

Ultrasonic acoustic health monitoring of ball bearings using neural network pattern classification of power spectral density

William Kirchner

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State
University in partial fulfillment of the requirements for the degree of

Master of Science
In
Mechanical Engineering

Dr. Steve Southward
Dr. Mehdi Ahmadian
Dr. Dan Inman

November 16, 2009
Danville, Virginia

Keywords: Ultrasonic, Acoustics, Emissions, Health Monitoring, Artificial Neural
Networks, Bearings, Artificial Training Data

Copyright 2009

Ultrasonic acoustic health monitoring of ball bearings using neural network pattern classification of power spectral density

William Kirchner

ABSTRACT

This thesis presents a generic passive non-contact based acoustic health monitoring approach using ultrasonic acoustic emissions (UAE) to facilitate classification of bearing health via neural networks. This generic approach is applied to classifying the operating condition of conventional ball bearings. The acoustic emission signals used in this study are in the ultrasonic range (20-120 kHz), which is significantly higher than the majority of the research in this area thus far. 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. Using the UAE power spectrum signature, it is possible to pose the health monitoring problem as a multi-class classification problem, and make use of a multi-layer 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. Artificial training data, based on statistical properties of a significantly smaller experimental data set is created using the combination of a normal distribution and a coordinate transformation. The artificial training data provides a sufficient sized data set 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.

Table of Contents

INTRODUCTION	1
CHAPTER 1 – ULTRASONIC ACOUSTIC EMISSIONS (UAE).....	2
Ultrasonic Microphone	3
Data Acquisition System	5
CHAPTER 2 - LABORATORY AND EXPERIMENTAL DATA	5
Laboratory Bearing Test Stand.....	5
Bearings and Preparation.....	7
Test Conditions	9
CHAPTER 3 - POWER SPECTRAL DENSITY AND SIGNAL PROCESSING.....	11
High Pass Filtering of UAE.....	11
Power Spectral Density of UAE.....	11
Results of Pre-Processing	12
Isolation and Averaging of Frequency Bands	15
Determination of Frequency Ranges	16
CHAPTER 4 - NEURAL NETWORK DESIGN	17
Multi-Class Pattern Recognition Problem.....	17
Motivation for Using Artificial Training Data	18
Artificial Training Data Sets.....	20
Artificial Training Data Sets using Linear Transformation	21
Neural Network Specifications.....	25
Neural Network Training Algorithm.....	29
Fuzzy Logic Classification Boundaries.....	31
Results of Neural Network	31
CHAPTER 5 - NORFOLK SOUTHERN TESTING	34
Ultrasonic Acoustic Emissions (UAE).....	35
Future Field Testing and Development	38
CONCLUSIONS.....	39
REFERENCES.....	41

List of Figures

Figure 1. Avisoft Bioacoustics CM-16 (© Avisoft Bioacoustics).....	4
Figure 2. Frequency Response of Avisoft Bioacoustics CM-16 (© Avisoft Bioacoustics).....	4
Figure 3. Polar Diagram for Avisoft Bioacoustics CM-16 (© Avisoft Bioacoustics)	4
Figure 4. Top View of Bearing Test Rig	6
Figure 5. Front View of Bearing Test Rig	6
Figure 6. Kollmorgen Drive Motor.....	7
Figure 7. Ball bearing and steel shavings used as contaminant.....	8

Figure 8. Test Fixture Diagram.....	10
Figure 9. Typical Time Series Data	10
Figure 10. Acoustic Signatures of Bearings (Round 1).....	13
Figure 11. Acoustic Signatures of Bearings (Round 2).....	13
Figure 12. Average of Acoustic Signatures with Error Bars Corresponding to 1 Standard Deviation (Round 1)	14
Figure 13. Average of Acoustic Signatures with Error Bars Corresponding to 1 Standard Deviation (Round 2)	14
Figure 14. Isolated Frequency Ranges from Rd. 1 of Testing.....	15
Figure 15. Isolated Frequency Ranges from Rd. 2 of Testing.....	16
Figure 16. Regions with Statistically Significant Differences in Power Spectrum.....	17
Figure 17. Network Error as a Function of Training Epochs	19
Figure 18. Network Error as a Function of Model Complexity	19
Figure 19. Network Error as a Function of Training Dataset Size	20
Figure 20. Artificial Training Data using Normal Distribution (Round 1 Experimental Data)	21
Figure 21. Artificial Training Data Created using Linear Regression Analysis and Linear Transformation (Round 1 Experimental Data).....	23
Figure 22. Combination of Methods to Create Artificial Training Data: Normal Distribution, and Normal Distribution of Data with Coordinate Transformation ((Round 1 Experimental Data).....	24
Figure 23. Artificial Training Data Created using Both Rounds of Testing	25
Figure 24. Neuron Model with R inputs	26
Figure 25. Tan-Sigmoid (Hyperbolic Tangent) Activation Function.....	26
Figure 26. Multi-layer Perceptron (Neural Network).....	27
Figure 27. Neural Network with no Hidden Layer: Updated using Both Rounds of Testing in Training.....	28
Figure 28. Neural Network with 8 Nodes in the Hidden Layer: Updated using Both Rounds of Testing in Training.	28
Figure 29. Neural Network Training Method Validation of Classification Boundaries in 2-D	33
Figure 30. Neural Network Application to Rd 2 Data with Visualization of Classification Boundaries in 2-D.....	33
Figure 31. Neural Network Updated using All Data (Rd 1 and Rd 2) in Training with Visualization of Classification Boundaries in 2-D	34
Figure 32. Representative Testing at Norfolk Southern (September 3, 2009)	35
Figure 33. UAE of NS-2547 at Notch 2	36
Figure 34. UAE of NS-6780 at Notch 2	37
Figure 35. UAE of NS-9769 at Notch 2	37
Figure 36. UAE of NS-2521 at Notch 2	38

Figure 37. BeenaVision’s Automatic Car Safety Appliance Inspection System39

List of Tables

Table 1. Bearing Specifications	8
Table 2. Summary of Bearing Conditions	9
Table 3. Summary of Test Conditions	10
Table 4. Classification Results (Training and Application)	32
Table 5. Mis-Classification Results (Training and Application).....	32

Introduction

In industry there is often a need and a significant amount of cost savings in the realization of the complete life of serviceable components. 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. Health monitoring and condition based monitoring are well studied. This paper attempts to leverage UAE as a means of health monitoring. The motivation behind this work is to develop a generic approach that is capable of assessing component health over a variety of systems and components that emit periodic UAE. The approach in this paper was developed with traditional ball bearings, but effort has been made to keep the approach as generic as possible.

In recent years, there has been an increased interest in condition based monitoring and prognostics. The aircraft industry especially has seen an increase in different types of active and passive non-destructive testing [1] as well as in an increase in real-time diagnostics [2]. The majority of research has focused on either structural health monitoring or engine health management [1][2][3]. Furthermore, additional work has been done to couple both passive contact based vibration based health monitoring with passive non-contact based techniques for improved results with what has been called vibro-acoustic health monitoring [1]. What most of these methods have in common, is that they are especially problem specific, and often there is much left to be desired with respect to generality. There has been some work that has focused on passive, non-contact, acoustic health monitoring with varying degrees of success [4] [5][6]. It is also often the case that the success of acoustic health monitoring is often hampered by low signal-to-noise ratios [6].

ANN's have become increasingly popular in recent years in the areas of pattern classification, machine intelligence and medical diagnostics. In fact neural networks have shown great success in multi-class pattern classification problems [7] as well as medical imaging technology and analysis [8]. It is also important to note the studies that compared traditional statistical approaches (maximum likelihood classifier, and discriminant analysis) in medical imaging to neural network applications, the neural network approach in general showed similar or better performance [8][9][10][11]. The use of these techniques, however, results in a need for an extensive set of training data. It becomes even more desirable to use a neural network in pattern classification applications when its outputs provide a measure of the probability of the occurrence of each class [12]. It is also important to note that there has been a lot of work improving the performance of neural networks, and more specifically overcoming the curse of dimensionality, by optimally setting the initial network parameters [13], as opposed to random assignment which is the more commonly used method.

In this thesis we will attempt to pose the analysis of bearing health as a pattern classification problem that is solvable using a multi-layer ANN. The power spectrum is used to pre-process the UAE for use as the input to the network. A method of artificially creating training data for the network using the underlying statistics of a significantly smaller experimental data set will be presented as a means of overcoming the curse of dimensionality common in ANN's. The approach outlined in this thesis shows a reasonable level of success classifying bearing health.

Chapter 1 – Ultrasonic Acoustic Emissions (UAE)

Acoustic measurement in the ultrasonic range requires very specialized sensors capable of reaching above a human's audible range, typically greater than

20 kHz. The inherent characteristics of microphones capable of measurements in this range, specifically the directionality, can be advantageous in environments with low signal-to-noise ratios. It is also important to point out the need for a data acquisition system that is capable of capturing signals at higher frequencies.

Ultrasonic Microphone

The microphone chosen for this work was Avisoft Bioacoustic's CM-16[14]. This microphone was chosen over others due to its high bandwidth, which in this case refers to the microphones' high frequency range. This particular condenser microphone is rated up to 150 kHz, well into the ultrasonic range and can be seen in Figure 1. It is also important to note the directionality of the microphone at higher frequencies. In health monitoring applications it is often desirable to isolate acoustic sources in noisy environments and using an ultra-high frequency directional microphone is one possible way to accomplish this. The frequency response of the microphone is presented in Figure 2, while the polar diagram is presented in Figure 3. It is interesting to note that the intended application for this particular microphone is to monitor bat communication. It is also important to note that a parabolic reflector, commonly used to concentrate acoustic signals to a single focal point therefore extending the range of a particular microphone, becomes less efficient with this particular microphone at higher frequencies (>80 kHz). This is evident after analysis of the polar diagram in Figure 3, and consideration of the "line of sight" issues at higher frequencies. For this reason a parabolic reflector or cone was not used for the work presented in this paper.



Figure 1. Avisoft Bioacoustics CM-16 (© Avisoft Bioacoustics)

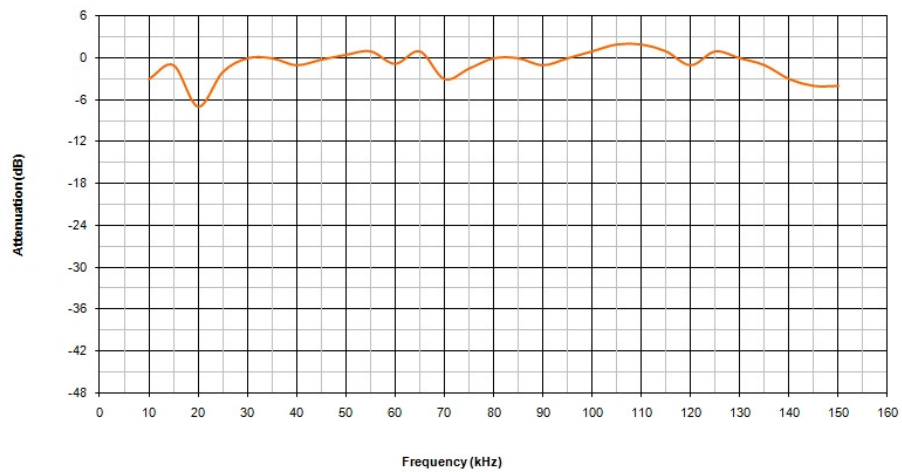


Figure 2. Frequency Response of Avisoft Bioacoustics CM-16 (© Avisoft Bioacoustics)

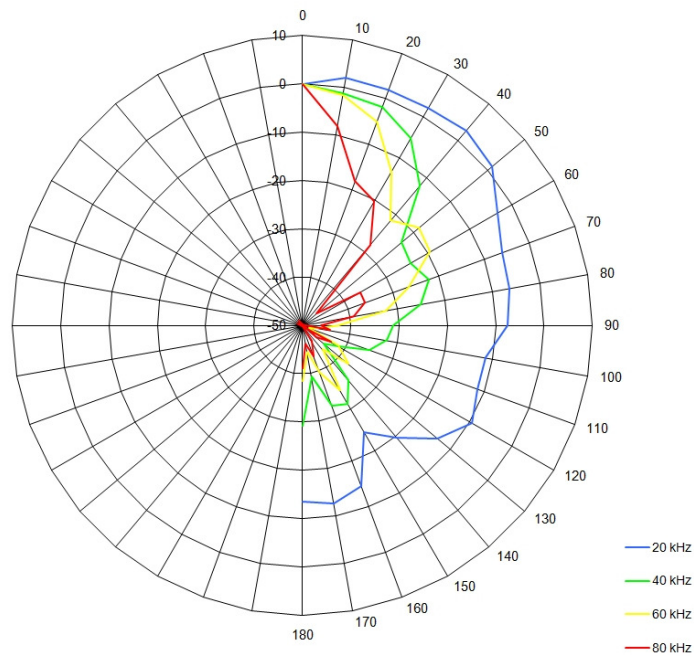


Figure 3. Polar Diagram for Avisoft Bioacoustics CM-16 (© Avisoft Bioacoustics)

Data Acquisition System

In order to fully utilize the microphones' bandwidth (150 kHz) it was necessary to use a data acquisition system capable of sample rates of 300 kHz or greater, following Nyquist's theorem, but more practically a system that could sample around two and a half times the microphones' bandwidth or 375 kHz.

Due to its availability in our lab, IOTech's Wavebook 516E was selected as the data acquisition system. The Wavebook 516E is a 16 bit data acquisition system providing sampling rates up to 1 MHz. It also provided 8-multiplexed channels allowing for additional sensors to be used during testing. Another advantage of using the Wavebook 516E and its supported software was that it allowed for direct streaming of data to the hard drive. When sampling at such high rates this became important, as the sample size was not limited by the size of the computers buffer or random access memory.

Chapter 2 - Laboratory and Experimental Data

A series of laboratory tests were conducted to collect a set of data appropriate to allow for the classification of bearing operating conditions. The intent of this work was to develop a method that was both passive and non-contact based, and this intent dictated the design of the experimental equipment and test procedure.

Laboratory Bearing Test Stand

Acoustic health monitoring of ball bearings is a well studied field. With the goal being, in this study, to classify several different conditions of a particular bearing using ultrasonic acoustic signature, a simple test rig was designed. The test rig was designed around equipment, largely, already available in our lab. The drive used was the Kollmorgen Servostar CD CR03250 – 3 Phase, with control being supplied via a PC with Kollmorgen's MOTIONLINK software. The bearing

test rig can be seen in Figure 4 and Figure 5, while the drive motor is presented in Figure 6.

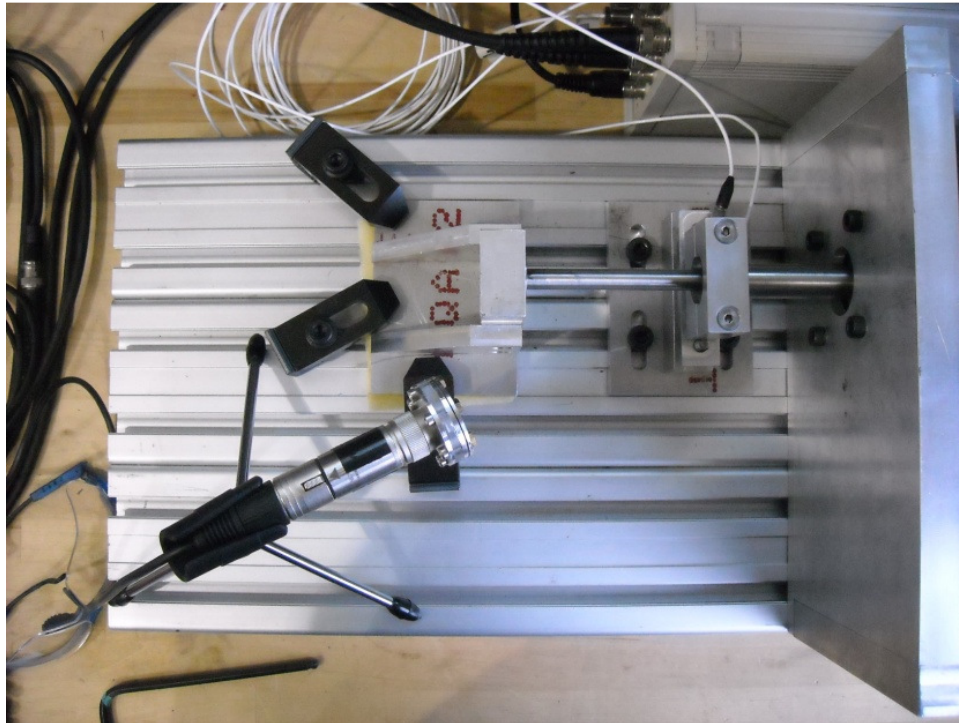


Figure 4. Top View of Bearing Test Rig

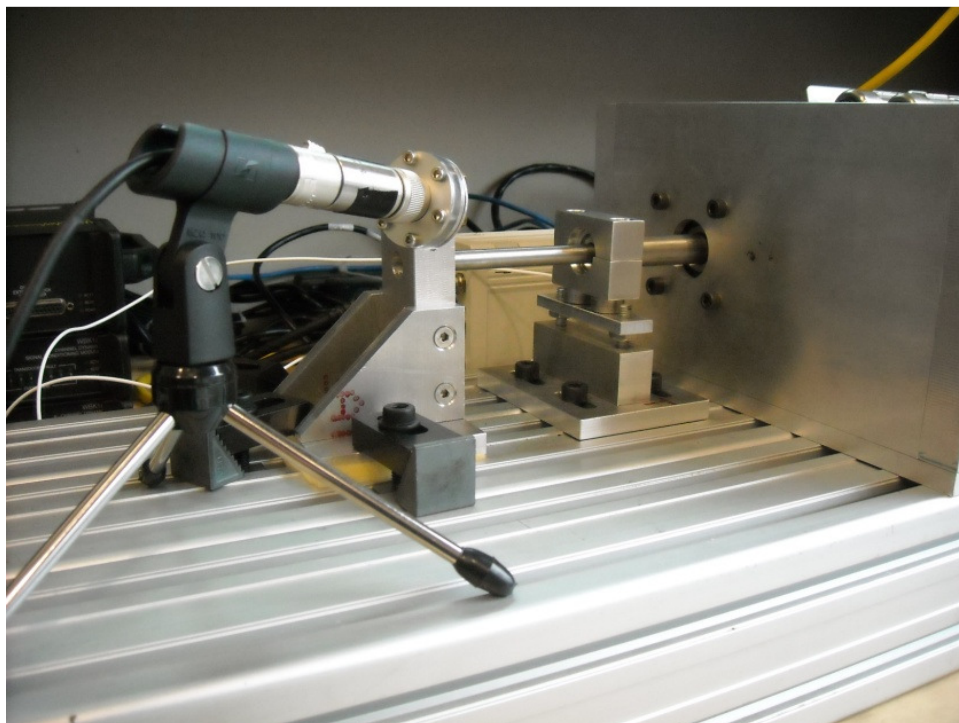


Figure 5. Front View of Bearing Test Rig

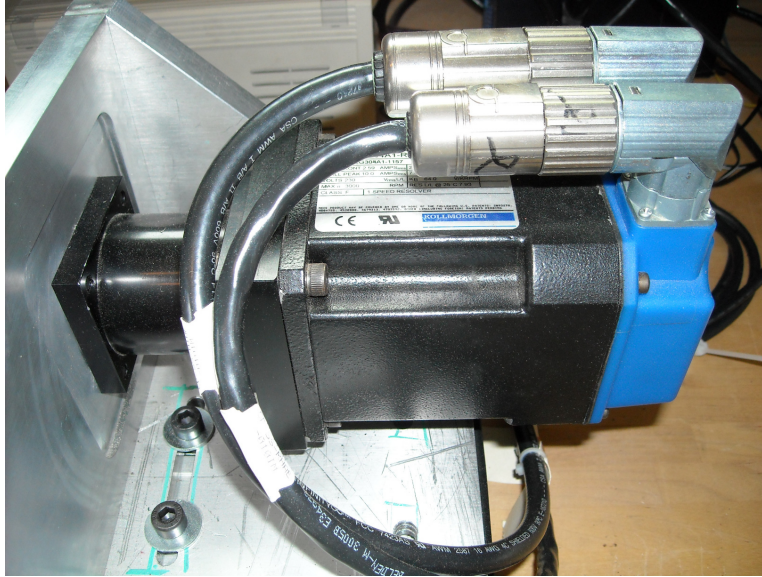


Figure 6. Kollmorgen Drive Motor

Bearings and Preparation

The bearings tested were hardened carbon steel, open ball bearings, designed for a 3/8" shaft, an outer diameter of 1" and a width of 5/16" and were manufactured by The General Bearing Corporation (Part# 21604-01-300 A2432). A full list of the bearing specifications is presented in Table 1. For each condition a total of 5 randomly selected bearings were used. The first condition that was analyzed was the bearings exactly how they arrived, with lubrication. The second condition was the un-lubricated condition. In order to remove the lubrication that the bearings were shipped with, each bearing was cleaned using brake parts cleaner, and a blend of fast acting solvents. The third condition analyzed was un-lubricated and contaminated. The same procedure was used to remove the lubrication and then a small amount of steel shavings were introduced to each bearing. The bearings tested as well as a sample of the steel contaminants used can be seen in Figure 7. A summary of the bearing conditions can be found in Table 2.

Table 1. Bearing Specifications

Ball Bearing Style	Open
Ball Bearing Type	General Purpose
System of Measurement	Inch
For Shaft	3/8"
Outside Diameter	1"
Width	5/16"
ABEC Precision Bearing	Not Rated
Dynamic Radial Load Capacity (lbs)	484
Dynamic Radial Load Capacity Range (lbs)	251 to 500
Maximum rpm	1,500
Maximum rpm Range	250 to 3,000
Temperature Range	-20 to +250 F
Bearing Material	Steel
Inside Sleeve Material	Hardened Carbon Steel
Outside Sleeve Material	Hardened Carbon Steel

Note: Test Bearings were manufactured by The General Bearing Corporation Part# 21604-01-300 A2432.

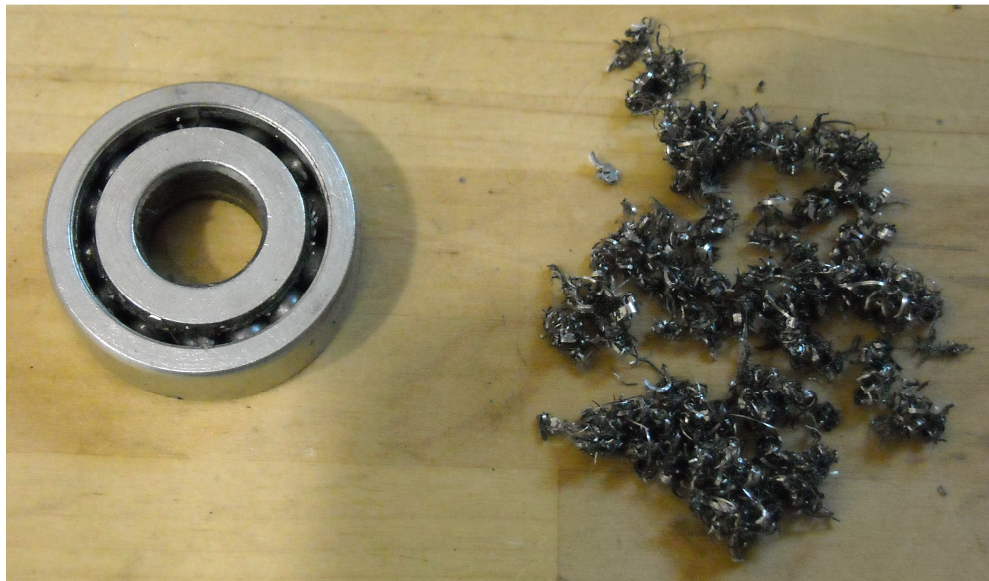


Figure 7. Ball bearing and steel shavings used as contaminant

Table 2. Summary of Bearing Conditions

Condition	Bearing (#)	Lubrication	Contamination
Batch 1	1	Yes	No
	2		
	3		
	4		
	5		
Batch 2	1	No	No
	2		
	3		
	4		
	5		
Batch 3	1	No	Yes
	2		
	3		
	4		
	5		

Test Conditions

All bearings were tested with an arbitrary loading at 1000 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. The data included the ultrasonic acoustic signal as well as two AC load cell measurements, the latter being collected simply as a precaution. It is important to note that the two AC load cells provide a measure of dynamic load and not static load. A summary of the test conditions can be found in Table 3. A diagram of the loading and load cell location is presented in Figure 8. It is important to note that $F_{a,1}$ and $F_{a,2}$ refer to the applied forces, and F_S refers to the resulting force on the shaft or the radial load on the bearing.

Table 3. Summary of Test Conditions

Condition	Value	Units
Shaft Speed	1000	rpm
Duration	2	sec
Loading	Arbitrary	NA

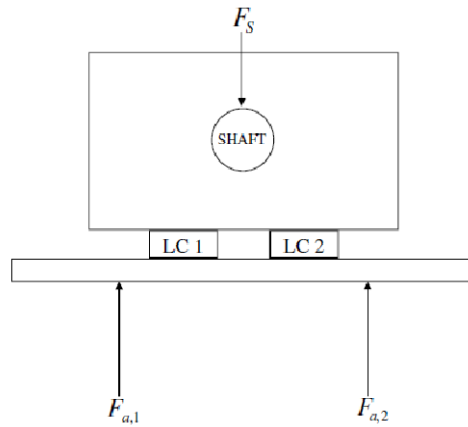


Figure 8. Test Fixture Diagram

A plot of the typical time series data collected for each bearing, using the aforementioned test conditions and hardware is presented in Figure 9.

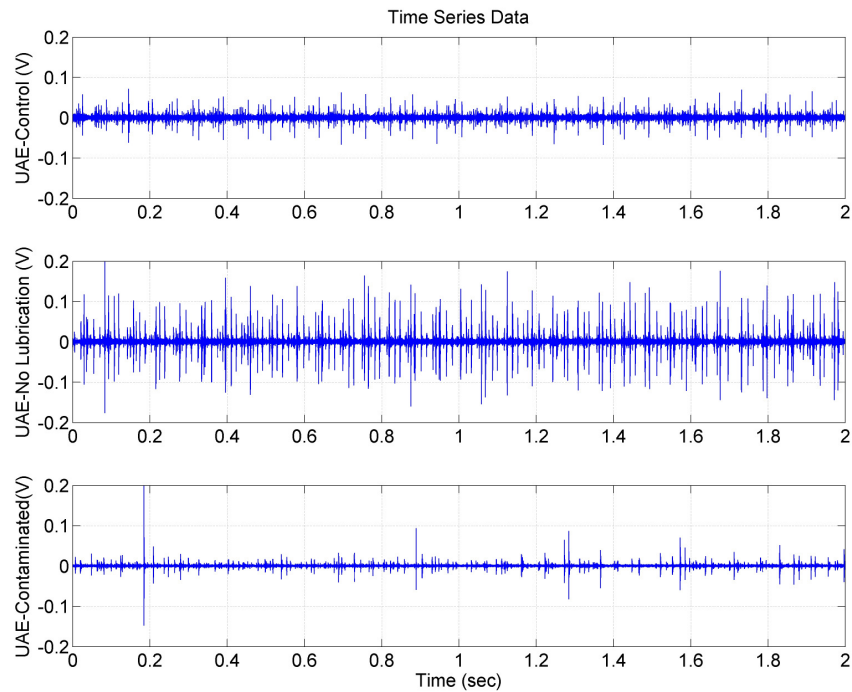


Figure 9. Typical Time Series Data

Chapter 3 - Power Spectral Density and Signal Processing

There is a fair amount of pre-processing of the UAE data that is necessary before it can successfully be used as an input to an artificial neural network. It is important to point out that the neural network will use these inputs to solve the classification problem. It is necessary to filter the input signal to ensure that we are not only looking at frequency ranges at which the equipment is capable of making meaningful measurements but also regions where there are meaningful differences in the signal content. Once the UAE has been filtered, the power spectrum is calculated which allows each signal to be analyzed in the frequency domain, as opposed to the time domain. It is the result of this transform, the power spectrum, which is used to feed the neural network.

High Pass Filtering of UAE

After analyzing the bandwidth of the microphone, Figure 2, it can be seen that the sensor attenuates signals below approximately 15 kHz. To isolate the ultrasonic range, commonly accepted as signal content above 20 kHz, a high pass second order Butterworth filter was used with a cutoff frequency equal to 20 kHz.

Power Spectral Density of UAE

The power spectrum of each acoustic signal was computed using Welch's averaged modified periodogram method for spectral estimation [15]. The two considerations made in the specifications of the frequency vector were the number of power spectrum points needed to make a meaningful comparison between two acoustic signals and the computational efficiency to do so. The signal was divided into windows of length 256 and windowed using a hamming window with 50% overlap. The frequency vector was constructed using 256 linearly spaced points with a range of 0-120kHz. This frequency range captured all observed trends and the number of points at which the power spectrum was calculated, 256 (2^8), provided sufficient resolution in capturing relevant trends. The results of the

power spectrum are later used as the inputs to the neural network, which casts and attempts to solve the multi-class classification problem.

Results of Pre-Processing

The results of the aforementioned pre-processing; high pass filtering and power spectrum are presented in Figure 10 through Figure 13. The results for individual bearings from the first and second rounds of testing are shown Figure 10 and Figure 11, respectively. It is important to note that the bearings are consistent through each round with the only difference between rounds 1 and 2 being the dates on which the data was collected. The average of each batch of bearings for rounds 1 and 2 are presented in Figure 12 and Figure 13, respectively. Perhaps more useful are the error bars that are presented in these two figures, which represent plus and minus one standard deviation for the batch. This is particularly important since the classification problem becomes significantly easier to solve when there are distinct boundaries between each class, or in this case the error bars from each class do not overlap those from the remaining classes. There are few if any regions where this is satisfied. However, there are regions that do come close and it is these regions that will be focused on.

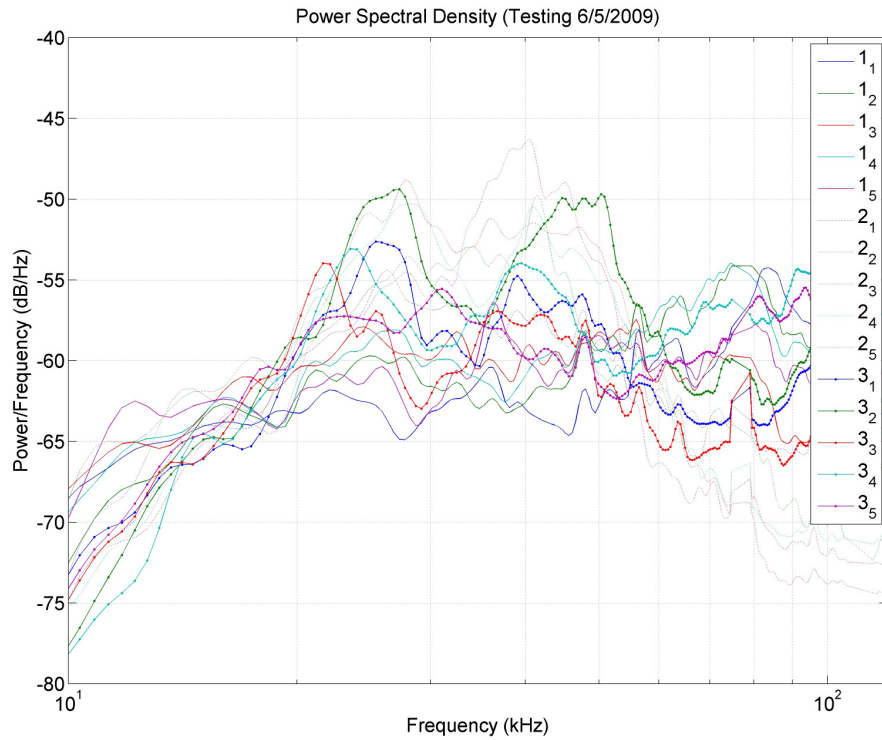


Figure 10. Acoustic Signatures of Bearings (Round 1)

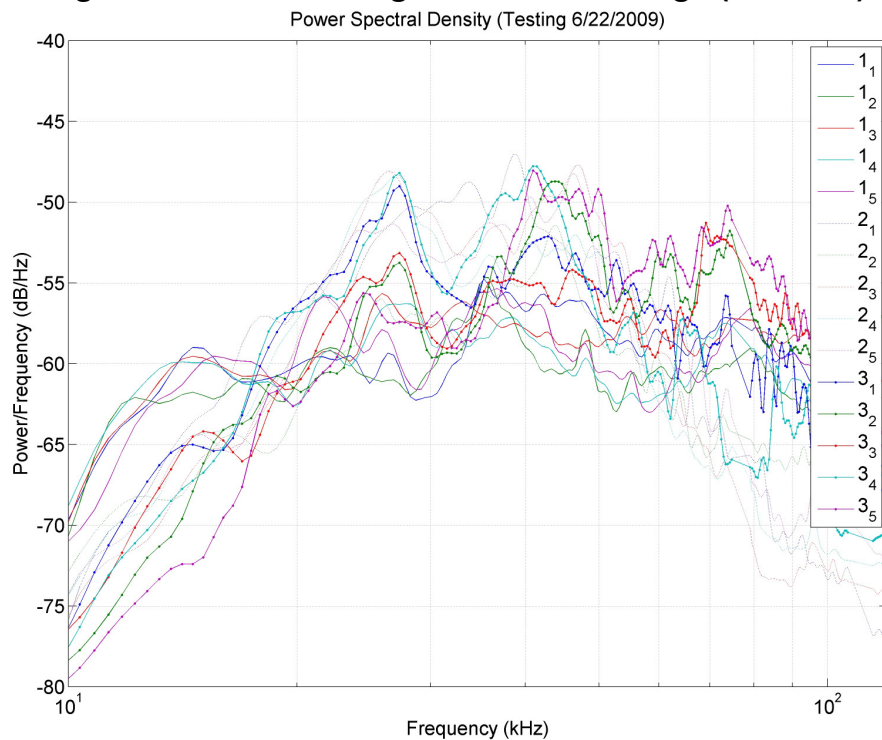


Figure 11. Acoustic Signatures of Bearings (Round 2)

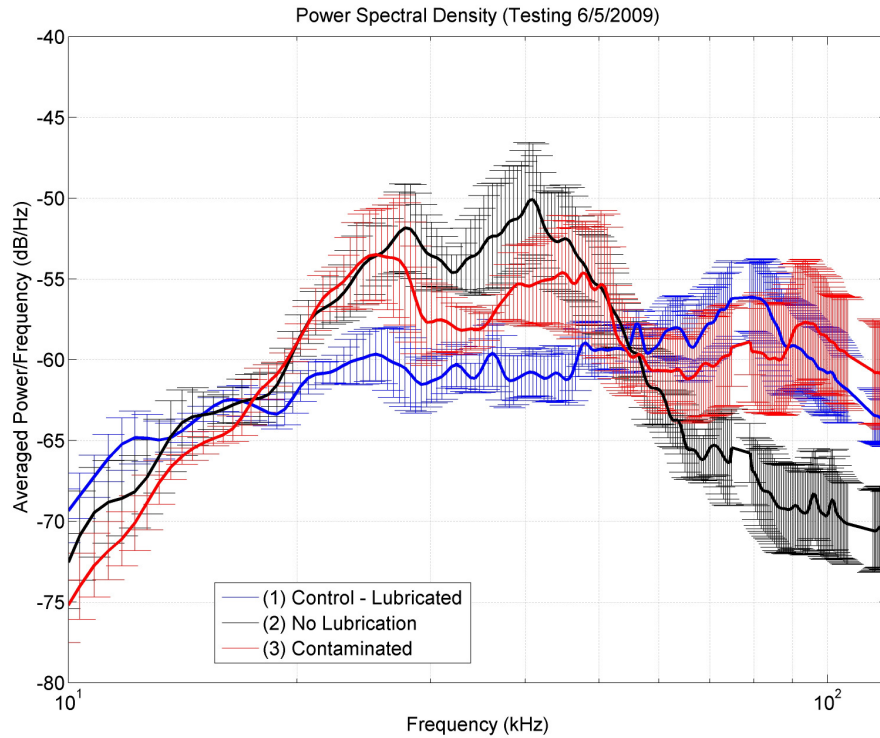


Figure 12. Average of Acoustic Signatures with Error Bars Corresponding to 1 Standard Deviation (Round 1)

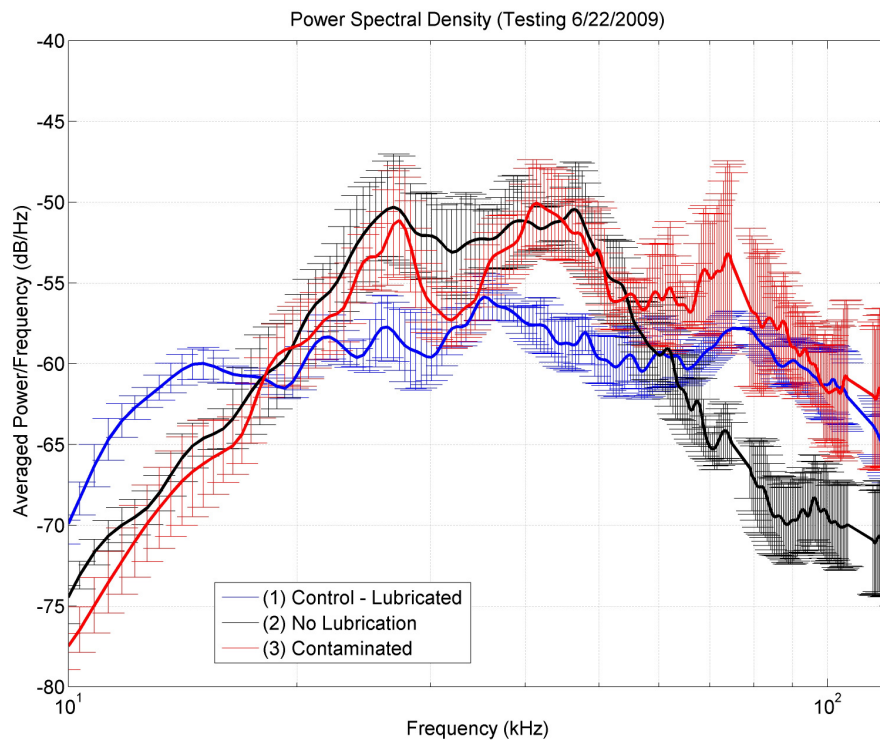


Figure 13. Average of Acoustic Signatures with Error Bars Corresponding to 1 Standard Deviation (Round 2)

Isolation and Averaging of Frequency Bands

It is desirable to keep the number of inputs to the neural network at a minimum, as complexity does not always result in a better performing classifier. The values of the power spectrum are averaged across a frequency range to obtain the inputs used in the neural network. The two regions ranging from 20-50 kHz and 90-110 kHz, respectively, become the inputs to the neural network. It is worth mentioning that both of these regions are well into the ultrasonic range, above the human threshold for hearing, which means that the microphones' directionality at higher frequencies is inherently being leveraged.

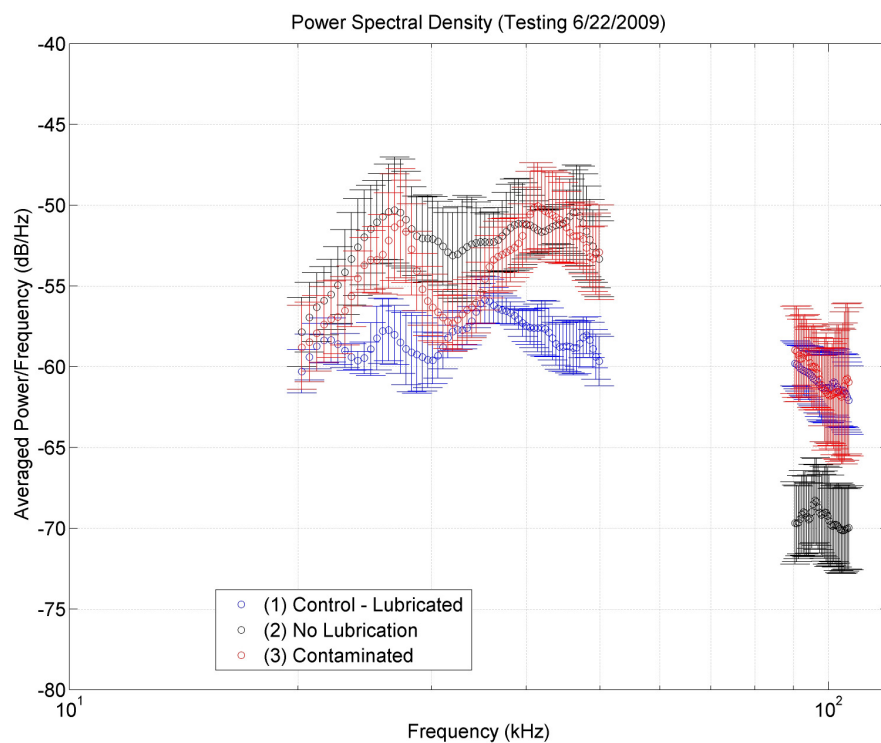


Figure 14. Isolated Frequency Ranges from Rd. 1 of Testing

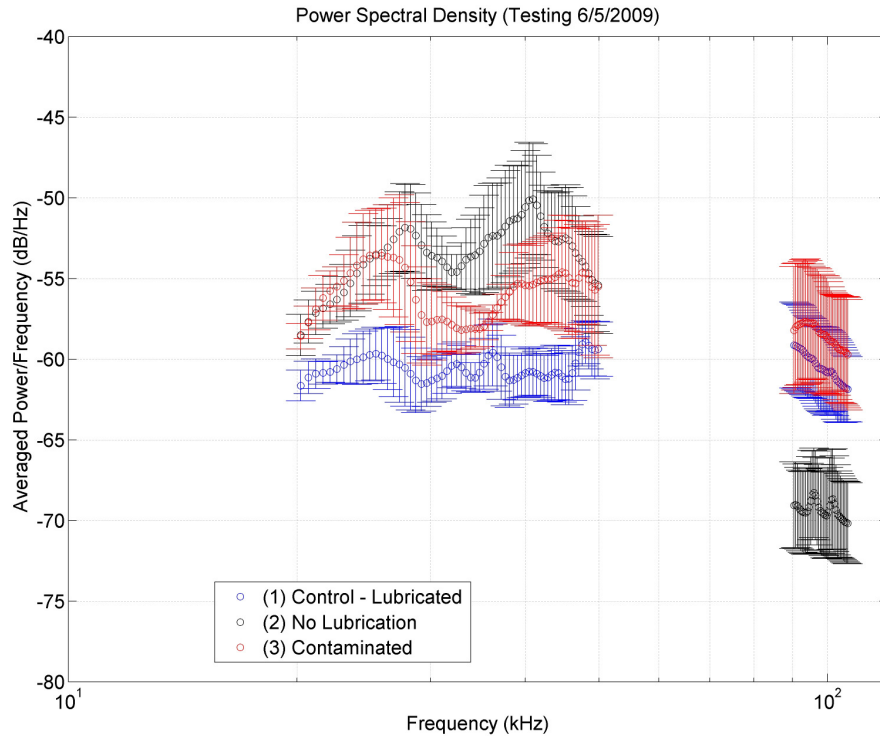


Figure 15. Isolated Frequency Ranges from Rd. 2 of Testing

Determination of Frequency Ranges

It is possible to statistically determine at which frequencies there are significant differences. A simple hypothesis test can be performed to determine whether two sets of data are statistically different. This is performed by calculating the mean of each data set and estimating the standard deviation. A t-test is used in this case because it is possible to readily calculate the standard deviation in each data set. The results of this particular form of hypothesis testing, with a confidence interval of 95%, are presented in Figure 16. A value of 1 corresponds to statistically different values between the two data sets (labeled on the y-axis), in other words, it is possible to reject the null hypothesis. The vertical lines represent the limits of the frequency regions used in the previous section and of the analysis performed throughout the rest of this work. It is important to note that throughout the regions chosen there is at least one pair of data sets that is statistically different either from round 1 or 2 of testing.

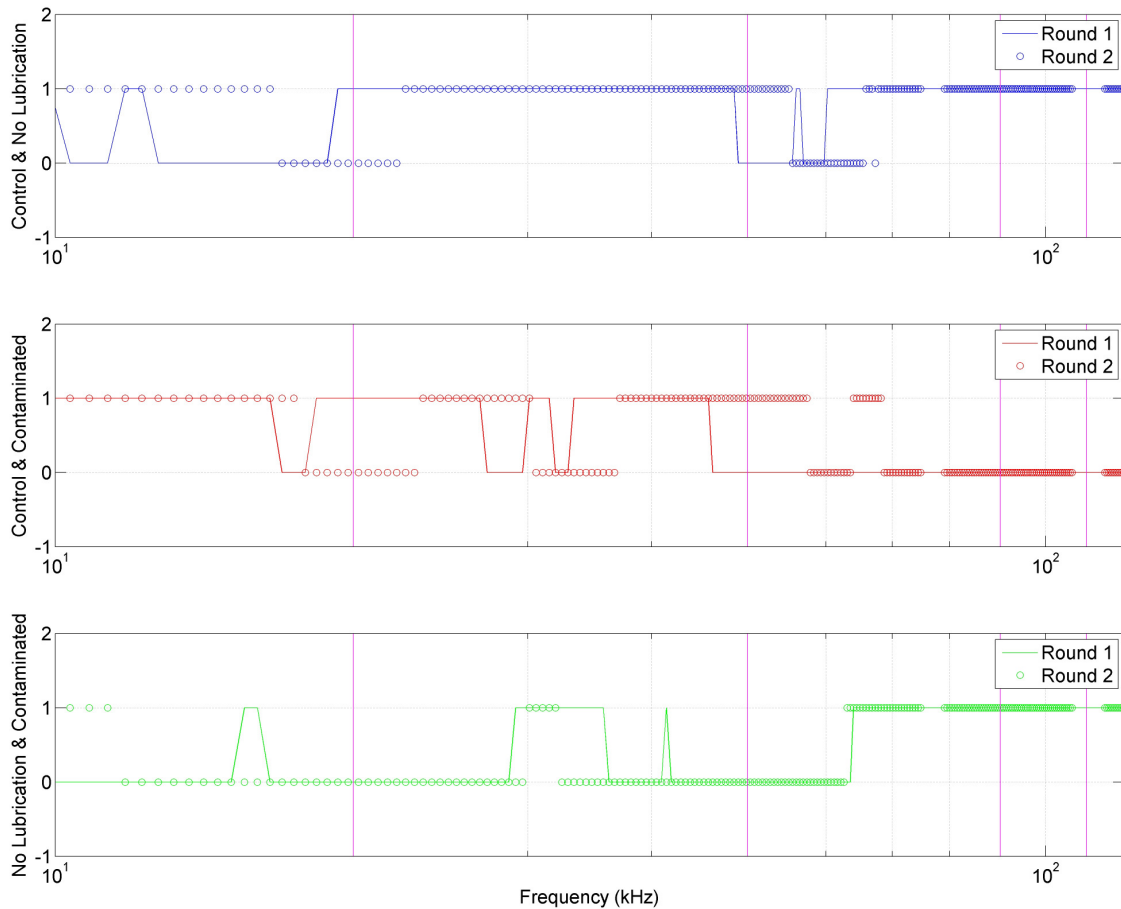


Figure 16. Regions with Statistically Significant Differences in Power Spectrum

Chapter 4 - Neural Network Design

Multi-Class Pattern Recognition Problem

An active area of research is the application of neural networks to multi-class pattern recognition problems. Until recently, the most common way to approach multi-class problems is to decompose them into several 2 class problems [7]. This is mostly because the two class classification problems have been extensively studied within the neural network community. It is also important to point out that often the best, most accurate and efficient neural network configuration happens to be problem dependent [7]. The subject of this paper is not to necessarily determine the optimum neural network structure for this specific problem but to design a network that is capable of classifying bearings using their

ultrasonic acoustic signature(s). It is also desired to keep the approach generic enough, to allow for application to other components and systems.

Motivation for Using Artificial Training Data

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 [16]. This generalization is presented in Figure 17. The same generalization can also be made about model complexity and validation error [16] and is presented in Figure 18, which can be attributed to the over-fitting problem in the design of ANN's. However, it is advantageous to increase the size of the training data set because of the generalization presented in Figure 19 [16]. This provides sufficient motivation for creating artificial training data to address the bearing operating condition classification problem. Using just 5 bearings per class leaves much to be desired with respect to adequate coverage of the design space during training. The idea is to create enough data where the error seen with the training subset of the network approaches that of what is seen with the testing subset. If the artificial data is carefully created, this will also minimize the effect that bad inputs, or outliers, have on the network training.

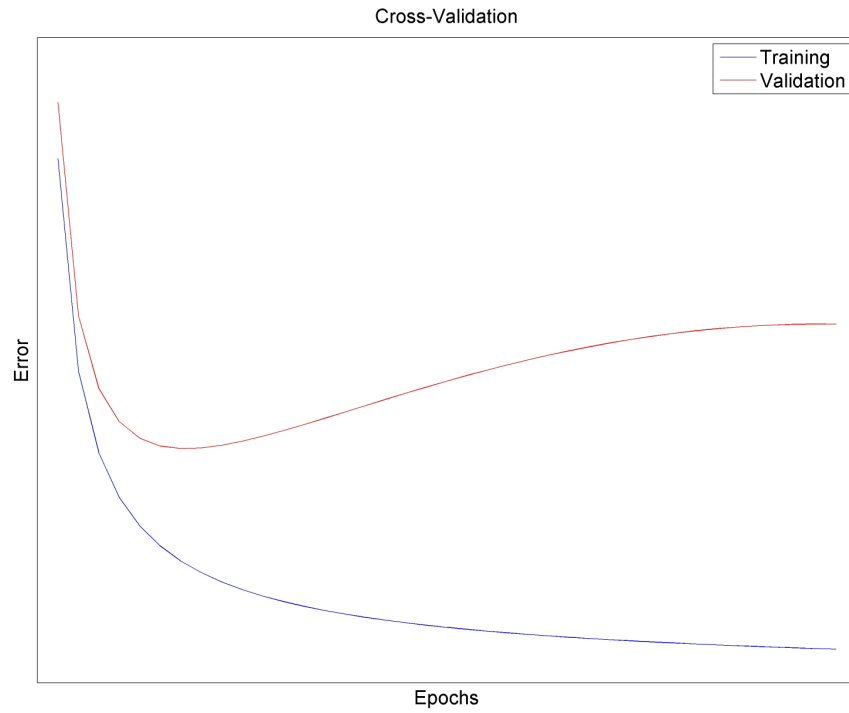


Figure 17. Network Error as a Function of Training Epochs

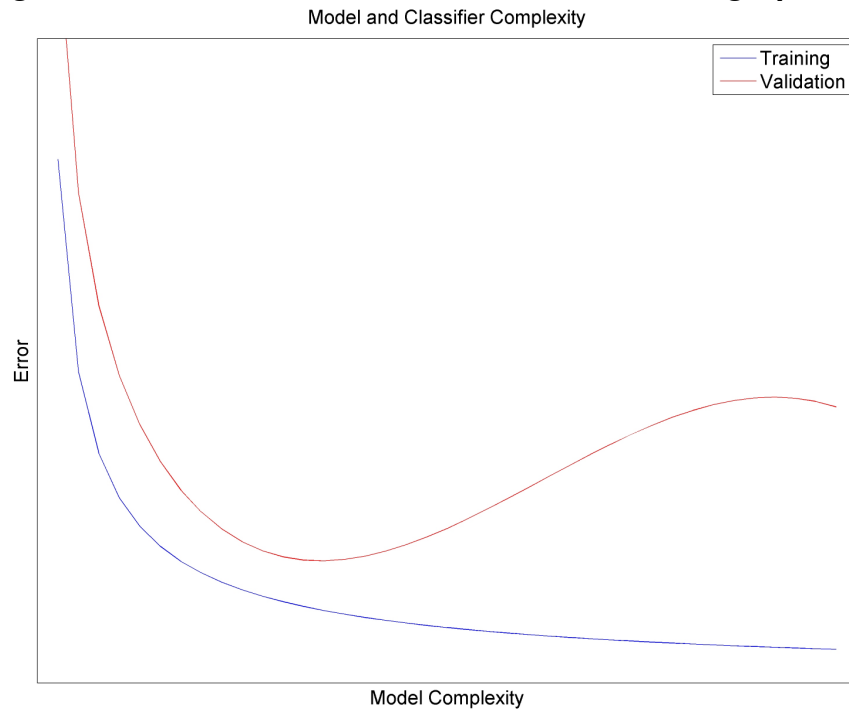


Figure 18. Network Error as a Function of Model Complexity

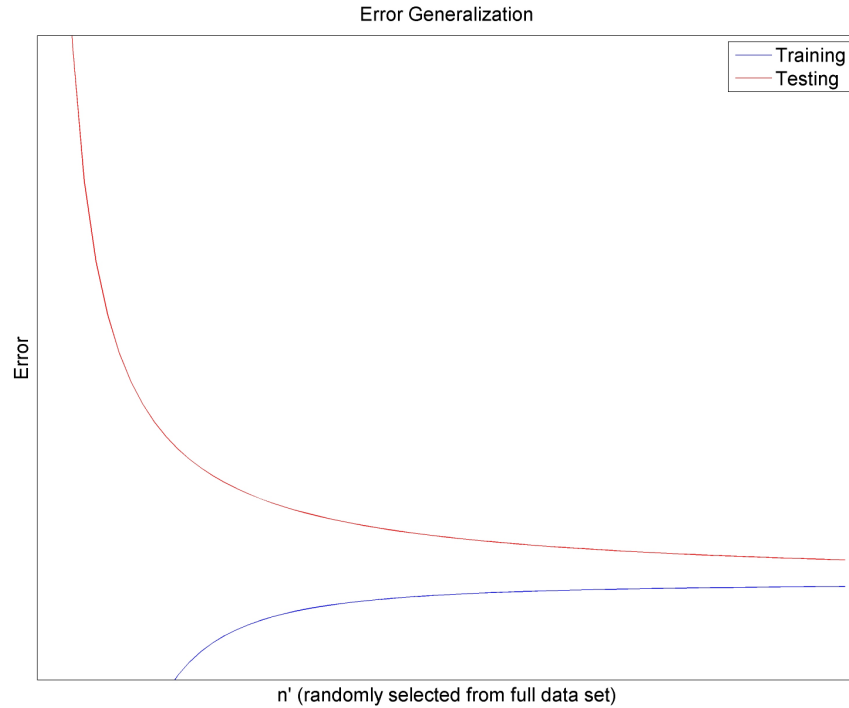


Figure 19. Network Error as a Function of Training Dataset Size

Artificial Training Data Sets

The method chosen to accomplish these two goals was to create a normal (Gaussian) distribution of random data points of each power spectral density value of the ultrasonic acoustic signature about the respective mean values. Using a normal distribution will both reinforce training boundaries with its increased coverage of the design space and will also minimize the effect of bad input data, or outliers, by reinforcing the average of each set of data, as opposed to giving equal weight to each experimental data point. As long as the same number of random points are used in each distribution it is possible to maintain a balanced set of training data. A 2-D visualization of the artificial training data created from the first round of testing using this method can be seen in Figure 20. The standard deviation used to create the normal distribution was the average standard deviation of all the power spectrum values in that particular frequency range. The mean

value is simply the mean of the power spectrum in the given frequency range. It is important to note that this method can easily be extended to N dimensions.

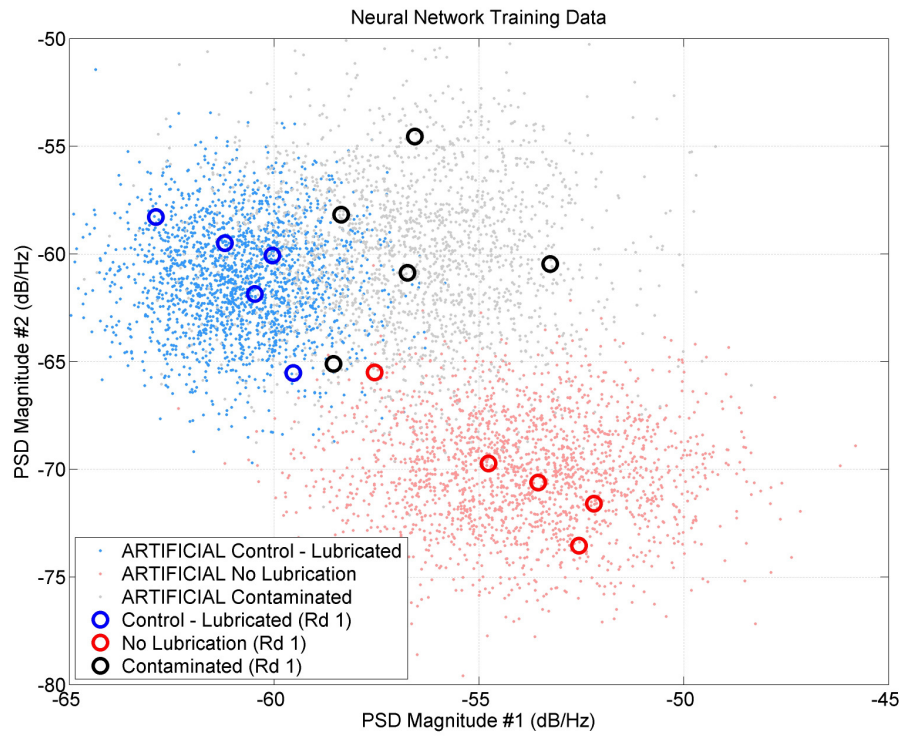


Figure 20. Artificial Training Data using Normal Distribution (Round 1 Experimental Data)

Artificial Training Data Sets using Linear Transformation

It is also possible to create the artificial training data using a linear transformation to align one of the coordinate axes used in calculating the standard deviation with the result of a linear regression analysis performed on the data set. 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. This method will result in very little difference in artificial training data from the method described in the previous section if the data points do not show a strong linear correlation. However, if there is a strong linear correlation in the data the artificial training data created using this method will differ greatly from the results presented in Figure 20. In two dimensions the linear

transformation is performed by first performing a linear regression analysis of each data set, and using the slope (m) of this analysis to calculate the desired transformation angle (θ) as seen in Equation 1.

$$\theta = \tan(m)^{-1} \quad (1)$$

It is also important to note the offsets (x_0, y_0) used in the transformation, as it is desired to perform the transformation about the mean values of each data set as seen in Equation 2.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \quad (2)$$

Once this transformation is complete it is possible to calculate the standard deviation of x' and y' to create a cloud of points representing a normal distribution based on experimental data set. Once this cloud is created it is necessary to transform the points back to the original coordinate system using Equation 3.

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}^{-1} \begin{bmatrix} x' \\ y' \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \quad (3)$$

The result of this method is to create a set of artificial training data that is presented in Figure 21, it is important to note the same laboratory data sets were used to create the artificial training data in Figure 20 and Figure 21.

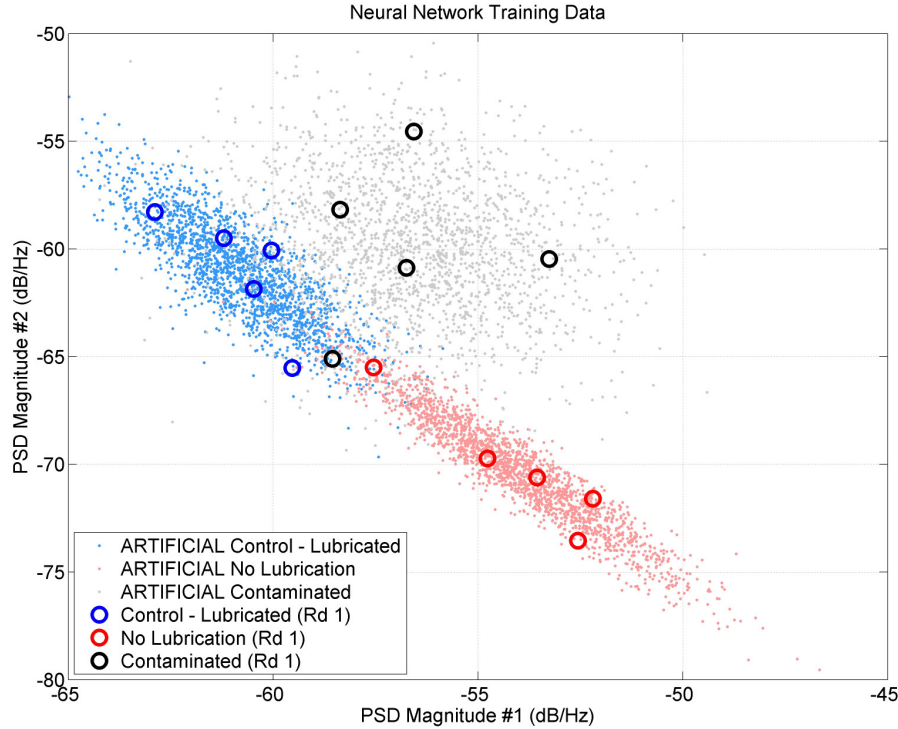


Figure 21. Artificial Training Data Created using Linear Regression Analysis and Linear Transformation (Round 1 Experimental Data)

This method can also be extended to N dimensions. In two dimensions the distribution represents an ellipse, in three dimensions it results in an ellipsoid and in N dimensions a distributions that represents a hyper ellipsoid is created.

The previously mentioned methods for creating training data for the neural network both have advantages. The normal distribution shown in Figure 20 provides superior coverage of the design space and there is sufficient overlap in the regions where a classification boundary would be expected. However, the use of the training data in Figure 21 is superior in the reinforcement of each pattern that displays linear trends using a regression analysis to identify the necessary coordinate transformation such that the standard deviation of both dimensions are at a minimum and maximum respectively. The combination of these two methods is shown in Figure 22.

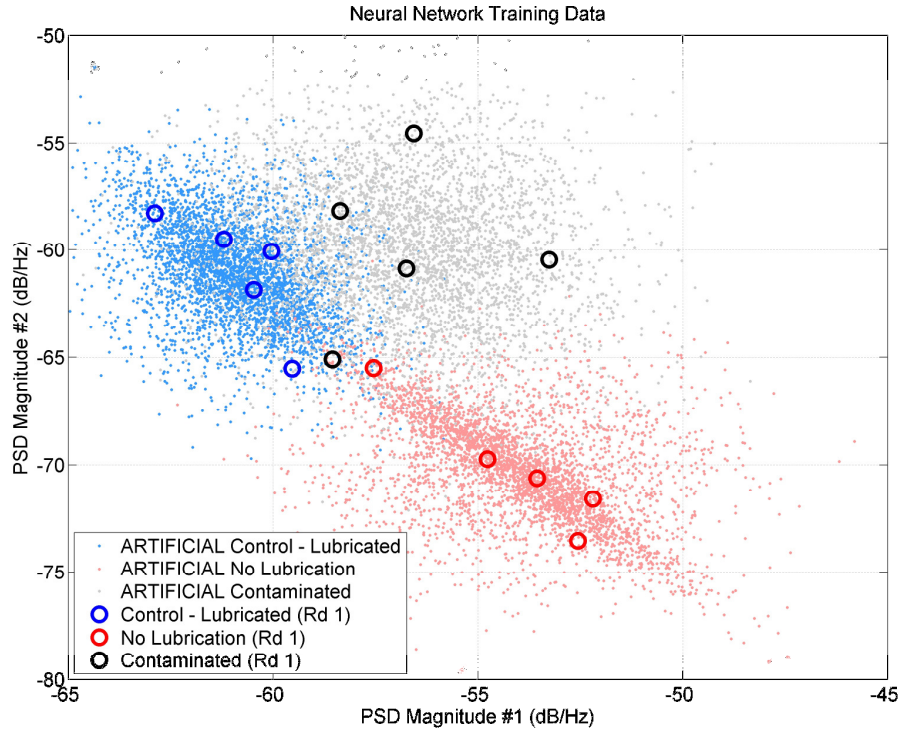


Figure 22. Combination of Methods to Create Artificial Training Data: Normal Distribution, and Normal Distribution of Data with Coordinate Transformation ((Round 1 Experimental Data))

Using this method to create artificial training data for both rounds of testing resulted in the set of artificial training data seen in Figure 23. It is important to note that each round of testing used the same number of artificial training data points, meaning the number of training points in Figure 23 is twice that of Figure 22.

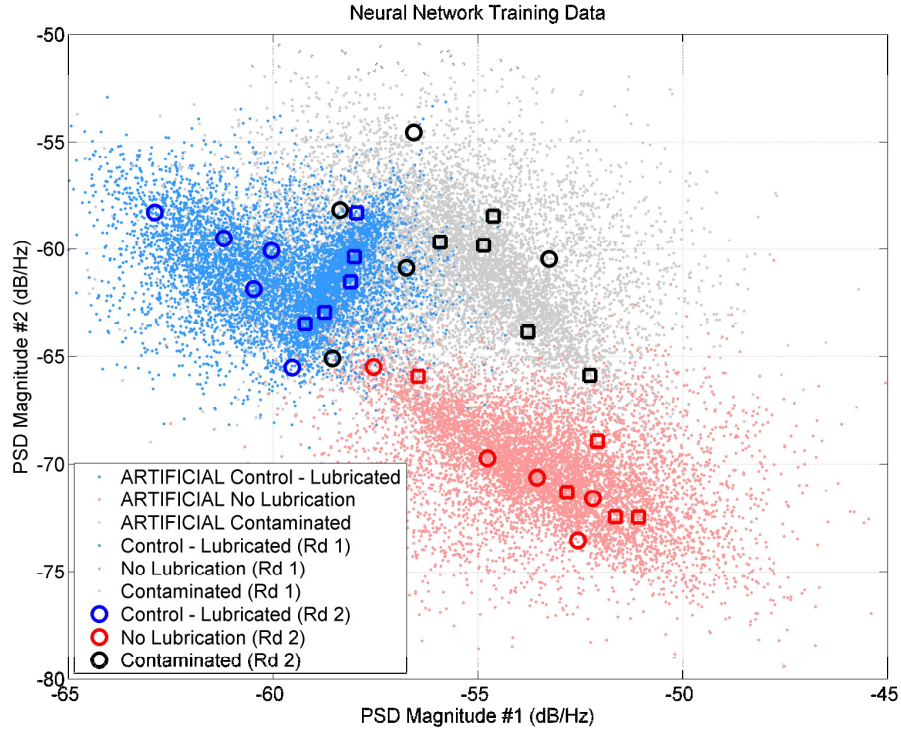


Figure 23. Artificial Training Data Created using Both Rounds of Testing

Neural Network Specifications

A two layer feed forward back-propagation network was used. The number of input nodes is based on the dimension of the input signal, which in this case is the number of data points from the power spectral density. Figure 24 is representative of each node with the output of each neuron being a function of the activation function, weights (w), biases (b) and inputs (p) [18]. The differentiable activation function was specified as the tan sigmoid, or hyperbolic tangent, for all nodes: hidden layer, and output layer. Since there are 3 classes, 3 output nodes are used with each output bit corresponding to a specific operating condition. The number of hidden nodes is largely based on the complexity of the pattern that is to be classified.

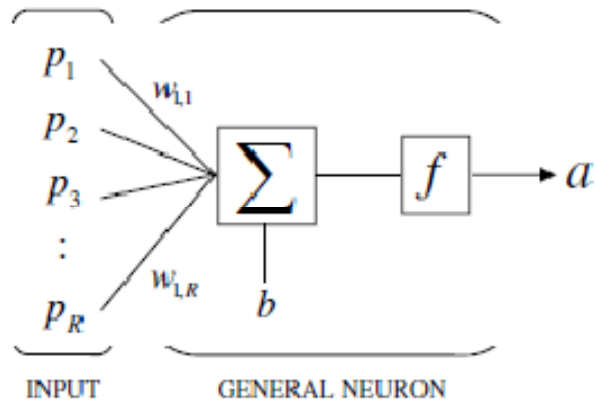


Figure 24. Neuron Model with R inputs

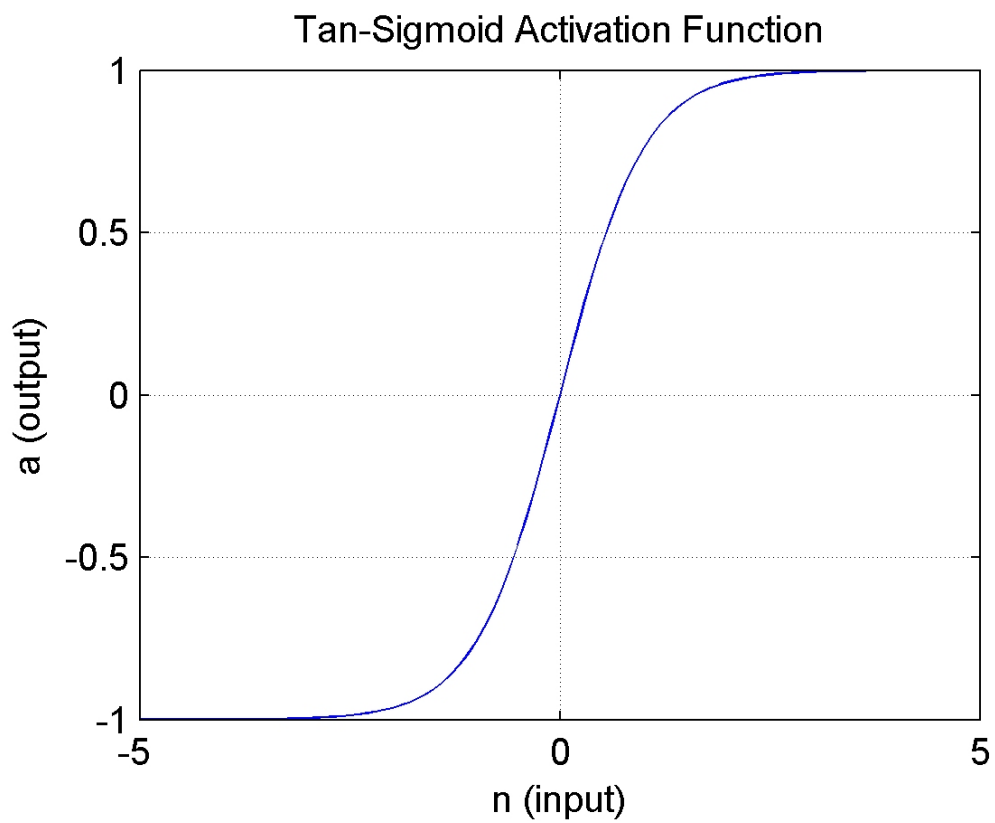


Figure 25. Tan-Sigmoid (Hyperbolic Tangent) Activation Function

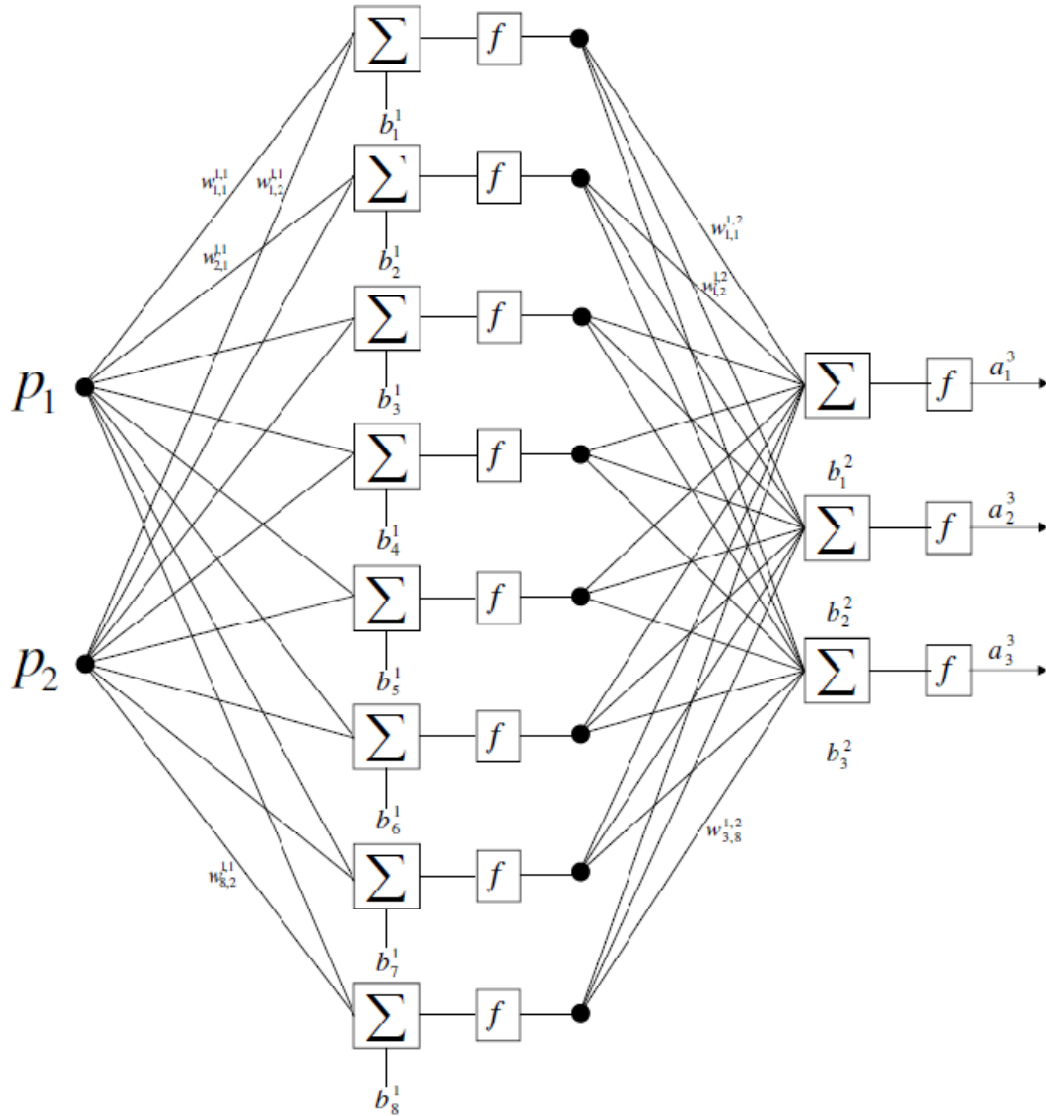


Figure 26. Multi-layer Perceptron (Neural Network)

The more complex the pattern, analogous to a higher degree of non-linearity, the higher the number of hidden nodes required for successful classification. The decision function that is used to determine the threshold values of each bit required for successful classification is discussed later. The next two figures, Figure 27 and Figure 28, are presented to show the difference in classification boundaries when increasing the number of nodes in the hidden layer.

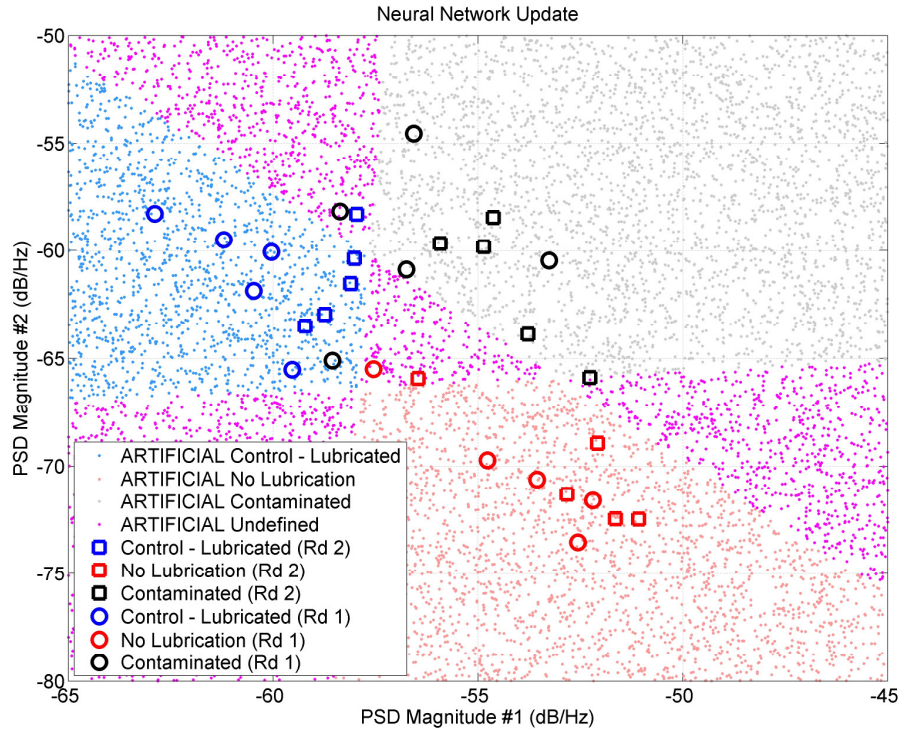


Figure 27. Neural Network with no Hidden Layer: Updated using Both Rounds of Testing in Training.

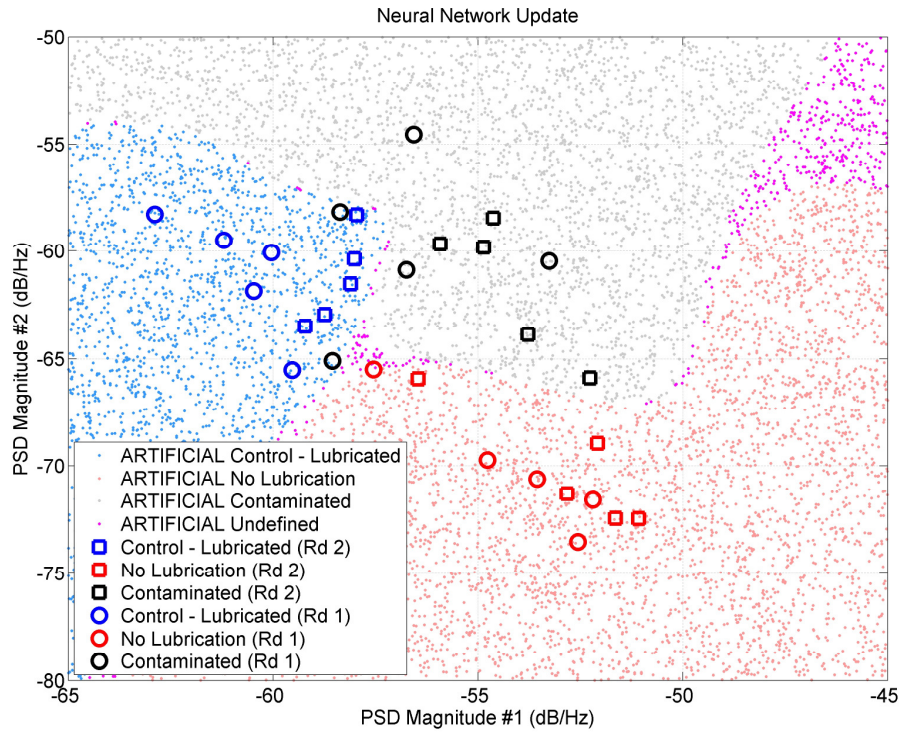


Figure 28. Neural Network with 8 Nodes in the Hidden Layer: Updated using Both Rounds of Testing in Training.

Neural Network Training Algorithm

One of the more traditional methods of training network parameters is the Widrow-Hoff (LMS) algorithm shown in Equation 4 and Equation 5 [18].

$$W(k+1) = W(k) + 2\alpha e(k)p^T \quad (4)$$

$$b(k+1) = b(k) + 2\alpha e(k) \quad (5)$$

A more efficient method is the Levenberg-Marquardt Algorithm. The Levenberg-Marquardt algorithm is similar to other quasi-Newton methods in that it attempts to approach the speed of algorithms that use second order derivatives, without actually calculating the Hessian matrix presented in Equation 6 [18][19]. Instead, as a result of using an error function (mean squared error) that is a sum of squares function the Hessian can be approximated using Equation 7 [19]. It is important to note here that the function e is the network mean squared error, and x_i are the network parameters (weights and biases).

$$H(f) = \begin{bmatrix} \frac{\partial^2 e}{\partial x_1^2} & \frac{\partial^2 e}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 e}{\partial x_1 \partial x_n} \\ \frac{\partial^2 e}{\partial x_2 \partial x_1} & \frac{\partial^2 e}{\partial x_2^2} & \dots & \frac{\partial^2 e}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 e}{\partial x_n \partial x_1} & \frac{\partial^2 e}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 e}{\partial x_n^2} \end{bmatrix} \approx J^T J \quad (6)$$

$$H(f) \approx J^T J \quad (7)$$

The update when using the Gauss-Newton optimization method then becomes equal to Equation 8. Where e is the network output error which in this case is the network outputs subtracted from their respective training targets, x are the network parameters and J is the Jacobian matrix shown in Equation 9. It is also important to note that the equations uses the gradient of the output error described as $J^T e$.

$$x_{k+1} = x_k - [J^T J]^{-1} J^T e \quad (8)$$

$$J(\underline{x}) = \begin{bmatrix} \frac{\partial e_1(\underline{x})}{\partial x_1} & \frac{\partial e_1(\underline{x})}{\partial x_2} & \cdots & \frac{\partial e_1(\underline{x})}{\partial x_n} \\ \frac{\partial e_2(\underline{x})}{\partial x_1} & \frac{\partial e_2(\underline{x})}{\partial x_2} & \cdots & \frac{\partial e_2(\underline{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial e_N(\underline{x})}{\partial x_1} & \frac{\partial e_N(\underline{x})}{\partial x_2} & \cdots & \frac{\partial e_N(\underline{x})}{\partial x_n} \end{bmatrix} \quad (9)$$

The advantage in using the Levenberg-Marquardt algorithm, Equation 10, lies in the simple modification to the update, where parameter u is adjusted after each step. This allows the algorithm to behave as the steepest descent algorithm for large values of u and the Gauss-Newton algorithm for small values of u [19]. This is helpful because as the algorithm gets closer to a minimum it transitions to the Gauss-Newton method, which is more efficient when close to an error minimum [18].

$$x_{k+1} = x_k - [J^T J + uI]^{-1} J^T e \quad (10)$$

The key to using the Levenberg-Marquardt algorithm is the computation of the Jacobian matrix. The connection must be made that the derivative of the error terms with respect to the network parameters will equal the derivative of the activation function, shown in matrix form in Equation 11. This becomes a key feature in neural network design, as it is necessary to backpropagate these derivatives from the final layer of the network to the first, in order to fully populate the Jacobian matrix. The backpropagation algorithm used for the Levenberg-Marquardt algorithm is defined as Equation 12 at the final layer M and Equation 13 for all other layers.

$$\dot{F}^k(\underline{n}^k) = \begin{bmatrix} \dot{f}^k(n^k(1)) & 0 & \cdots & 0 \\ 0 & \dot{f}^k(n^k(2)) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \dot{f}^k(n^k(Sk)) \end{bmatrix} \quad (11)$$

$$\Delta^M = -\dot{F}^M(n^M) \quad (12)$$

$$\Delta^k = -\dot{F}^k(n^k)x^{k+1}\Delta^{k+1} \quad (13)$$

Fuzzy Logic Classification Boundaries

As no network is ideal, when applied to the test points it is highly unlikely to return ideal classifications, where one logical state is equal to 1 and the others zero. Rather will have chosen to use the fuzzy logic to classify the outputs of the neural network using Equation 14.

$$C_i \in \begin{cases} a_i > 0.5 \\ a_j < 0.5 \end{cases} \quad \begin{matrix} i = \text{class} \\ j \neq i \end{matrix} \quad (14)$$

It is necessary to use a decision function to classify the bearing conditions based on the values of the three output nodes. This decision function is tunable, as a good starting point the threshold values of 0.50 were used. In order for a bearing to be associated with a certain class, the logical state specific to that class had to exceed 0.50 and the other two logical states had to be less than 0.50. If these conditions were not satisfied for a particular case than the bearing or region of design space was considered undefined.

Results of Neural Network

The correct classification rates and the mis-classification rates for each class are presented in Table 4 and Table 5, respectively. The classification rate refers to the number of bearings in class i that were successfully classified as being in class i . The mis-classification rate, in this case, refers to the number of bearings there were classified as class i but in fact do not belong to class i . A visualization of the results in 2-D can be seen in Figure 30. It is important to note that the training data and artificial training data used for the boundaries seen in Figure 30 are solely a function of the first round of testing. Both round 1 and round 2 of testing were used to train the neural network results shown in Figure 31, 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 circle, 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.

Table 4. Classification Results (Training and Application)

		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 5. Mis-Classification Results (Training and Application)

		Mis-Classification 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	

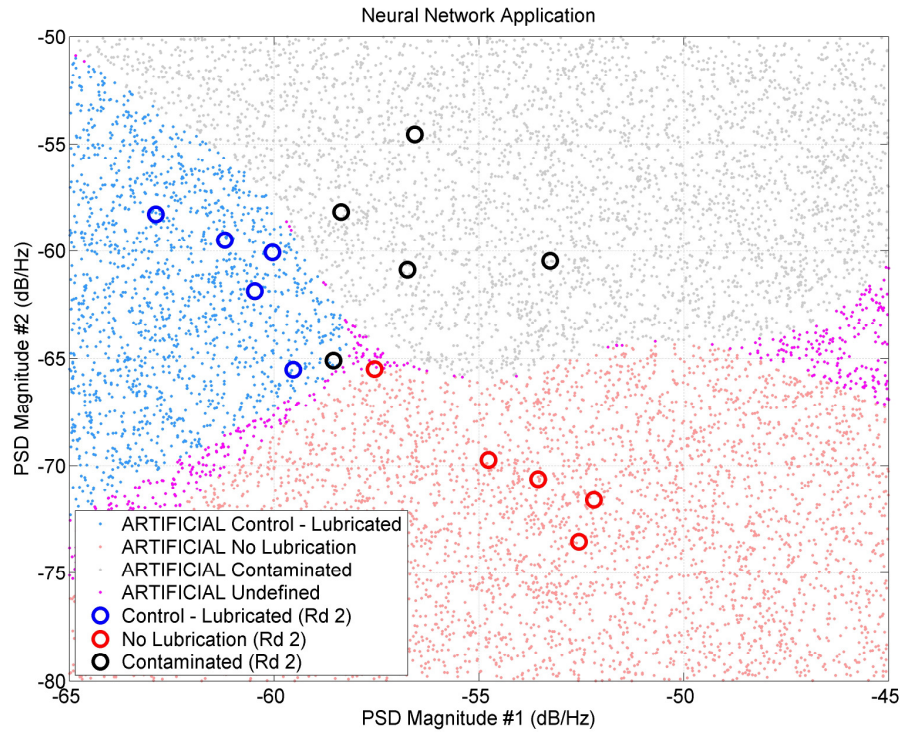


Figure 29. Neural Network Training Method Validation of Classification Boundaries in 2-D

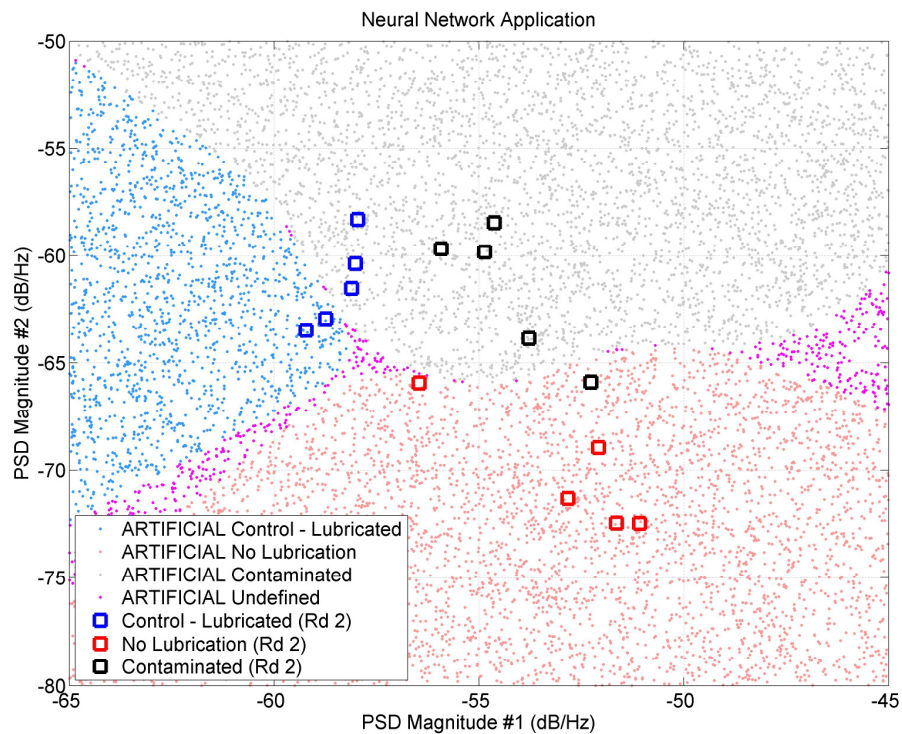


Figure 30. Neural Network Application to Rd 2 Data with Visualization of Classification Boundaries in 2-D

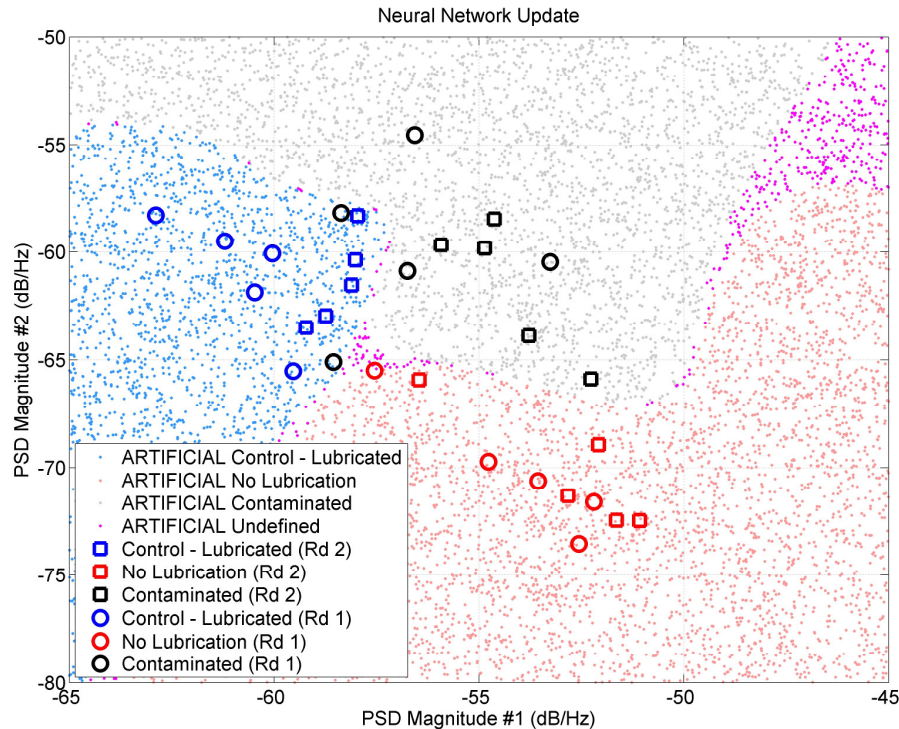


Figure 31. Neural Network Updated using All Data (Rd 1 and Rd 2) in Training with Visualization of Classification Boundaries in 2-D

Chapter 5 - Norfolk Southern Testing

It is both desirable and necessary to test both properly functioning components as well as failing or failed components to collect the necessary data to eventually train the neural network. In the absence of UAE from both types of components it is extremely difficult to train a neural network to successfully solve the pattern classification problem. However, there is significant value in determining if UAE exists in the field and if it could be identified. The next section is a brief analysis of data collected from properly functioning diesel locomotives at Norfolk Southern's Shaeffer's Crossing service depot on September 3, 2009. A representative picture of experimental setup in the field is presented in Figure 1Figure 32



Figure 32. Representative Testing at Norfolk Southern (September 3, 2009)

Ultrasonic Acoustic Emissions (UAE)

The aforementioned parameters for filtering the UAE of bearings as well as the power spectral density calculations were repeated for the UAE from a variety of diesel locomotives. The UAE of several components were measured, in an effort to show that UAE exists in the field. It can be seen in Figure 33 through Figure 36 that there noticeable differences in the UAE signature across a variety of locomotives (NS-2547, NS-6780, NS-9769 and NS-2521) and a variety of components. In this case “Notch” refers to the engine speed of the diesel locomotive. Not only can differences in the broadband levels of the spectrums in Figure 33 through Figure 36, but the different components display different tonal frequencies with corresponding amplitudes. These are the types of features in the

power spectrum that can be used to classify the health of the components. However, a good versus bad comparison on a small subset of parts is needed before a neural network can be trained to discriminate between multiple conditions of a single component. It is important to note that each locomotive was a different model and therefore the power spectrum of each locomotive is presented independently.

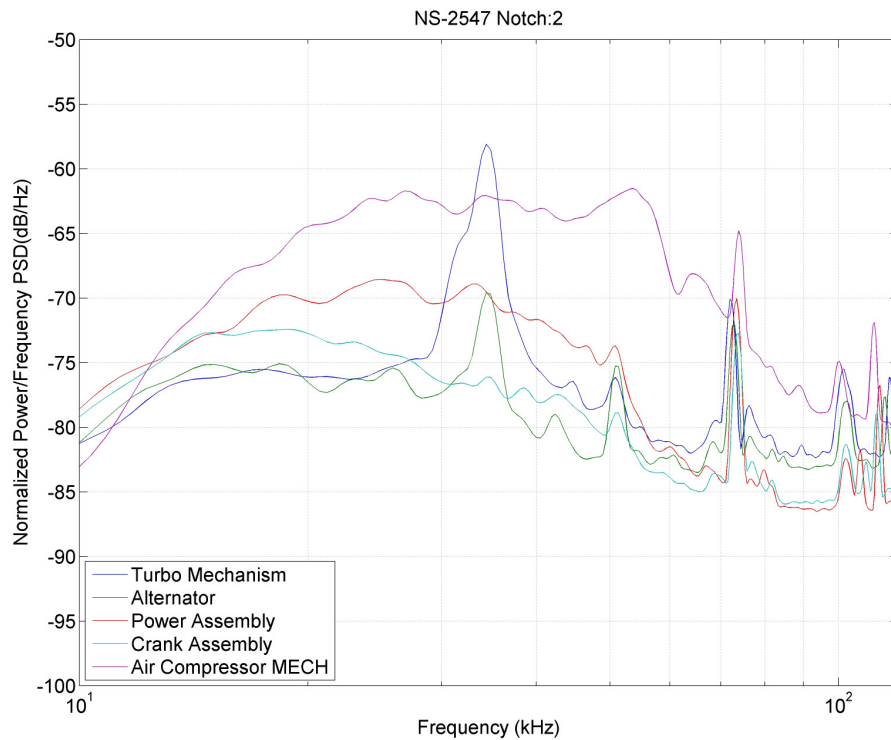


Figure 33. UAE of NS-2547 at Notch 2

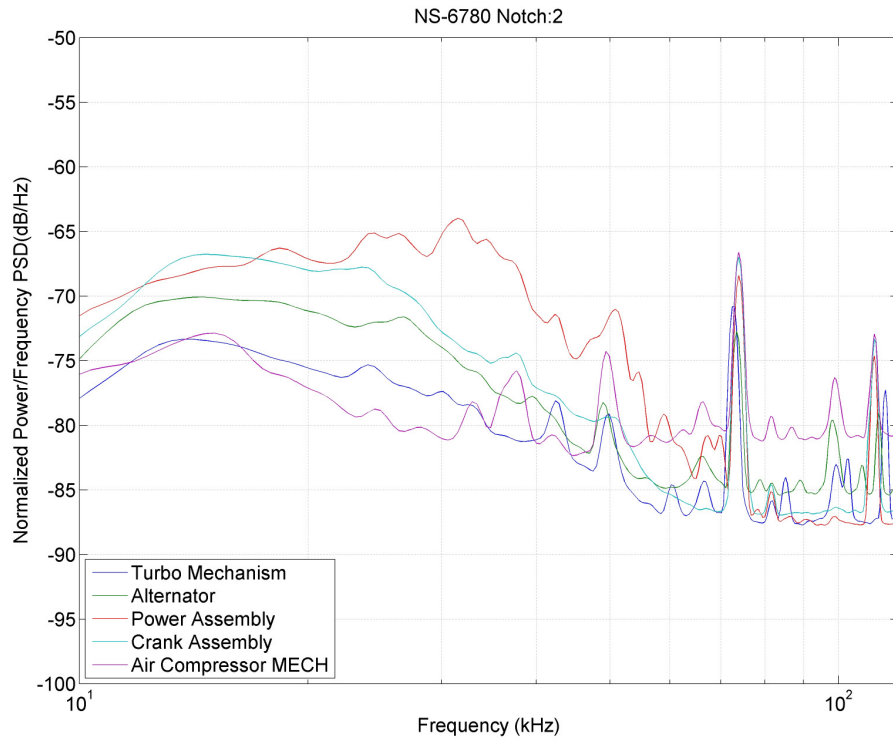


Figure 34. UAE of NS-6780 at Notch 2

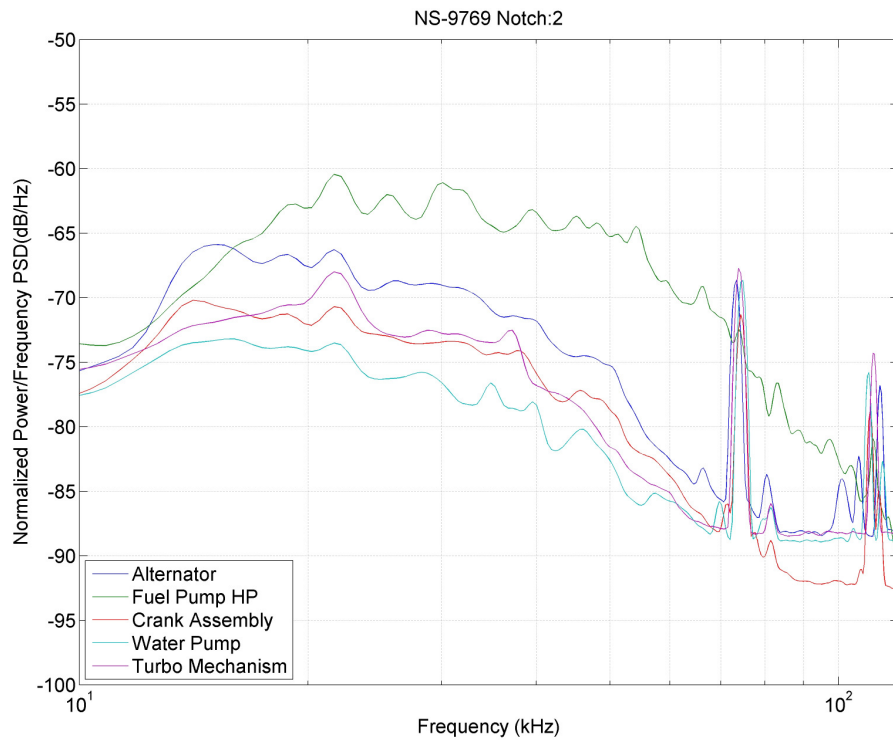


Figure 35. UAE of NS-9769 at Notch 2

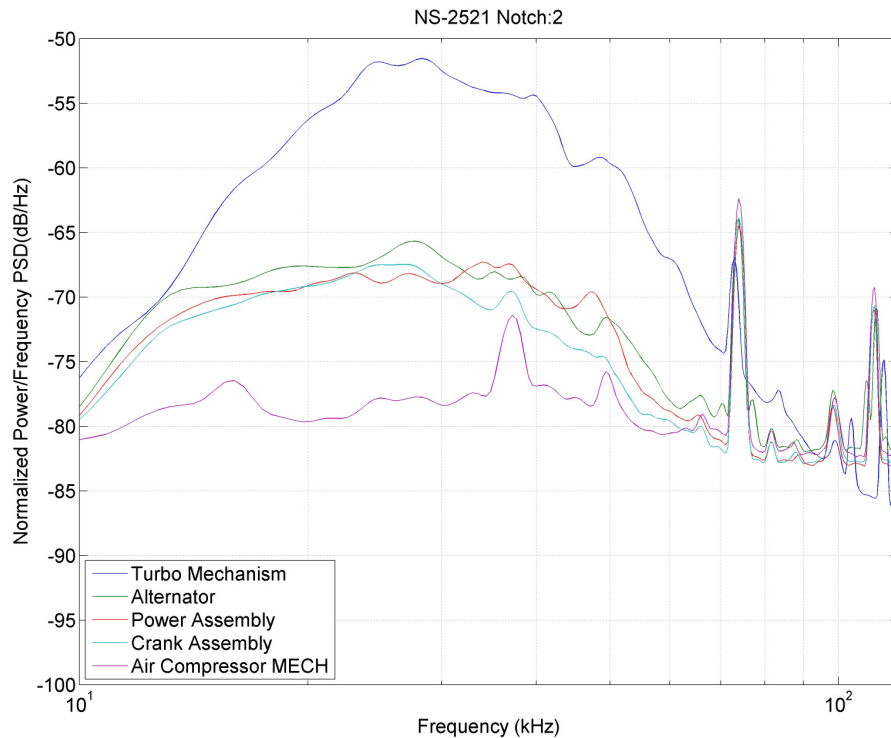


Figure 36. UAE of NS-2521 at Notch 2

Future Field Testing and Development

There are a couple of different directions that could be taken with this project. The first is to start to develop a handheld unit that allows technicians to collect baseline data in a service depot environment, similar to existing technology used in factory environments for isolating air system leaks, electric motor health, etc.

The second approach would be to invest in an array of microphones to be placed in the field, possibly incorporated into existing structures used by the railway community for health monitoring. One such suitable structure would be BeenaVision's Automatic Car Safety Appliance Inspection System ©, which is presented in Figure 37. In this manner, it would be possible to collect acoustic data from a variety of areas and subsystems on both the railcars as well as locomotives.



Figure 37. BeenaVision's Automatic Car Safety Appliance Inspection System

Conclusions

UAE exist both in bearings in a laboratory environment as well as in the field on a variety of components on diesel locomotives. It is possible to pre-process the UAE by means of the power spectrum. Provided the power spectrum shows differences between different operating conditions 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 data set. This method of creating artificial training data is easily extendable to N-dimensions creating a hyper ellipsoid of training points. This method significantly helps in overcoming the curse of dimensionality, common in neural networks. 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. It would also be necessary to validate this method in a high noise environment with a proper experimental data set. In general the idea of using a passive and non-contact measurement, in this case the UAE of the test specimen (bearings), becomes very attractive due to its non-intrusive 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.

References

- [1] Matt Watson, Jeremy Sheldon, Sanket Amin, Hyungdae Lee, and Carl Byington. "A Comprehensive High Frequency Vibration Monitoring System for Incipient Fault Detection and Isolation of Gears, Bearings and Shafts/Couplings in Turbine Engines and Accessories." *ASME Turbo Expo 2007: Power for Land, Sea and Air*. Montreal, CA, 2007. 885-894.
- [2] S. Borguet, O. Leonard. "Coupling principal component analysis and Kalman filtering algorithms for on-line aircraft engine diagnostics." *Control Engineering Practice* 17, 2009: 494-502.
- [3] Michael J. Roemer, Carl S. Byington, Gregory J. Kacprzynski, and George Vachtsevanos. "An Overview of Selected Prognostic Technologies with Application to Engine Health Management." *ASME Turbo Expo 2006: Power for Land, Sea and Air*. Barcelona, Spain, 2006. 707-715.
- [4] M.R. Banan, J. Ghaboussi and R.L. Florum. "Neural Networks in railway Engineering: Acoustic Wayside Fault Detection." *Intelligent Engineering Systems Through Artificial Neural Networks* 4 (1994): 1147-1153.
- [5] G.D. Meegan, H.R. Nelson, M.L. Barlett, G.R. Wilson. "Fault detection in engines through higher order spectral analysis of acoustic signatures." *Journal of Acoustical Society of America* 106(1), July 1999: L1-L6.
- [6] Peng Xu, Andrew K. Chan. "Fast and Robust Neural Network Based Wheel Bearing Fault Detection with Optimal Wavelet Features." *IEEE*, 2002: 2076-80.
- [7] Guobin Ou, Yi Lu Murphey. "Multi-class pattern classification using neural networks." *The Journal of The Pattern Recognition Society (Pattern Recognition)* 40, April 2006: 4-18.
- [8] C.S. Pattichis, A.G. Constantinides. "Medical Imaging with Neural Networks." *IEEE: Neural Networks for Signal Processing*, 1994: 431-440.
- [9] R. Nekovei, Y. Sun. "Classification of Digital Angiograms Using Artificial Neural Networks." *Proc. of the Ann. International Conference of the IEEE Engineering in Medicine and Biology Society* 13 (1991): 1440-1441.
- [10] R.H. Sherman, AS Noetzel. "Image Processing and Pattern Recognition in Ultrasonograms by Back-Propagation." *Neural Networks*, 1990: 593-603.
- [11] M. Wan, B.M. Dawant, RJ Maciumas. "Neural Network Based segmentation of Multi-Modal Medical Images: A Comparative Study." *IEEE Transactions on Medical Imaging* 12 (1993): 531-544.
- [12] Gish, Herbert. "A probabilistic approach to the understanding and training of neural network classifiers." *International Conference on Acoustics, Speech, and Signal Processing, ICASSP-90* 3 (1990): 1361-1364.
- [13] Zuyi Wang, Yue Wang, Jianhua Xuan, Yibin Dong Marina Bakay, Yuanjian Feng, Robert Clarke, Eirc Hoffman. "Optimized multilayer perceptrons for molecular classification and diagnosis using genomic data." *Bioinformatics* 22, no. 6 (2006): 755-761.
- [14] Specht, Raimund. *Avisoft Bioacoustics*. www.avisoft.com (accessed October 9,

- 2009).
- [15] Welch, Peter D. "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 (June 1967): 70-73.
 - [16] Richard Duda, Peter Hart, David Stork. *Pattern Classification*. Second Edition. John Wiley & Sons, 2001.
 - [17] Howard Demuth, Mark Beale, and Martin Hagan. *Neural Network Toolbox 6 Users Guide*. The MathWorks Inc, 1992-2009.
 - [18] Martin T. Hagan, Mohammad B. Menhaj. "Training Feedforward Networks with the Marquardt Algorithm." *IEEE TRANSACTIONS ON NEURAL NETWORKS* VOL. 5 , no. NO. 6 (1994).
 - [19] Richard Duda, Peter Hart, David G Stork. *Pattern Classification*. 2nd. New York : John Wiley & Sons Inc., 2001.