

The work described in this document was performed by Transportation Technology Center, Inc.,
a wholly owned subsidiary of the Association of American Railroads.

Ultrasonic Acoustic Health Monitoring of Ball Bearings using Neural Network Classification of Power Spectral Density

William Kirchner, Dr. Steve Southward, and Dr. Mehdi Ahmadian, Virginia Tech
Hayden Newell and Nathan Stoehr, Norfolk Southern Corporation

Summary

Through the Association of American Railroads (AAR) Strategic Research Initiatives (SRI) Technology Scanning Program, researchers at Virginia Tech and the Transportation Technology Center, Inc. have developed an effective method of classifying ball bearing health using an ultrasonic acoustic emissions (UAE) inspection technique in the range of 20 to 150 kHz, which is significantly higher than the generally accepted limit of human audio perception at 20 kHz. This methodology is feasible for evaluating railroad engine and running gear components.

Using a range of lubrication conditions, in the laboratory, the inspection system was able to determine the condition of ball bearings. Lubrication condition, from dry to contaminated to fully lubricated, is an early indicator of potential bearing problems. Thus, the ability to detect these conditions will aid in performing timely, preventive maintenance.

A direct benefit of working in this frequency range is the inherent directionality of the microphones capable of measurement in this range, which becomes particularly useful when operating in environments with low signal-to-noise ratios. By using the UAE power spectrum¹ signature, it is possible to pose the health monitoring problem as a multiclass classification problem and to make use of a multilayer artificial neural network (ANN) to classify the UAE signature.

One major problem limiting the usefulness of ANN's for failure classification is the need for large quantities of training data.² Artificial training data, based on statistical properties of a significantly smaller experimental dataset, is created using the combination of a normal distribution and a coordinate transformation. The artificial training data provides a sufficient sized dataset to train the neural network, as well as overcome the curse of dimensionality. The combination of the artificial training methods and ultrasonic frequency range being used results in an approach generic enough to suggest that this particular method is applicable to a variety of systems and components where persistent UAE exist. An intended use of the technology is locomotive engine diagnostics and repair verification.

The Railway Technologies Laboratory at Virginia Tech is an Affiliated Laboratory under the AAR Technology Scanning SRI Program. The mission of the Technology Scanning Program is to exploit new technologies and provide technical experts for the railroad industry.



INTRODUCTION

There is often a need and a significant amount of cost savings in the realization of the complete life of serviceable components in the rail industry. Traditionally components like bearings are serviced, and often replaced, after a specified service interval. However, there is often an economic advantage to service these components only when they show signs of performance degradation. The motivation behind this work is to develop a generic approach that is capable of assessing component health over a variety of systems and distressed components that emit periodic UAE.

The rail industry has commercially available systems to monitor axle roller bearing health and attempts have been made to successfully monitor other components as well,^{3,4} while this work was developed with traditional ball bearings; an effort has been made to keep the approach as generic as possible. The result is a method that can be applied to any situation where persistent UAE exists.

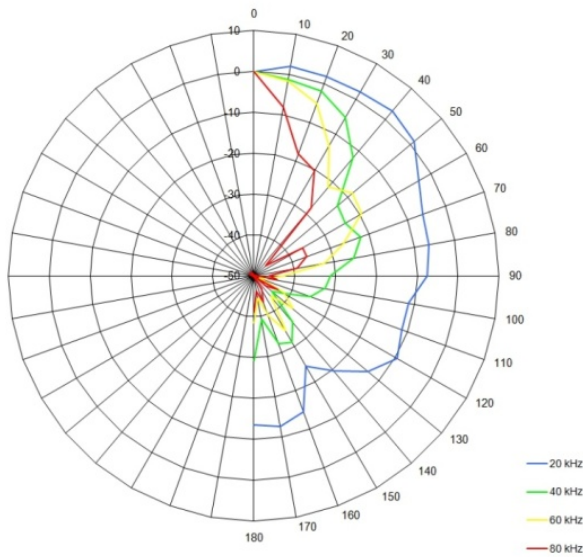


Figure 1. Frequency Response and Polar Diagram of Avisoft Bioacoustic's CM-16 Microphone

All bearings were tested with an arbitrary loading at 1,000 rpm. It is important to remember the intent of this work is to create a generic passive health monitoring approach using UAE. It is, therefore, desirable to have the method be independent of load. A total of 2 seconds of data was collected for each bearing. Three conditions were tested; lubricated (control), no lubrication, and contaminated. The latter two conditions are precursors to bearing failure. Two rounds of testing were conducted with the same batch of bearings (5 lubricated, 5 no lubrication, and 5 contaminated).

RESULTS

Analysis of Power Spectral Density and Isolation of Frequency Ranges

As Figure 2 shows, the power spectrum¹ for each bearing was computed. The results of the power spectrum¹ are used later as the inputs to the neural network, which casts and attempts to solve the multiclass classification problem.

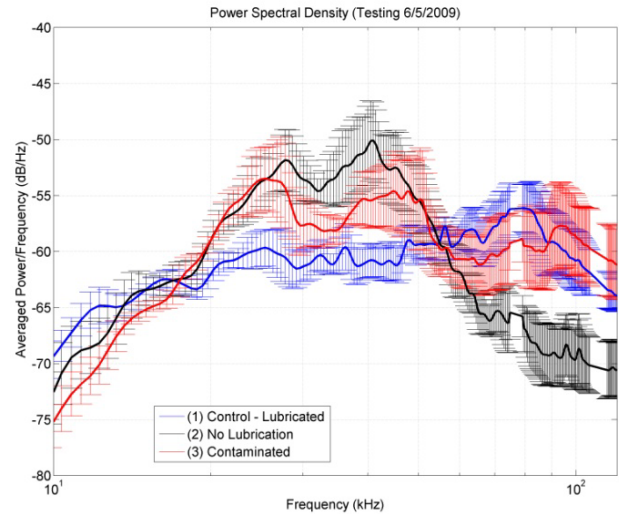


Figure 2. Power Spectrum from Round 1 of Testing

It is advantageous to only use the regions in the power spectrum¹ that aid in the classification between different operating conditions. It is possible to determine regions that are statistically different using a student t-test. Figure 3 shows the results of this hypothesis test. A value of 1 represents the ability to reject the null hypothesis of two samples being random samples from the same dataset. The vertical purple lines represent the limits of the frequency ranges that were chosen to use as the inputs to the neural network. Figure 4 shows the isolated frequency ranges of the power spectrum.¹ It is important to note that each of these regions is averaged, creating only two neural network inputs.

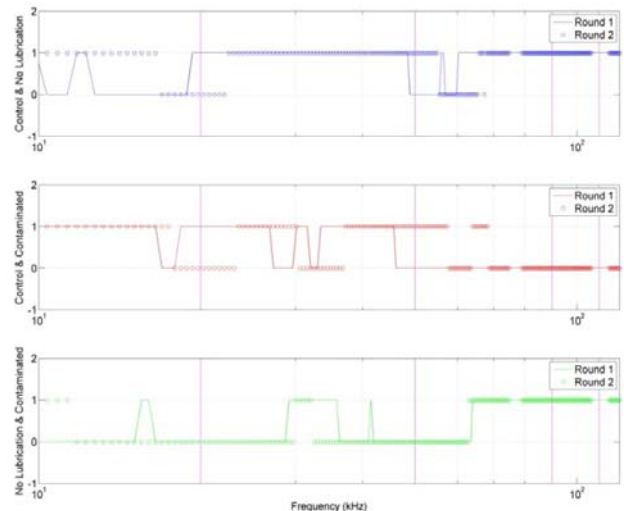


Figure 3. Statistically Different Regions in Power Spectrum

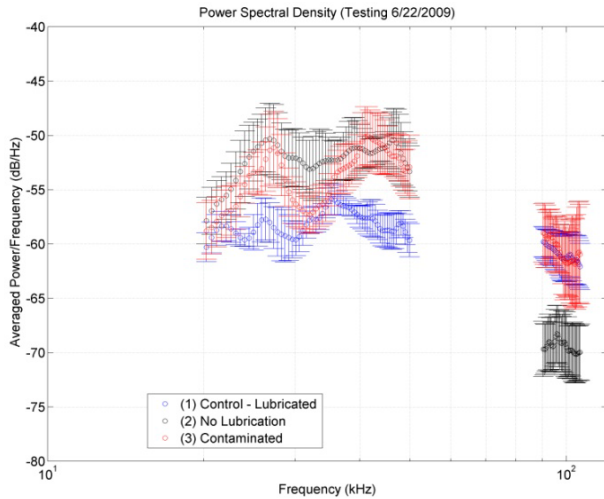


Figure 4. Isolated Frequency Ranges of Power Spectrum

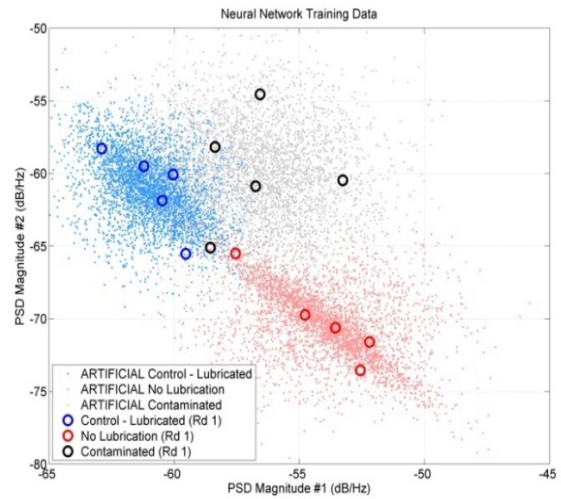


Figure 5. Artificial Training Data Created with Experimental Data from Round 1

Artificial Training Data Generation

In general, a random sampling of the training data is used to divide the training data into three groups. The first group (60% of data) is used for the neural network training, the second group (20% of data) is used to cross-validate the results of the training, and the third group (20% of the data) is used to test the neural network or generalize the error of the network.

It is generally accepted that increasing the number of Epochs or iterations of network training does not always result in a better performing network.² The same generalization can also be made about model complexity and validation error, which can be attributed to the over-fitting problem in the design of neural networks.²

However, it is advantageous to increase the size of the training dataset.² This provides sufficient motivation for creating artificial training data to address the bearing operating condition classification problem. It is also possible to create the artificial training data using a normal distribution with and without a linear transformation. The linear transformation is used to align one of the coordinate axes used in calculating the standard deviation with the result of a linear regression analysis performed on the dataset.

This was done as a result of the observed linear trends in the experimental data. If the trend was something other than linear, a different coordinate transformation would have been used. A combination of the two methods (50% normal, 50% linear transformation) was used to create the training data seen in Figure 5.

Neural Network Classification of Bearing Health

Table 1 and Table 2 list the correct classification rates and the misclassification rates for each class, respectively. The classification rate refers to the number of bearings in class *i* that were successfully classified as being in class *i*. The misclassification rate, in this case, refers to the number of bearings that were classified as class *i* but in fact do not belong to class *i*. Figure 6 shows a two-dimensional visualization of the results.

Table 1. Successful Classification Rate

		Classification Rate	
Training data	Class	Training	Application
Experimental Data (Round 1 Only)	Lubricated	5/5	2/5
	No Lubrication	5/5	5/5
	Contaminated	4/5	4/5
Experimental Data (Round 1&2)	Lubricated	10/10	
	No Lubrication	9/10	
	Contaminated	8/10	

Table 2. Misclassification Rate

		Misclassification Rate	
Training data	Class	Training	Application
Experimental Data (Round 1 Only)	Lubricated	1/6	0/2
	No Lubrication	0/5	1/6
	Contaminated	0/4	3/7
Experimental Data (Round 1&2)	Lubricated	2/12	
	No Lubrication	0/10	
	Contaminated	0/8	

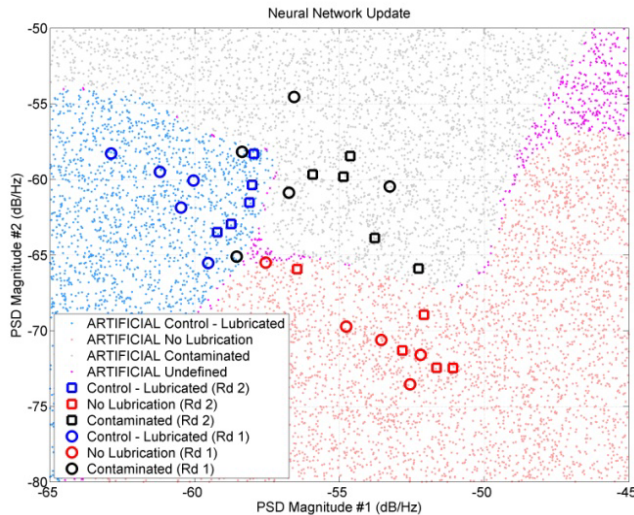


Figure 6. Classification Boundaries; Application After Round 1 of Testing and Update Using Both Rounds of Testing

Both Round 1 and 2 of testing were used to train the neural network results shown in Figure 6, with the experimental points shown for comparison. When analyzing the figures, it is important to realize the small dots are presented purely as a means to visualize the classification boundaries. The actual experimental data is represented as squares or circles, corresponding to each round of testing. A correct classification is made if an experimental data point (square or circle) lies in a region of small dots of similar color.

CONCLUDING REMARKS

It is possible to preprocess the UAE by means of the power spectrum. Provided the power spectrum shows differences between different operating conditions, it is possible to train a neural network to classify the components using artificial training data that is created using the statistical properties of a much smaller experimental dataset.

A key limitation of this method of health monitoring is the initial investment in gathering a statistically significant amount of experimental data that can be used to train the neural network. In general, the idea of using a passive and noncontact measurement — in this case the UAE of the test specimen (bearings) — becomes very attractive due to its nonintrusive nature. A smaller initial investment to get the statistically significant experimental dataset is also advantageous and can be accomplished by the aforementioned method of generating artificial training data.

FUTURE WORK

We have demonstrated the ability of the UAE technique to discriminate bearing lubrication condition (one precursor to bearing failure) in the laboratory. Additional work is needed to expand the range of potential engine component defects that can be found. Additional effort will also be needed to extend the technique to other rotating or noise generating vehicle components. If successful, a prototype will be developed for evaluation in the locomotive shop environment.

ACKNOWLEDGEMENTS

The Association of American Railroads, Transportation Technology Center, Inc. funded and supported this project under the Technology Scanning Strategic Research Initiative.

REFERENCES

1. Welch, Peter D. June 1967. "The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time-Averaging Over Short, Modified Periodograms." *IEEE Transactions on Audio and Electroacoustics* AU-15 No. 2: 70-73.
2. Duda, Richard, Peter Hart, and David G. Stork. 2001. *Pattern Classification*. New York : John Wiley & Sons Inc.
3. Xu, Peng and Andrew K. Chan. 2002. "Fast and Robust Neural Network Based Wheel Bearing Fault Detection with Optimal Wavelet Features." *IEEE Proceedings of the 2002 International Joint Conference on Neural Networks*, pp. 2076-80.
4. Banan, M.R., J. Ghaboussi, and R.L. Florom. 1994. "Neural Networks in Railway Engineering: Acoustic Wayside Fault Detection." *Intelligent Engineering Systems Through Artificial Neural Networks 4*, pp. 1147-1153.