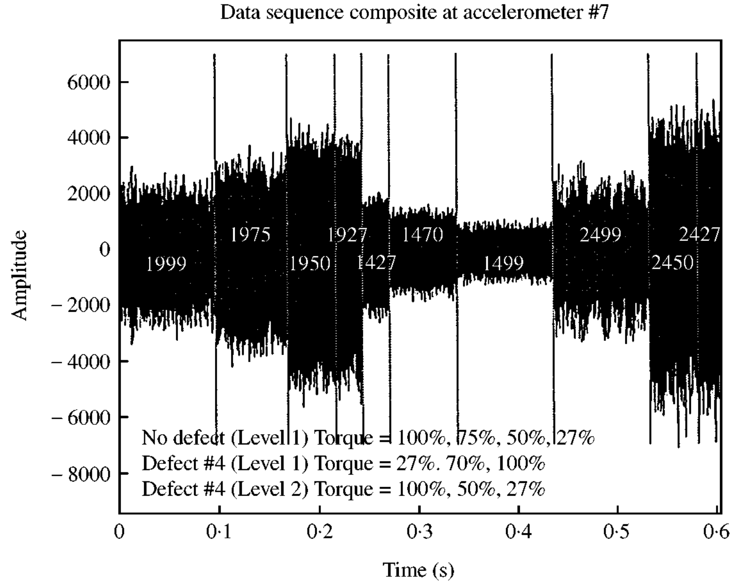


Figure 7. Illustration of sequenced data from port quill shaft sensor.

each rectangle is labelled by the defect type and level it represents. Finally, inside each rectangle the torque level is increasing from the lower left to the upper right of the box.

As can be seen in Fig. 4 the classification of the 68 cases only has minor errors consisting of two torque-level misclassifications. In particular, for the level 2 defect #8 the torque at 50% is misclassified to be 45% and that of 60% is taken to be 50%. Thus, it can be concluded that using eight-dimensional Gaussian distributions to model the data is sufficient to identify defect types and almost sufficient to identify all torque levels. This remains true when the observed data set is reduced to only 1000 samples as is seen in Fig. 5. However, as shown in Fig. 6, when the number of samples is reduced to only 200 the classification of the 68 cases starts to show a more significant number of errors. Nevertheless, these errors continue to consist of only torque-level classification errors and none are defect-type errors.

The next test shows how the above classification procedure might work on a data set from an operating helicopter gearbox. The tests are made using a composite data set which was created by splicing together data from 10 of the 68 different operating conditions. An example of this data set for the port quill shaft sensor is illustrated in Fig. 7. The figure shows the 10 segments delineated by vertical lines. Also shown on each of the segments is a four-digit code of the form XYZZ for which X gives the defect level (1 or 2), Y gives the defect type (1–9), and ZZ gives the torque level.† The sequence of segments is also summarized in the lower-left corner of the plot. The left-most segment of the data set is thus the no-defect case at 100% torque. This is followed by a diminishing of the torque to 75, 50, and then 27% at which point defect #4 (spiral bevel input opinion spalling) is introduced also at a torque level of 27%. The torque level is then increased to 70 and then 100% with this defect. The situation then becomes more serious evolving to a level-2 defect #4 at 100% torque. The torque level then decreases to 50 and then 27% for this defect level.

†The torque level designated by 99 actually indicates a torque of 100%.

It is interesting to note that in the sequence presented in Fig. 7 the amplitude of vibration is higher for low torque levels than for high torque levels. This was already noted previously in the examination of power spectra (see Fig. 1 and the associated comments).

The sequenced data set for all eight sensors (of which only one is illustrated in Fig. 7) were used to perform three classification experiments. Each experiment was conducted by applying a sliding window (with no overlap) of length 1k, 200, and 100 samples, respectively, over the data. The models used in the prior examples were used to classify the data output from the sliding window at each step. The results are shown in Figs 8–10.

Figure 8 is a plot of the classification of the data as a function of time superimposed on the true model plot for a sliding window of 1k samples. The dependent axis is the model index and each segment of the sequenced data is also labelled by the four-digit code as described for Fig. 7.

In Fig. 8 there are very few errors of classification for the case where a sliding window of 1k samples is used. Since model index is organized by torque as was illustrated in Figs 4–6 it can be seen that the several small errors are torque-level classification errors only. For the case of the sliding window using 200 samples shown in Fig. 9 there are more errors but they remain small and are still only slight torque-level discrepancies. Finally, in Fig. 10 which illustrates the case of the sliding window with 100 sample points, there are several important errors which give rise to misclassifications in the type of defect. These errors are indicated by circles and are labelled by their type codes.

The tests presented in the preceding figures indicate that the 68 models are quite robust and are quite capable of performing accurate classification of defects and even classification of torque level. Furthermore, the examples show that the use of more samples is preferable since it improves the accuracy of classification and proportionally diminishes the amount of data to consider. It is important to note that the computational burden necessary to obtain the decision statistics is quite light depending only on a number of multiplies proportional to the number of samples as defined in equation (2).

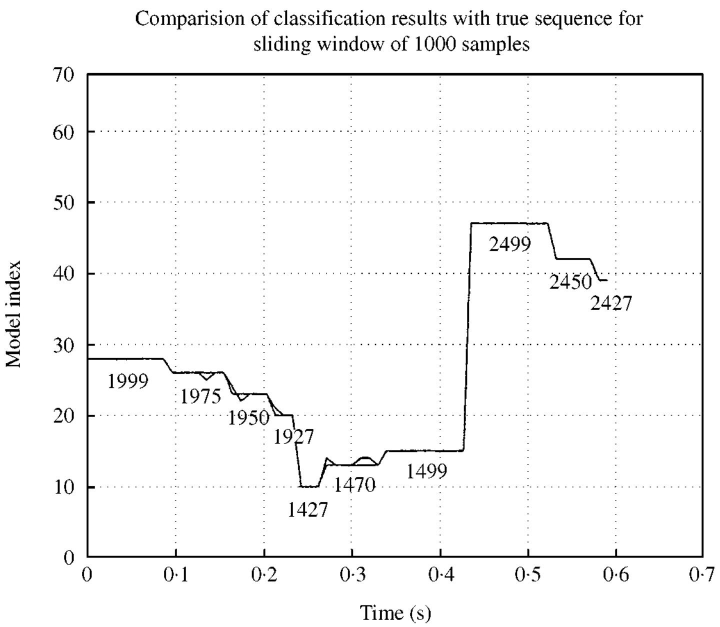


Figure 8. Comparison of classification and true sequence with a sliding window 1000 samples wide.

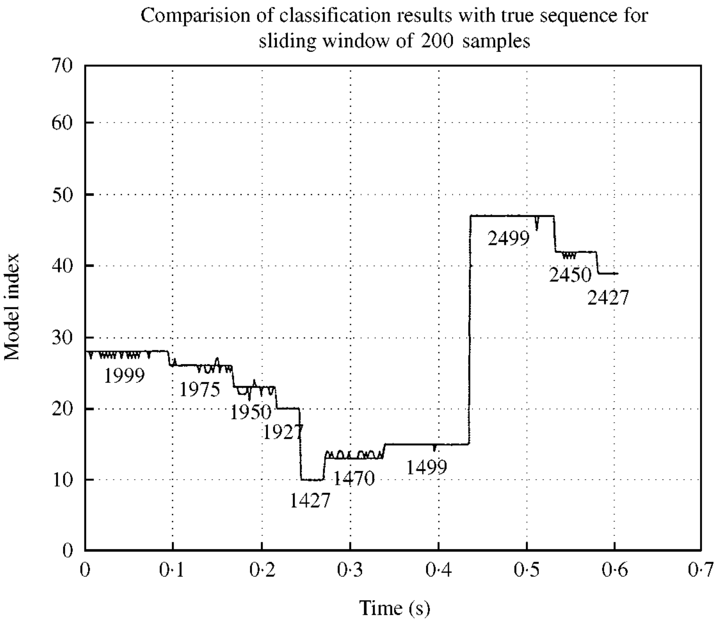


Figure 9. Comparison of classification and true sequence with a sliding window 200 samples wide.

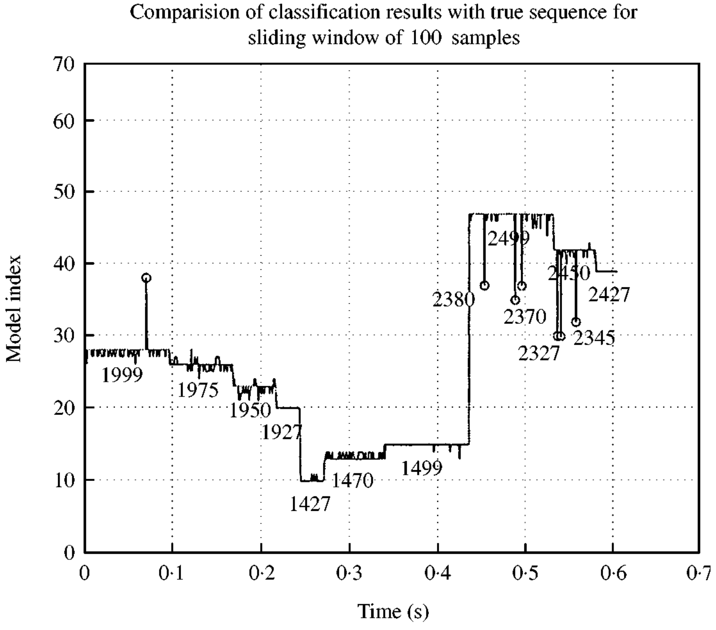


Figure 10. Comparison of classification and true sequence with a sliding window 100 samples wide.

4.2. MODELLING STATE TRANSITION PROBABILITIES

The construction of a state transition model improves classification performance over that obtained from simply using a finite set of unrelated models. That is to say the additional information obtained from knowing what are the relative state visitation frequencies provides a smoothing mechanism to the estimation of state (i.e. classification). Detection and estimation algorithms basedon HMMs incorporate state transition information in a Bayesian way.

In the previous section, a description of how to model state probability densities for the Westland helicopter gearbox data was described and the resulting models were shown to be effective in discriminating between various types of seeded defects. The models were also shown to be robust and sufficiently discriminating to determine the operating torque levels of the gearbox. These results seem to be sufficiently good to forego the need for the construction of a state transition model. However, the Westland data set is a high-quality laboratory test set. Under true helicopter-operating conditions the recorded data would certainly be a lot noisier and would most likely contain features not apparent in the test data.

Another important point is that for general problems of machine health monitoring there will not always be laboratory test data from which state probability densities can be estimated as was done in the previous subsection. Under these circumstances it may be necessary to determine these densities from explicit modelling using only the knowledge of the physical characteristics of the various machine components (for a description of how to do this with gears see reference [2]). Consequently, additional statistical information obtained from the state transition model is very important for the performance of any classification computations.

This section discusses a state transition model for the gearbox data. This model should be based on prior information about the frequency of occurrence of each defect and the average time spent operating at each of the torque levels. Although, for the Westland helicopter gearbox data set, this information is not available it is still useful to discuss a potential model.

The model describes how a simple state transition diagram might reasonably be constructed. The model is illustrated in Fig. 11 which shows three rows of states the top one for no-defect, the middle one for a level 1 defect, and the bottom one for a level 2 defect. Each row contains nine states representing the nine torque levels under which the Westland helicopter gearbox was operated. The arrows between torque levels are associated with probabilities for transitioning from one torque level to the next. As shown in the diagram they indicate that the torque can only increase or decrease by increments. The torque level can also stay the same. The arrows between defect levels are associated with probabilities for transitioning from no defect to a level-1 defect and subsequently to a level-2 defect.

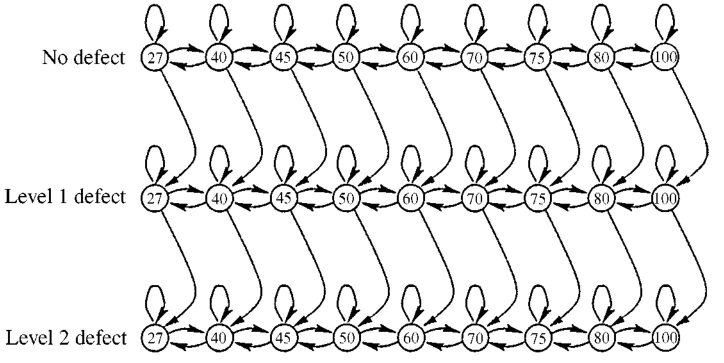


Figure 11. Transition diagram for Westland helicopter gearbox.

Notice that once a no-defect state is left there is no way to return. This is also true for the transition from a level-1 defect to a level-2 defect.

The state diagram in Figure 11 illustrates a model for transitions from a no-defect to a single-defect case. Since the Westland data set consists of seven distinct defect types the full-state transition diagram would be based on Fig. 11 where the bottom two rows would be repeated six more times and each attached in parallel to the top row. The full-state diagram model would then permit a transition from the no-defect case to any one of the seven defect cases.

The state transition model could then be made more complex by allowing multiple defects to be simultaneously present. Under these circumstances, additional branches representing the multiple defect cases would be added to the state transition diagram. The new branches would be appropriately attached to the single defect states. For example, simultaneous defects A and B could be accessed by transitioning from either the single defect A or the single defect B (not necessarily with the same probability).

4.3. PROGNOSTICS

An important problem in CBM is planning essential maintenance. Maintenance must be performed before the expiration of the remaining useful life of critical machine components. Nevertheless, performing machine maintenance before it is needed gives rise to unnecessary costs and can lead to maintenance-caused defects (damage inflicted on the machine via the maintenance procedure). Thus, the estimation of mean useful life, known as prognostics, is an interesting and important problem.

HMMs provide a natural framework within which problems of prognostics can be formulated. To illustrate this, Fig. 12 shows an HMM which can be used for estimating the mean remaining useful life. In the figure, each state represents the degree of incipient failure of a particular machine defect. The random process of each state is determined by a row of Fig. 11. That is, state 1 in Fig. 12 is associated to a random process generated by the top row of Fig. 11 (i.e. for no-defects). State 2 in Fig. 12 gets its random process from the middle row of Fig. 11 and so on. For each state n there is a probability of remaining in that state, a_{nn}, and a probability of leaving it, $a_{n,n+1}$. Since defects always become worse and never improve the chain only allows transitions in the direction of increasing defect.

Thus, if state N represents the state of zero remaining useful life then the mean time, t^*, to this state given that the current state, n, is calculated as a function of the mean number of steps required to go from state n to state N. From the definition of expectation this is accomplished by calculating the probability, p_k, of going from state n to state N in exactly k steps and then forming the sum

$$t^* = \sum_{k=1}^{\infty} k p_k. \tag{3}$$

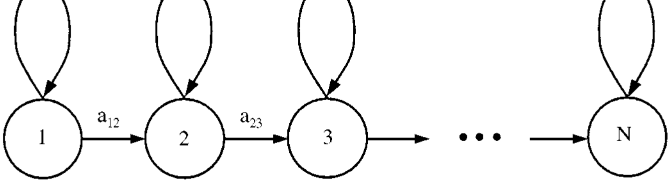


Figure 12. Hidden Markov model for prognostics.

This shows, at least in principle, how the problem of prognostics can be formulated if the appropriate transition probabilities can be determined. Transition probabilities are determined from experiential knowledge. As an example, it is well known that timing belts on cars fail at around 70 000 miles. This sort of information is known for many types of machine components and can be incorporated into the HMM in Fig. 12 as transition probabilities. Used in conjunction with observations of vibration data it helps to determine the current state in the HMM and the transition probabilities can then be used to compute a mean remaining useful life estimate.

This also illustrates how HMMs can be used in a hierarchical way to build up complex models from simpler ones.

5. DISCUSSION

This paper discusses using hidden Markov models (HMMs) for condition-based maintenance (CBM). The paper makes this point by comparing the many similarities of CBM to that of speech processing. Since speech processing enjoys a wide range of successful commercial products based on HMMs, it is concluded that HMMs have a strong potential for constructing practical and robust algorithms for CBM.

The advantages of HMMs are that they are fully probabilistic models which directly incorporate quasi-stationarity (an important issue for CBM) as a feature. HMMs can be used to incorporate all available prior knowledge in a Bayesian formulation and because of their Markovian structure they give rise to computationally efficient signal processing algorithms.

This paper describes the construction of an HMM for the Westland helicopter gearbox data set and presents examples of torque level, defect level, and defect-type classification. The classification is shown to be quite robust which is important since signal-to-noise ratio for practical CBM is often low. It is also shown how HMMs can be used to formulate the problem of prognostics. Most importantly, it is shown how HMMs provide a natural framework within which questions of machine health diagnostics and prognostics can be formulated.

This paper does not answer all questions of how to apply HMMs to CBM and many issues remain to be resolved. In particular, it is important to explore the practical question of how to train HMMs for CBM. Since it is impractical to test each machine for each operating condition that might be encountered, this paper suggests a dynamic modelling approach for synthetically generating vibration data. Much work needs to be done in order to implement such an approach and to demonstrate that the resulting data can be used to accurately perform diagnostics.

Furthermore, the application of transition probabilities between HMM states requires additional data, which are unavailable at this time. It has yet to be demonstrated that a prior knowledge of transition probabilities can be reasonably estimated or that they will have a significant effect on the diagnostics and prognostics abilities of HMMs.

Finally, although there remain many technical challenges we infer that HMMs are extremely flexible modelling tools and thus present a powerful framework within which practical CBM algorithms can be developed. In particular

- Observation sequences for HMMs are completely general and can consist of any combination of data features. This means that any previously proposed defect detection algorithm can easily be integrated into an HMM formalism.
- HMMs can incorporate both input and output observations. In this way, shaft speeds, loads, temperature, etc., can be easily integrated.

- As discussed in Section 4, HMMs can be used in a hierarchical fashion where a state of an HMM can contain another entire HMM. This allows for a lot of flexibility and has been used with great effect in the field of speech processing.

REFERENCES

1. L. A. LIPORACE 1982 *IEEE Transactions on Information Theory* **IT-28**, 729–734. Maximum likelihood estimation for multivariate observations of markov sources.
2. W. D. MARK 1992 In *Noise and Vibration Control Engineering: Principles and Applications*, pp. 735–770. New York: John Wiley & Sons. Elements of gear noise prediction.
3. A. B. PORITZ 1982 *Proceedings of ICASSP, Paris, France, May*, 1291–1294. Linear predictive hidden markov models and the speech signal.
4. L. R. RABINER 1989 *Proceedings of the IEEE* **77**, 257–285. A tutorial on hidden markov models and selected applications in speech recognition.
5. L. R. RABINER and R. W. SCHAFER 1978 *Digital Processing of Speech Signals*. Englewood Cliffs, NJ: Prentice-Hall.
6. H. L. VAN TREES 1968 *Detection, Estimation, and Modulation Theory: Part I*. New York: John Wiley & Sons.