N93-25965

Detection of Bearing Failure in Mechanical Devices Using Neural Networks

Richard A. Burne, Paul F. Payer, R. Paul Gorman, Dan T. Horak
Microelectronics and Technology Center
AlliedSignal Aerospace Company
Columbia, MD 21045

Abstract

We present a novel time-domain method for the detection of faulty bearings that has direct applicability to monitoring the health of the turbopumps on the Space Shuttle Main Engine. A feed-forward neural network was trained to detect modelled roller bearing faults on the basis of the periodicity of impact pulse trains. The network's performance was dependent upon the number of pulses in the network's input window and the signal-to-noise ratio of the input signal. To test the model's validity, we fit the model's parameters to an actual vibration signal generated by a faulty roller element bearing and applied the network trained on this model to detect faults in actual vibration data. When this network was tested on the actual vibration data, it correctly identified the vibration signal as a fault condition 76% of the time.

1.0 Introduction

A critical aspect of the Space Shuttle Main Engine (SSME) as a reusable space vehicle is the durability of its components. One major inadequacy has been the insufficient life of the bearings in the SSME's turbopumps. The life expectancy of the turbopump was designed to be 55 missions, but actually the pumps require an overhaul every one to three missions As a result, a significant ground test program has been required to provide "flight-qualified" turbopumps. One means of reducing the cost associated with ground testing is to provide a preflight, non-invasive monitoring procedure that can detect subtle bearing failures without requiring the firing of the SSME. This paper describes a novel bearing failure detection technique that is suitable for preflight inspection of SSME components.

The most common failure modes of rolling element bearings are local defects in the outer race, the inner race, or a rolling element. As the bearing rotates, whenever the defect passes through the element-to-race contact area, a short duration impact is generated that can be detected by accelerometers or acoustic emission sensors mounted near the bearing. A typical accelerometer signal generated by a faulty bearing is shown in Figure 1. This signal is characterized by transient events caused by bearing imperfections. These transients occur against a background of minute transients whose sum is approximated well by a Gaussian distribution. The fault transients typically exhibit a quasi-periodicity governed by the rotational speed and the bearing geometry[1]. The interval between such transients is typically much longer than the duration of the transient itself. Such impact transients have been recorded from SSME turbopumps using acoustic emission sensors[2]. Because the structure of each fault transient is generally random, the challenge associated with the early detection of bearing faults is to detect the fault transients' periodicity.

PRECEDING PAGE BLANK NOT FILMED

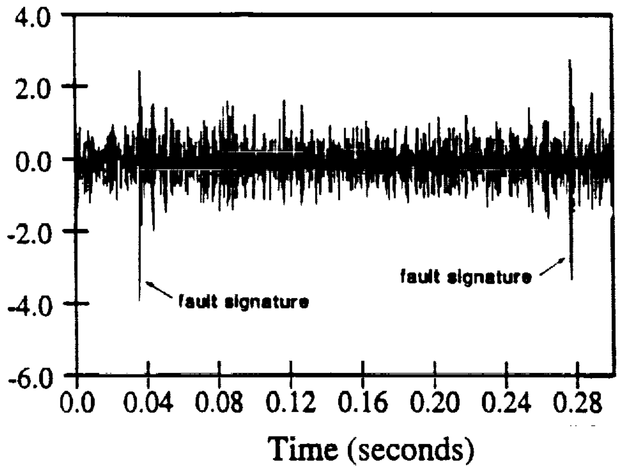

Figure 1 A typical faulty vibration signal. Impact transients are indicated.

In the past, spectral analysis had been used to analyze the acceleration signals from faulty bearings[1]. The basis for this analysis is that the frequency corresponding to the spacing of the fault transients shows up as a peak in the frequency domain. The problem with this method is that the noise in the interval between fault transients tends to dominate the power spectrum due to the temporally local nature of the fault transients. Consequently, only large bearing faults can be detected using this method. This has been shown to be the case for detecting turbopump bearing cage failures. A severely damaged turbopump bearing exhibits peaks in the accelerometer signal's power spectrum at 214 Hz and 428 Hz. These are the primary and secondary harmonics of the bearing cage at 104% of the turbopump's rated power level. Even under severe fault condidions, these spectral peaks cannot be reliably detected.

More recently, time-frequency methods, such as wavelet transforms, have been applied to transient signals in an attempt to address the averaging problems associated with Fourier techniques[3]. Such techniques have proven quite useful in characterizing temporally local events. However, due to the local nature of their basis functions the time-frequency techniques are inappropriate for detecting periodicity in the signal.

As an alternative to spectral and time-frequency techniques, we propose to use time-domain analysis as a means of detecting the inter-pulse interval associated with bearing fault transients.

The simplest time-domain algorithm for detection of pulse trains in the presence of noise is averaging of points in the vibration signal which are one period of the pulse train apart[4]. This averaging enhances the pulses by reducing the effect of random background noise. While this method is effective for enhancing deterministic signals in rotating machinery, such as cylinder pressure in internal combustion engines, its use is limited in the bearing fault detection application. The main problem is that the pulses are random, being a sum of stress waves that reach the sensor through multiple paths between the bearing contact points and the sensor mounting point. The rolling element-to-race impact that produces the pulse is itself random, being dependent on the exact orientation and vibrations of all the bearing components at that instant. Consequently, the pulse spacing is not exactly constant and no two pulses have the same shape. As the result of the randomness, the averaging process attenuates the pulses just as it attenuates the background noise, and does not improve significantly the detectability of the fault.

Another practical limitation of the averaging method is the need for accurate data alignment during the averaging process. In the engine cylinder pressure example, a one-per-revolution signal is sufficient for perfect alignment of signals of any length. In the bearing case, the frequency of the pulses is a product of the rotational speed and a geometric constant. Therefore, a one-per-period signal cannot be generated, making it impossible to align long records accurately.

The contribution of this paper is the use of a feed-forward neural network as an alternative time-domain detection method for pulse trains generated by faulty bearings. The three main features of our method overcome the limitations of the averaging method. First, the fault transients are not required to possess a specfic structure. Second, there is no need for data alignment. Finally, our algorithm can tolerate moderate variations in pulse spacing. In summary, our method can detect pulse trains in noise without excessive sensitivity to the features and repeatability of the pulses.

The next section details the signal model used to train the neural network. Section 3.0 describes the neural network experiments conducted on the modelled data. Section 4.0 provides the results of those experiments and Section 5.0 presents the results of an experiment applying the network detector to actual vibration data generated by a faulty roller bearing. This data was collected from a simple test device. Finally, Section 6.0 discusses the conclusions drawn from these results.

2.0 Vibration Signal Model

To develop the neural network-based fault detector, we modelled the vibration signal generated by a faulty bearing as a pulse train embedded in Gaussian noise. The pulse train possessed a specific periodicity. These pulse train signals generated using various signal-to-noise ratios (SNRs) were used to train the neural network. Once trained, the neural network was tested using actual vibration data collected from a faulty ball bearing in our laboratory.

In order to train a neural network to serve as a generic fault detector for rolling element bearings, a general signal model was developed. Faulty vibration signals are characterized by quasi-periodic impact transients. Figure 1 shows a faulty vibration signal with two impact transients indicated. We were interested in a signal model that provided only quasi-periodicity information as a classification cue. Therefore, we used the same Gaussian statistics to generate both the pulses and the background noise. The only difference between a pulse and noise was the mean amplitude of their respective distributions.

Two classes of vibration signal were generated. The first class of signals possessed pulses whose inter-pulse interval was random (uniform distribution between zero and twice the mean interval). The second class was designed to represent a vibration signal generated by a faulty bearing. This signal possessed a pulse train that exhibited a quasi-periodicity (a Gaussian distribution with a variance equal to 20% of the mean inter-pulse interval). The pulse width to inter-pulse interval ratio was 0.22 and the position of the initial pulse was chosen randomly (a uniform distribution between zero and the mean inter-pulse interval).

The signal-to-noise ratio of a model signal was computed as follows

$$SNR = 10\log\frac{A_S^2}{A_N^2} \qquad (1)$$

where A_S and A_N are the means of the Gaussian distributions used to generate the pulse and the noise, respectively. The final signal was generated by adding the Gaussian noise to the pulse train signal.

3.0 Neural Network Training

Feed-forward neural networks with two layers of modifiable weights were used throughout the study. Two sets of experiments were conducted. In the first experiment, we trained the network on input signals containing an average of ten pulses. In the second experiment, the network was trained on signals containing only three pulses on average. The network trained on ten pulses contained 192 input units, 20 hidden units, and 2 output units. The network trained on three pulses contained 63 input units, 10 hidden units, and 2 output units. The size of the input layer was determined by the number of signal sample points required to provide the appropriate number of pulses to the network. The number of hidden units was chosen to provide a sufficient number of degrees of freedom to solve the classification problem. In both cases, the desired output was [1 -1] for the good bearing and [-1 1] for the faulty bearing.

The network was trained using the back-propagation learning algorithm[5]. The learning rate and the momentum were set at 0.01 and 0.9, respectively. These values provided the best convergence and training rates for the two networks.

During training, network weights were adjusted so as to minimize the difference between desired and actual output values. Each pattern presented to the network was generated at the time of presentation and, therefore, the network never saw the same pattern twice. Training continued until the average improvement in performance over weight updates fell below a fixed threshold. This is termed asymptotic performance.

4.0 Network Performance

For the ten-pulse and the three-pulse networks, training was conducted on signals with a given SNR. Performance figures for each network was obtained for various SNR values. Table 1 provides asymptotic performance levels for the ten-pulse network at the SNRs indicated. Although the SNR for the last two experiments was less than zero, the average amplitude of the pulse was larger than the background noise as a result of adding the pulse vector to the noise vector to produce the final signal.

Ten-Pulse Neural Network	
SNR [dB]	Performance [%]
20.0	87.7
14.0	86.4
6.0	81.9
0.0	80.2
-3.5	79.6
-6.0	54.6

Table 1 Detection performance for ten-pulse network (% correct classification).

As can be seen from the performance values, the network's ability to detect the quasi-periodic pulse train degrades less than 10% as the SNR is decreased from 20.0 dB to -3.5 dB. However, at an SNR of -6.0 dB the performance falls to near chance. This represents a precipitous drop in performance below an SNR of -3.5 dB.

Table 2 presents performance values for the three-pulse network. This network consistently performed 10% to 15% below the ten-pulse network. This indicates that

Three-Pulse Neural Network	
SNR [dB]	Performance [%]
20.0	75.4
14.0	76.5
6.0	74.0
0.0	69.7
-3.5	65.0
-6.0	60.0
-12.0	53.0

Table 2 Detection performance for three pulse network (% correct classification).

the network perfoms better as a function of the number of pulses within its input window, as expected.

5.0 Vibration Data

To test the model against actual vibration signals, we obtained vibration data from a faulty ball bearing. The data was acquired from an accelerometer mounted on the bearing housing which held the outer race. The inner race of the bearing was mounted to a rotating shaft which was driven by an electric motor. The bearing was disassembled and the outer race was damaged with a grinding tool. During data collection, the shaft was rotated at a constant RPM and the vibration signal was digitized and recorded by a personal computer equipped with an A/D converter.

The faulty vibration signal exhibited quasi-periodic pulses similar to the model signal. The frequency of the pulse train was proportional to the RPM of the rotating shaft. We estimated the signal-to-noise ratio of the vibration signal to be 12.5 dB. We computed this value by measuring the mean amplitude of the signal within the inter-pulse interval which we used as the mean of the noise. We then measured the mean amplitude of the pulses and subtracted the mean amplitude of the noise to obtain the mean amplitude of the signal. We then used Equation 1 to compute the signal-to-noise ratio.

A model of the vibration signal was developed by fitting the parameters governing the periodicity of the modelled fault signal to the actual statistics of the vibration signal. In this case the pulse width to inter-pulse interval ratio was 0.054 which was a factor of 4 times smaller than the original model. The variance in the interval between each pair of pulses was a 20% of the mean inter-pulse interval. This value was used previously to generate the modelled signal for the simulation experiments described above.

A three-pulse neural network was trained on model signals as described above. This network achieved an asymptotic performance of 89% correct classification on the modelled data. The network weights were then fixed and tested by presenting the network actual vibration data obtained from the faulty roller element bearing. The network classified the fault signal as a fault 76% of the time.

6.0 Discussion

A feed-forward neural network was trained to detect modelled roller bearing faults on the basis of the quasi-periodicity of impact pulse trains. The network's performance was dependent upon the number of pulses in the network's input window and the signal-to-noise of the input signal. To test the model's validity, we fit the model's parameters to an actual vibration signal generated by a faulty ball bearing. We then applied the network trained on this experimental model to the detection of faults in an actual vibration signal.

The performance of the three-pulse network trained on the modelled signal whose parameters were fit to actual vibration signal statistics performed much better than the three-pulse network trained during the original set of experiments on signals with the same SNR. This is accounted for by the difference in the pulse width to inter-pulse interval ratio between the two cases. In the simulation, model the ratio of the pulse width to the interval between the pulses was four times as large as the same ratio derived from the actual vibration signal. Therefore, the percentage of confusable patterns generated using random inter-pulse intervals was significantly larger for the simulation model.

However, the performance of the same network applied to the actual vibration signal was much closer to the performance of the network trained on the simulation model. This suggests that perhaps the actual variance in the inter-pulse interval exhibited by the actual vibration data should have been measured and used as a model parameter in the experimental model. In any case, the differential in classification performance on the modelled and the actual signal data suggests that a more accurate signal model is required.

It should be pointed out that the performance figures presented in this study were obtained by requiring the neural network to make a decision based on a very small portion of the signal. In the case of the actual vibration signal, the network's decision was based on a signal segment only 3.75 ms in length. We could improve the performance of a neural network-based fault detector significantly by using a time-delay neural network which would allow us to scale the amount of information available to the network a couple orders of magnitude.

The current model completely ignores any characteristic structure of the impact pulses. This was done to ensure that the network detector would be applicable for a variety of bearing faults and systems being monitored under various environmental conditions. However, if the application were restricted sufficiently to allow the use of characteristic impact pulse features, a second neural network could be used to extract such features allowing the detection of faults at much lower SNRs. The capability of neural networks to detect transients in noise was demonstrated in a previous paper[6]. This work showed that a neural network trained to detect a transient with specific structural characteristics consistently out-performed a matched filter designed for the same purpose.

In future work, we plan to apply this technique to the monitoring of bearing failure for the Space Shuttle Main Engine turbopumps. This application would allow us to monitor the system under controlled conditions ensuring that the RPM of the pump was held to a fixed value. However, in some practical applications, where the RPM of the rotating shaft could vary widely, it would be necessary to either restrict the range of RPMs monitored by a neural network fault detector or use a bank of network detectors each tuned to detect faults in a specific RPM range. This is due to the fact that the network cues on periodicity information which must be restricted to a finite range in order to distinguish a periodic pulse train from random pulses.

References

1. Taylor, J. I., "Identification of Bearing Defects by Spectral Analysis," ASME Journal of Mechanical Design,Vol. 102, April 1980, pp. 199-204.

2. Hawman, M. and Galinaitis, W., "Acoustic Emission Monitoring of SSME-ATD Roller Bearings", AIAA 89-2849, 25th Joint Propulsion Conference, Monterey, CA., July 10-12, 1989.

3. Friedlander, B., and Porat, B., "Performance Analysis of Transient Detectors Based on a Class of Linear Data Transforms," IEEE Trans. on Information Theory, Vol. 38:2, pp 665-673, 1992.

4. Braun, S. G., and Seth, B. B., "On the Extraction and Filtering of Signals Acquired from Rotating Machines," Journal of Sound and Vibration, (1979), 65(1), pp. 37-50.

5. Rumelhart, D. E., Hinton, G. E., and Williams, R. J., "Learning Internal Representations by Error Propagation," in Parallel Distributed processing: Explorations in the Microstructure of Cognition, Vol. 1, MIT Press Cambridge, MA, 1986.

6. Burne, R. A., Baindur, S., Gorman, R. P., "Detection of Transients in Gaussian Noise a Using Neural Network," to be submitted to IEEE Trans. Aerospace and Electronic Systems.