UNIVERSITY OF CALIFORNIA, SAN DIEGO

A Performance Comparison of Features Used in Vibration-Based Health Monitoring of a Complex Mechanism

A Thesis submitted in partial satisfaction of the
requirements for the degree Master of Science

in

Engineering Sciences (Mechanical Engineering)

by

Luke Mathieson Zsiga

Committee in charge:

Professor Raymond de Callafon, Chair
Professor Bob Bitmead
Professor Michael Todd

2013

The Thesis of Luke Mathieson Zsiga is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

Chair

University of California, San Diego

2013

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ROC	Receiver Operating Characteristic
SHM	Structural Health Monitoring
SNR	Signal to Noise Ratio
TPR	True Positive Rate
FPR	False Positive Rate
ETFE	Empirical Transfer Function Estimate
DAQ	Data Acquisition
NDE	Non Destructive Evaluation
AR	Autoregressive
AUC	Area Under the Curve

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge several individuals who have made my work on this thesis possible. First and foremost among them, I acknowledge my advisors Mike Todd and Raymond de Callafon who provided the opportunity to pursue this topic and guided me from beginning to end. I am indebted to Professor Todd for his patient tutoring, provision of the means for this project, willingness to travel to meet with my project leads, and most importantly for teaching me to expand my critical thinking process. I have also considered it an honor to have worked with and studied under Professor de Callafon, who has taught me to apply coursework principles to practice.

I am indebted to my Structural Health Monitoring colleagues Dustin Harvey, Scott Ouellette, Colin Haynes, Luke Robinson, Zhu Mao, Richard Do and Eric Kjolsing. Without their support and assistance I would still be at square one. The program SeDaq created by lab mate Dustin Harvey was very useful in acquiring data as well as Dustin's guidance. My lab mates made our lab both an excellent learning environment and an enjoyable workplace, and I have appreciated the friendships we have all created. Last I would like to thank my wife Rebekah Zsiga for encouraging and supporting me to pursue higher education and patiently enduring all the nights that I came home well after dark.

VITA

2006	Airframe & Powerplant License, Tulsa Technology Center
2007-2008	Aircraft Mechanic, Oklahoma State University
2009	Aircraft Mechanic, Summer Institute of Linguistics
2011	Bachelor of Science, Mechanical Engineering, Oklahoma State University
2011	Bachelor of Science, Aerospace Engineering, Oklahoma State University
2011-Present	Engineer, Sandia National Laboratories
2013	Master of Science, Mechanical Engineering, University of California, San Diego

ABSTRACT OF THE THESIS

A Performance Comparison of Features Used in Vibration-Based Health Monitoring of a Complex Mechanism

by

Luke Mathieson Zsiga

Master of Science in Engineering Sciences (Mechanical Engineering)

University of California, San Diego, 2013

Professor Raymond De Callafon, Chair

The purpose of this study is to use vibration response data to inspect a certain safety mechanism owned by the Department of Energy. These mechanisms are required to have a long accurately predicted service life. However, these mechanisms are permanently enclosed in metal and installed in a higher level assembly hindering inspection.

This thesis considers observing the vibrations output from the mechanism by applying the structural health monitoring process. Features that correlated with the change in health of the machine (basic statistics and model parameters) were extracted from the measured vibrations and compared for detection performance. Specifically, failure modes were

introduced into a prototype of the device, and supervised learning (involving training data) was applied to assess the features' ability for detection and localization.

Performance results were quantified by using the Mahalanobis distance measure to determine how much the features have changed. This distance was used in a receiver operating characteristic curve to compare damage detection results. It was discovered that the use of this structural health monitoring process enabled accurate discrimination of the mechanism as being healthy or unhealthy. A side study of sensor location vs. damage detection indicated which locations sensors were most effective in measuring damage.

The principal conclusion was that the structural health monitoring process should be implemented by the department of energy on their mechanism. The benefits include detecting damage before it is severe enough to cause a failure, the ability to locate damage, and a possible reduction of unnecessary inspection labor.

Chapter 1
Structural Health Monitoring

1.1. Introduction

Structural health monitoring (SHM), in one of its major intended uses, is the autonomous (or increasingly autonomous) informing of maintenance inspections. SHM utilizes a variety of sensing technologies to generate in-situ data from which features are extracted through data processing to analyze the health of a structure or mechanical system [1]. Traditional maintenance methods usually involve visual, time-based inspections and maintenance [2]. Implementing SHM can provide performance enhancement, economic benefit, or life-safety benefit over these traditional methods, as it may be run continuously in-situ, often even in a hostile environment, and can empower condition-based maintenance inspection. These benefits include possibly providing a substantially more accurate record of useful life (when coupled with predictive modeling), reducing the amount of labor associated with inspections and part replacement, optimizing maintenance costs and actions, and increasing life safety in applications where the costs of failure are substantial.

A four-step statistical pattern recognition paradigm has become the recognized general methodology in the field to implement SHM [3,4]. The major steps in this paradigm, and some of the primary issues/challenges with each step, include:

1. Operational evaluation
    a. What is the "business case" for using SHM?
    b. What damage mode(s) is/are most relevant to the given application, and what level of detection, localization, and/or quantification is required for those damage mode(s)?
    c. What are the expected environmental and operational conditions that constrain the performance of the SHM system?
    d. What other constraints are present (economic, technical, legal, regulatory) that limit the use/deployment of the SHM system?
2. Data acquisition
    a. Sensing (and possibly actuation) selection and trade-offs
    b. Deployment issues
    c. Power consumption/requirements
3. Feature extraction
    a. What features resulting from signal processing of acquired data most correlate to the targeted failure mode(s)?
    b. Data processing limitations
    c. Length and time scales that features must be computed for appropriate decision-making
    d. Very integrated with the data acquisition process (the two are highly coupled steps)

4. Feature discrimination (decision-making)
    a. Supervised vs. unsupervised learning (classification) methods
    b. Establishment of appropriate thresholds for normal operation (decision boundaries)
    c. Data normalization (nulling feature changes due to non-damage events)

There are no sensors that can directly measure damage: it is the proper selection and use of sensors combined with intelligent feature extraction and classification that inform the kind of decision-making required of SHM that are capable of detecting damage [5]. Although similar and often comparable, non-destructive evaluation (NDE) is a field separate from SHM. SHM is a damage detection process that's operated in-situ while NDE's damage detection process is more of a tool used when a part has been taken out of service. Furthermore, NDE typically involves a lot more localized monitoring, where there is a priori knowledge of damage-prone areas [6].

1.2. Structural Health Monitoring Process

1.2.1. Operational Evaluation

The operational evaluation stage is a time to plan the SHM process. This part of the process is very important because there are no predefined solutions for SHM and each case is unique. The first step in planning the SHM process is to evaluate what justifies its use. These justifications can vary, but typical justifications include saving money, increasing safety, and enabling inspection to occur in previously un-inspectable areas. The SHM system can then be designed to maximize the benefits of the particular justification. Relevant damage types are

identified and examined. Their importance to the given application and the required level of detection, localization, and or quantification is studied to dictate the types of sensing and raw data that will be most useful for feature extraction [7]. If damage is detected what actions can and should be taken, should the power be cut off from the device? Or should an email be sent to a maintenance technician? Last constraints like environmental and operational conditions that will affect equipment selection and operation should be carefully reviewed [8]. Many constraints could be present to name a few; economic, technical, legal, and regulatory.

1.2.2. Data Acquisition

All required hardware for a SHM system is carefully reviewed and chosen. The required hardware includes sensors, transmission hardware (cables or wireless equipment), power supplies, weather proof enclosures, pre-designed sensor mounts, and a central computing system. If necessary, equipment for actuation/indication is considered as well. The central computing system is responsible for operating sensors, acquiring their data, and processing this data for a damage detection report. When selecting equipment, important performance specifications to consider are frequency range, sensitivity, and sampling rate. It is important to know what frequencies are excited when the machine operates or is externally excited. This operating frequency range can be determined by modeling, trial and error, or from previous knowledge [9,10]. In some cases power for the system is not readily available and transmitting power to the system would be expensive and present additional maintenance requirements. This has created an exciting field of research for SHM that includes ideas such as harvesting electrical current produced by corrosion in piers [11] and remote transmitting of power / data by unmanned vehicles [12].

1.2.3. Feature Extraction

Data storage and processing time can become a significant hindrance due to typical sampling frequencies in the tens of thousands of hertz, and some even in the megahertz range. Data must be condensed so that a single decision can be made on the health of the component and to accommodate storage requirements. Condensing is accomplished by extracting features that describe the data. A feature set is the final form of the reduced raw data that will ultimately be used to guide the detector's decision making [7]. These features can be parameters from a time series model created from the measured data or a single statistic pulled from the data, like the average noise level. The feature selection process is a critical component of SHM. The features chosen should contain as much pertinent information about the system as possible with the fewest features. However, compressing data into a small number of features has the inherent tradeoff of losing information.

To help determine the appropriate features three questions should be visited: "What features capture relevant physical properties of the structural systems behavior that indicate the targeted failure mode?" [7], "What is the most efficient way to model or condense the data?" and "What limitations are there on feature extraction?" The feature extraction decision is highly coupled with that of data acquisition. The desired feature dictates what types of sensors are chosen while the availability/limitations of the selected sensors can limit how effective the features are. Limitations for sensors include the measuring range, sampling frequency, and other properties as described above in section 1.2.2. The selected features and their discrimination process determine the required memory for data storage. Features can be processed in real time and a decision can be made immediately eliminating the requirement for storage, or the discrimination process can rely on stored data requiring the features to be saved. Once the feature is selected its variability must be statistically modeled to help in

differentiating change due to standard variance and change due to damage. This process and more is described below in the feature discrimination section.

1.2.4. Feature Discrimination (Decision Making)

For an automated decision on structural health to occur, a threshold must be set that discriminates the acquired feature or features as coming from a damaged unit or an un-damaged unit. This threshold or thresholds are determined with statistical modeling of the feature's variability. If available, features extracted from damaged unit's data can also be statistically modeled to set additional thresholds that indicate what type of damage has occurred in the machine and the severity of the damage. This is called supervised learning, if information on how damage changes features is not used then the only thing that can be indicated is a change from the un-damaged state which is referred to as unsupervised learning [13]. These thresholds are drawn from the statistical model of the feature's known variance and vary from thresholds that minimize false positive indications to thresholds that minimize false negative indications. A specific threshold is picked using the justifications determined in step one, operational evaluation. Once the features have been extracted and compared to the thresholds the health of the machine can then be indicated to the user. If desired, and the severity of damage is high enough to require, additional automated sequences can be built in that terminate operation of the machine or prohibit use of the structure.

1.3. Example

A classic SHM example is condition monitoring of bearings. This is a well-established SHM application [14-17] that industries have adopted fairly consistently. The four step SHM paradigm for this example follows. A diagram of a standard ball bearing is shown in Figure 1 (below).

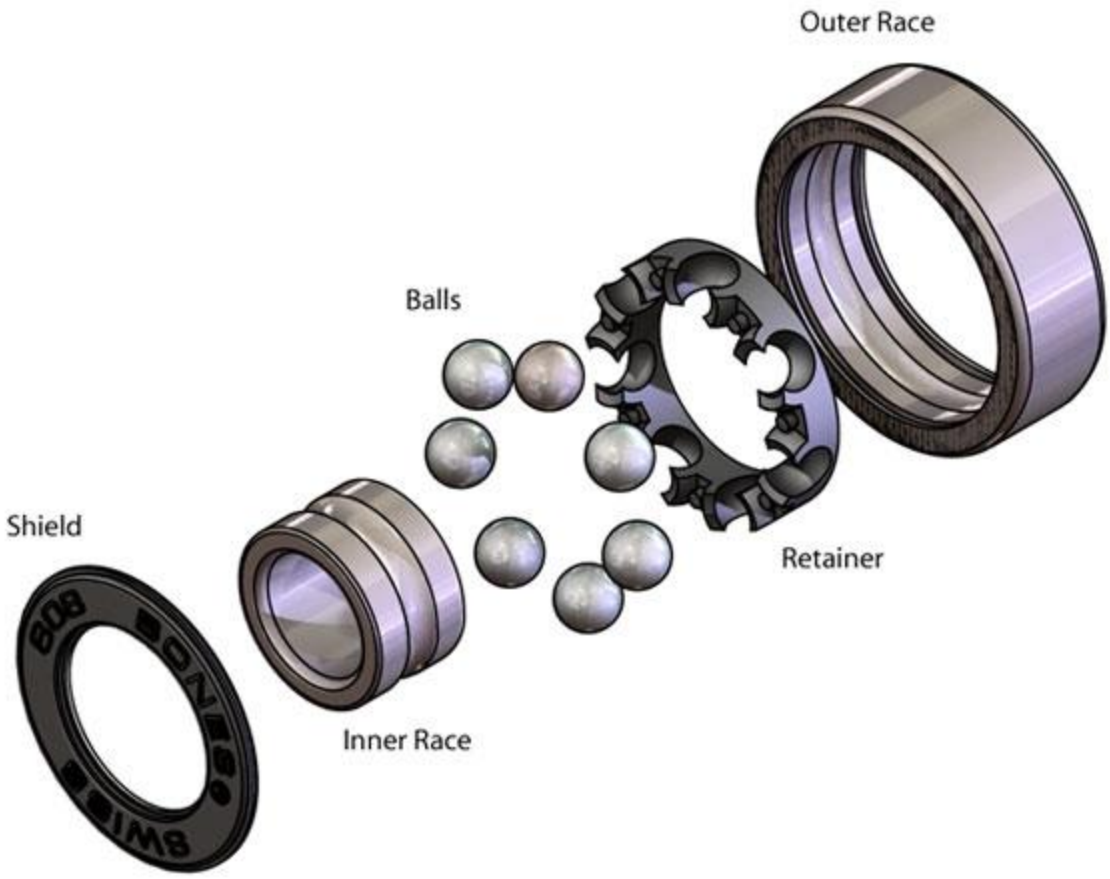

Figure 1: Ball bearing diagram [18]

1.3.1. Operational Evaluation

Bearings are vital components that are widely used in machinery. Certain applications for bearings require an accurate indication of its health to avoid long shut down times due to un-predicted failures. The loss of operating time and money associated with the shut down time provides the justification for implementing a SHM system. Reducing costs while providing an accurate indication of health are the business needs to be maximized. Three major categories of failures for bearings are: wear, abnormalities (cracks, corrosion pits or indentations), and melt down. These relevant damage modes need to be detected before they cause failure.

1.3.2. Data Acquisition

The damage modes wear and abnormalities both produce excessive vibrations in the rotating plane of the bearing prior to and after failure. Two accelerometers are placed in this same plane 90 degrees apart on a solid structure as close to the bearing as allowable. This location receives the most energy from the bearing's vibrations and is most sensitive to the first two damage modes. The accelerometers and their accompanying data acquisition unit must be selected to handle sampling frequencies that are twice that of the maximum frequency that could be produced by the bearing (A function of maximum rotational speed, the number of ball bearings and the geometry of the bearing). Meltdown of a bearing is preceded by an increase in temperature. A temperature sensor is placed as close as possible to the bearing to monitor its temperature.

1.3.3. Feature Extraction

Frequency analysis is performed on the measured acceleration data and the frequency content is normalized to the rotating speed of the bearing's race that is being driven (in revolutions per minute [rpm]). Any peaks in the frequency content are extracted, and their respective normalized frequency is used as features to detect and characterize any abnormalities. The noise level of the measured acceleration data while the machine is operating is recorded as a feature. This operational noise level will increase with wear.

1.3.4. Feature Discrimination (Decision Making)

Supervised classification methods are possible with bearings as information regarding the bearings is readily available, this process is described below. The normalized frequency features indicate the location of the abnormality while the noise level indicates the severity of the wear. If there is damage present in parts from the bearing, a spike will show up in the

frequency content at a specific normalized frequency associated with that part. The specific normalized frequency for each part in the bearing can be calculated based off the geometric dimensions of the bearing and the number of ball bearings. For instance, if there is an indentation in the outer race, and the geometric dimensions of the bearing determine that the retainer spins at $1/10^{th}$ the speed of the inner race, and there are 7 balls in the bearing, then there will be 7 accelerations caused by each ball hitting the indentation for every 10 rotations of the inner race. This will show up as a spike in the normalized frequency content at 7/10/ (inner race rpm) indicating a damaged outer race. As wear occurs the fit of parts in the bearing are no longer tight and thus increase the noise level. The threshold indicating the amount of wear can be set as a percentage increase from baseline noise level or more accurately it can be set by measuring the noise level of a damaged bearing of the same model. The threshold for maximum allowable temperature the bearing can reach before a meltdown occurs is determined by specifications set by the bearing manufacturer. The bearing and temperature sensor's distance and medium separating them are important parameters that will have to be accounted for as they delay the true indication of a bearings temperature.

1.4. Scope of Thesis

The Department of Energy has the lifecycle responsibility for safety systems in weapon applications. The safety systems must meet stringent reliability requirements when used in the intended mode, but preclude unintended energy from reaching critical system features. In addition, it must provide this type of performance for multiple decades without being operated frequently or even infrequently. The safety systems consist of numerous mechanisms. The safety mechanism considered in this thesis serves a simple purpose of closing a switch and does so "safely" by requiring a set of unique mechanical actions to be

performed before the switch closes. This safety mechanism is a unique, complex, precision machine and is housed in a controlled environment resulting in a more predictable service life that is hard to quantify. Predicting service life is difficult for a mechanism that also requires a long service life because reliability may decrease as the mechanism ages. The mechanism is sealed in a hermitic metal housing making visual inspection impossible and maintenance very difficult thus furthering the difficulties with predicting the service life. Structural Health Monitoring (SHM) is ideal for use in this scenario as it will allow inspection to occur inside the housing by monitoring the vibrations generated by the mechanical actions of the device. SHM provides a method of inspection that is non-destructive to the unit, allows important information to be extracted about its health, and does not compromise performance as SHM is non-invasive. Since the safety mechanism is mounted in a higher level assembly, costs can be saved related to disassembly and destruction of the higher level assembly as it will not need to be disassembled for the SHM process to occur.

Applying the SHM process to this safety mechanism from the DOE has been thoroughly researched and is laid out in this thesis. Design of the machine and the operational evaluation step of the SHM paradigm are discussed in chapter two. Data acquisition is outlined in chapter three while feature extraction and discrimination is outlined in the following chapter.

Chapter 2
Design and Operational Evaluation

2.1. Test-bed design

Due to the limited availability of the target mechanism (which is a mechanical safety mechanism) owned by the Department of Energy and its high cost, a surrogate test-bed was designed that replicates the types of vibrations and other primary performance behaviors that occur in the actual mechanism. The SHM process was completed on this replacement to verify whether or not SHM is viable for this application. This project focused solely on the health of the mechanical portion of the safety mechanism, which has a primary function of closing a switch in a way that minimizes an accidental or un-planned closure. This is accomplished by using a mechanical combination lock that must be un-locked before the switch can be closed. This combination lock is designed so that when an incorrect combination is entered the device will physically lock up and become inoperable until it is disassembled and repaired. Requiring a combination to be input, prior to switch closure, reduces the chance that an event other than a planned unlock sequence will close the switch. For instance, if a faulty electronic component outputs random signals to the device, the device will lock up if the input signal no longer

matches the un-lock combination. If a surge of current strong enough to jump electrical contacts occurs the device will not be able to jump the contacts until the combination is complete and the switch is physically aligned. The design for the test-bed used in this project is described in detail below.

2.1.1. Prototype

The safety mechanism was housed in a sealed environment enclosed in thick metal. This was replicated in design by placing the device in a machined metal housing that consisted of a top and bottom half. Parts designed to be the combination lock portion of the electrical switch were placed inside this casing. The component located in the center of the device will be a rotating disc that, when rotated to the proper position, will align a portion of the wheel that closes the switch. The initial prototype for this disc is shown below in Figure 2.

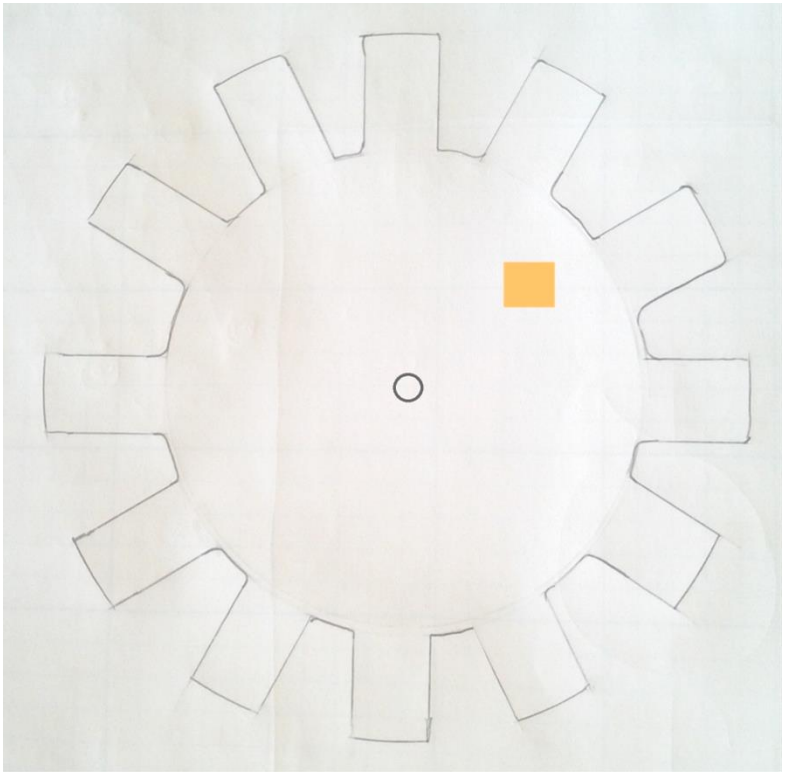

Figure 2: Main disc prototype with electrical contact

The main disc is rotated by two swinging arms that contact spokes on the outer edge of the disc. Grooves cut into certain spokes of the main disc create a combination of spokes that either lock the machine in place or advance the main wheel. The main disc has to be advanced in the correct order by the two swinging arms and if it is advanced incorrectly, e.g., arm 1 advances the main disc with a spoke that has a groove cut into it, the swinging arm lodges in the cut groove. The swinging arm is held in place, and locked, by a spring attached to the swinging arm. This technique is pictured below in Figure 3.

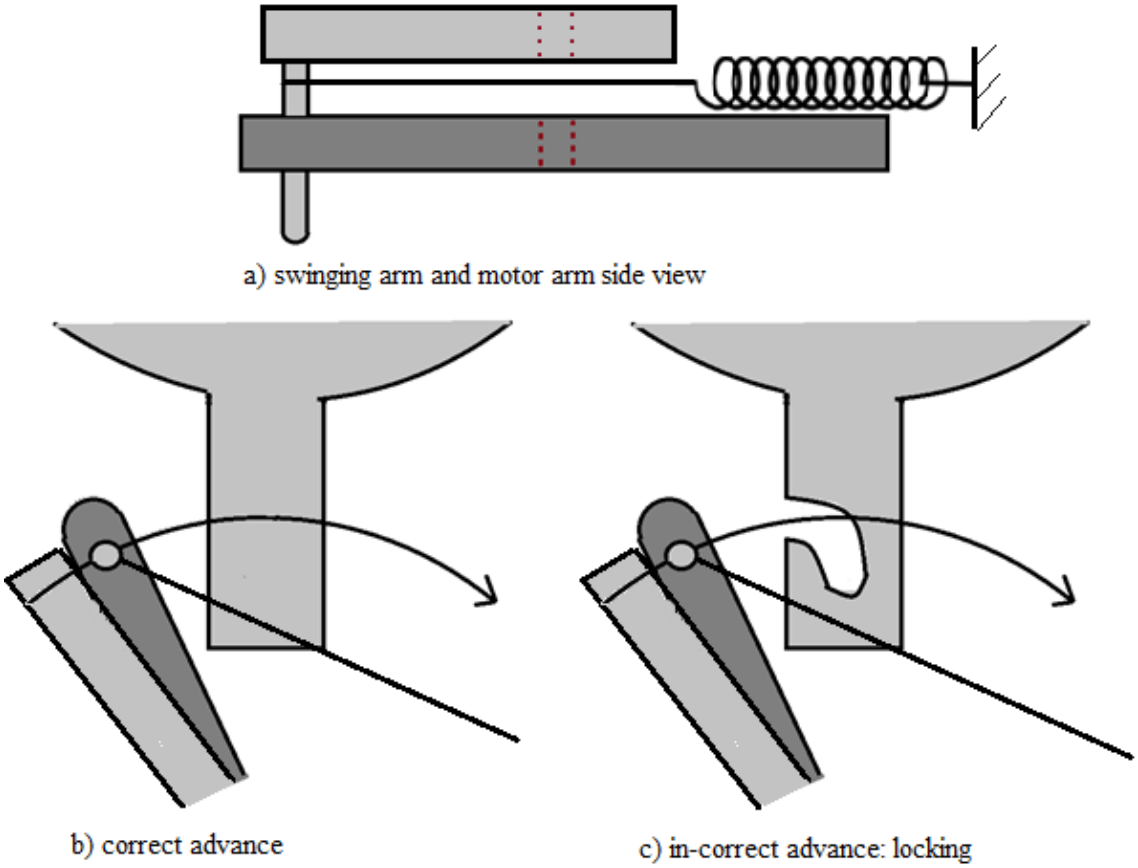

Figure 3: Main wheel locking design

The swinging arms pictured above in Figure 2 have 360 degrees of movement. The arm is powered by both a spring and an arm attached to a motor. The motor rotates the arm by

pushing the upper peg on the swinging arm to rotate it, stretching the spring until the spring powers the arm for the other half of the rotation. By completing a full revolution, the arm never has to move backwards and thus can only advance the disc forward, unless the motor turns the arm in the opposite direction. The motor arm, swinging arm and main disc working together are pictured below in Figure 4.

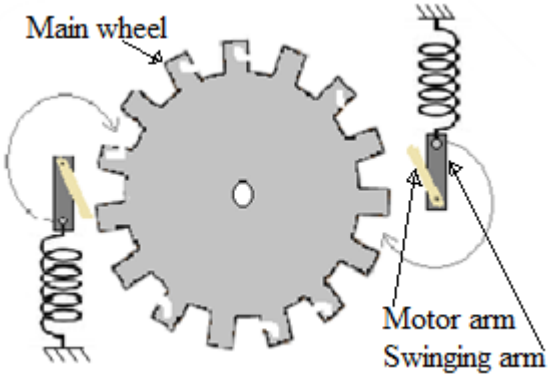

Figure 4: Motor arms, swinging arms and main disc

The swinging arms and the motor arms are pictured above. The motors in figure 3 are located directly above the two motor arms, but are not in the picture for visibility purposes. If a more complex advancement combination is desired a second main disc could be added stacked above the first and the two swinging arms could be offset to line up on opposite main discs.

To help control the dynamics of the moving components, dimples are designed to be machined onto the main disc in a circular pattern with a radial spacing identical to the spokes. A small ball is held against the main disc with a spring in the same location as the dimples when a dimple rotates by the ball lodges in the dimple holding the main disc in place until it is advanced again. This prevents the disc from over rotating and helps limit the main discs movement to pre-specified increments. The ball and spring's contact force needs to be

sufficient to arrest the main wheel's motion, but not so high that it prevents the main code wheel from being moved.

The scale of the device was considered carefully. It would be difficult to construct the safety mechanism on the same scale as that of DOE's safety mechanisms due to their small size. A medium scale device is easiest to work with and machine, but the main code wheel inertia scales with its radius to the fourth power. This impacts the aforementioned stopping mechanism (ball and spring) significantly but more importantly it limits the speed with which the device can be operated. The Department of Energy's mechanism operates at a speed that was not achievable within a reasonable budget and timeframe, but the designed device still needed to be as close as reasonably possible, within manufacturing and ease-of-use constraints. The size and speed of the machine were duplicated as closely as possible however they were not the most important aspects of the machine to duplicate. A focus was placed on designing a machine that replicates the same types of vibrations by designing it to perform the same function with similar parts.

2.1.2. Final design

To facilitate a smaller design size, the main wheel was changed from a design of spokes to notches cut in the main wheel. The swinging arms were placed at a distance away from the main wheel that caused them to collide with the main wheel and prevent advancing unless the portion of the main wheel with a notch cut in it was aligned. This new design serves the same purpose as the spoke design but allows for better machinability and a smaller size. Only the health of the safety mechanism's mechanical portions are of interest, so the closure of the switch was disregarded in the design of the main wheel and any electrical malfunctions were not considered as potential damage types. Aluminum was initially chosen as the material

for the wheel to keep inertia down but had to be changed to a harder material--stainless steel--to avoid issues with the notches wearing (not a failure mode of interest).

There were no bearings available that were small enough to fit the close spacing of the swinging arm, spring, and motor arm. Custom bearings were designed for the swinging arm and the main wheel. Bronze was chosen as the material for these bearings because of its availability and low coefficient of friction against steel. The bearing was designed to have a flush surface to allow the spring to move across the top of it.

Aluminum was chosen as the material for the case of the device for its machinability, reducing the cost to make it. The housing is designed with all flat surfaces on the exterior and with a 10 mm thick shell of aluminum around all parts to allow holes to be drilled and taped anywhere on the housing for sensor mounting locations. Extra spring pin mounting locations are designed in the housing to allow adjustment of the spring's pre load and direction of force. Adjusting the spring's location gave the user more control of the machines operating speed. (extra mounting locations can be seen below in Figure 5)

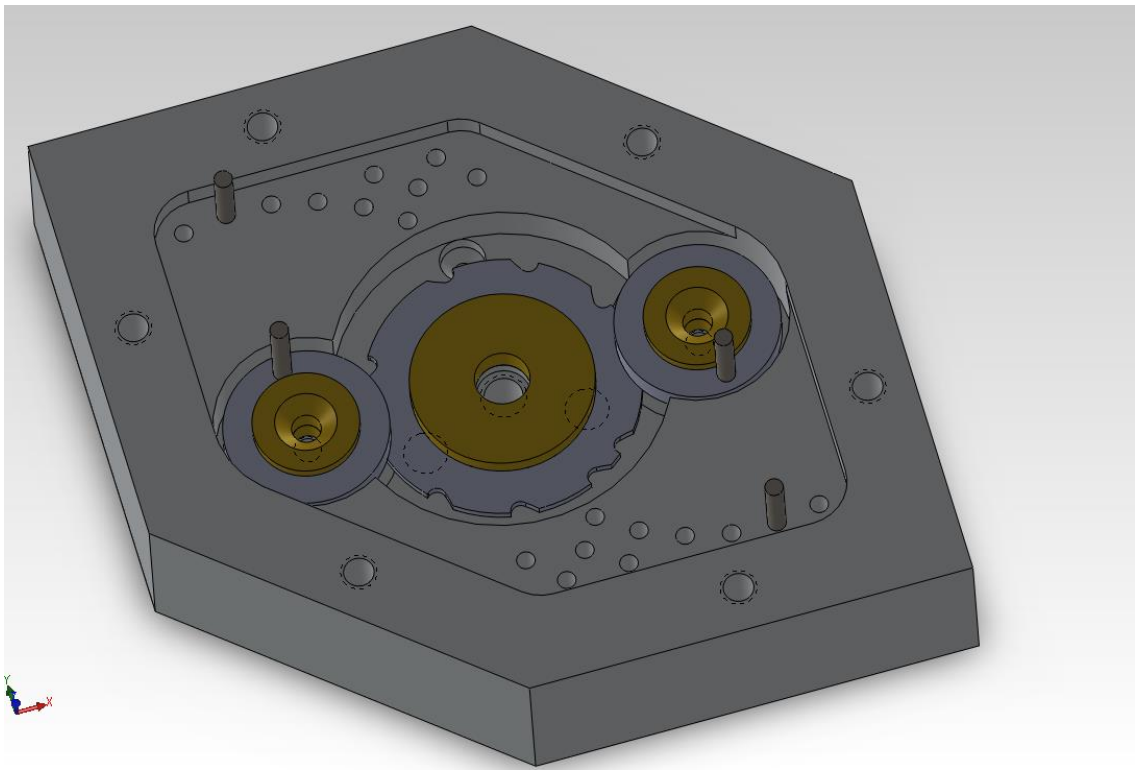

Figure 5: Bottom half with parts

To vary the operating speed of the machine and to insure that the main code wheel does not over rotate, a ball and spring plunger is installed in line with a radial set of holes drilled in the main wheel. The smallest ball and plunger available with the lightest spring force was chosen from the vendor (Figure 6). Holes drilled in the main wheel were chosen over dimples to save money in machining costs. The size of the holes in the main code wheel where chosen based off the size of the ball and its available travel length. Four ball and spring plunger locations were drilled in the housing so that there is an option of installing one to four ball and spring plungers. Adding extra ball and spring plungers multiplies the contact force on the main wheel helping with the adjustment associated with changing the machine's operating speed.

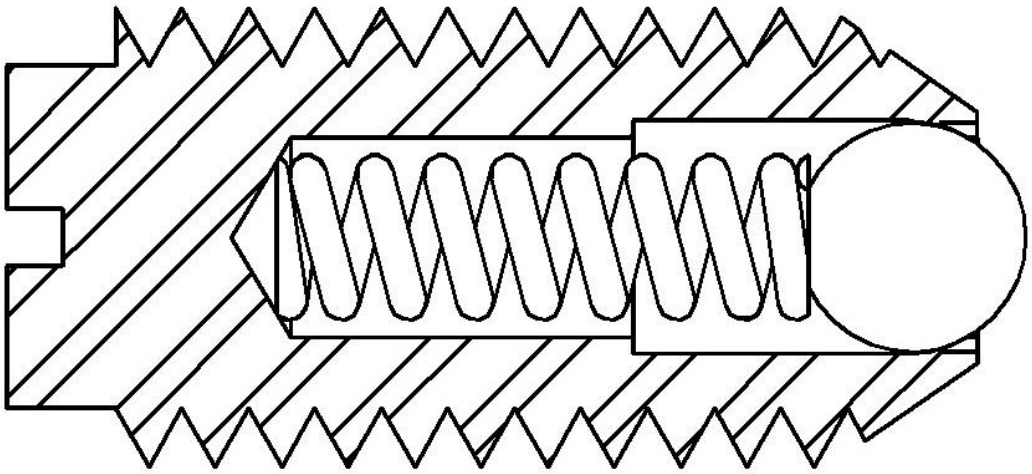

Figure 6: Ball and spring plunger [19]

The device's parts were designed to have the smallest dimensions possible while still being reasonably easy to manufacture. The swinging arm's radius was minimized based on the diameter of the bearing and the size of the pin in the swinging arm. The radius of the main wheel was minimized based on the width of the notches that had to be cut in the wheel. This width was equal to the diameter of the pin that would ride in the notch. Ample space was left between each notch so the impact of the swinging arm would not break the wheel. Once the radius of the swinging arm and main code wheel were determined, their distance apart was calculated based on the desired rotation of the main code wheel for each event. The device was to complete a full revolution in one cycle, with 20 events in one cycle, resulting in the main wheel rotating 18 degrees each time the swinging arm moved it. The distance of the swinging arm to the main code wheel and the depth of the cut were calculated based on the required 18 degree rotation; a diagram detailing this calculation is shown below in Figure 7.

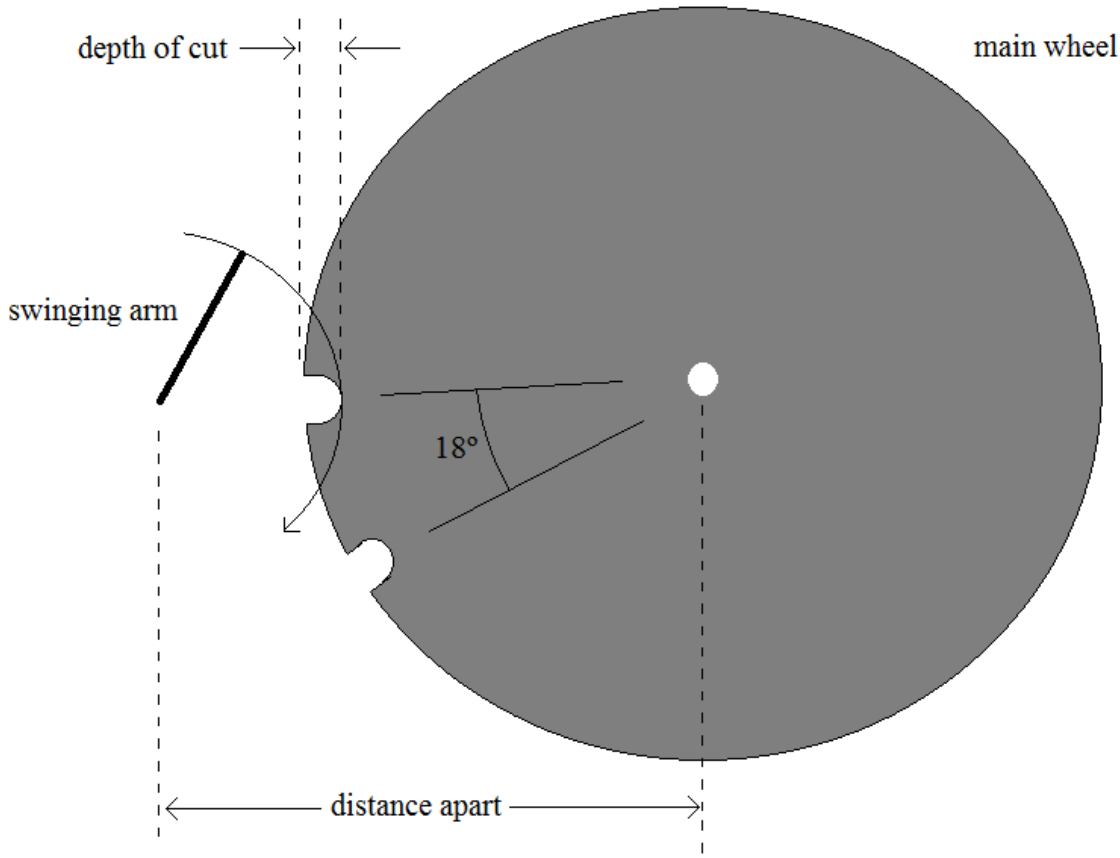

Figure 7: Swinging arm distance & notch cut diagram

2.1.2.1. Electronics Selection

Stepper motors were chosen to power the device for their accurate control and ease of use [20]. Stepper motors are widely available with preprogrammed controllers that are suitable for this application. The use of a stepper motor also eliminated the need for an encoder to be installed and programed. Stepper motors allow precise movements; in this particular case movement accuracy of up to .1125 degrees was attained. AllMotion's EZ 4AXIS controller was used along with its recommended power supply to control the stepper motors. The specific stepper motors were chosen based on the required torque to operate the device. The range of required torques was calculated based on the geometry of the various spring mounting locations and spring constants of springs that were available for sale. The force from

the spring applied to the swinging arm was calculated in excel for the whole rotation of the swinging arm.

After calculating the required torque, there were two choices of motor size to choose from: a Nema size eight and a Nema size eleven. The higher rated Nema eight stepper motors produced less than double the required torque when operating at full speed. Nema eleven stepper motors produced upwards of twelve times the required torque when operating at full speed. The smaller motors were ordered first, since minimizing the size of the machine was a design goal, and they worked but only up to 60% of the machine's top speed, from 60% of top speed and up, the required torque is too much for the motors as the resistive force of dampening from the bearing's lubrication increased with speed. The Nema size eleven motors were ordered next and were able to operate the machine at full speed.

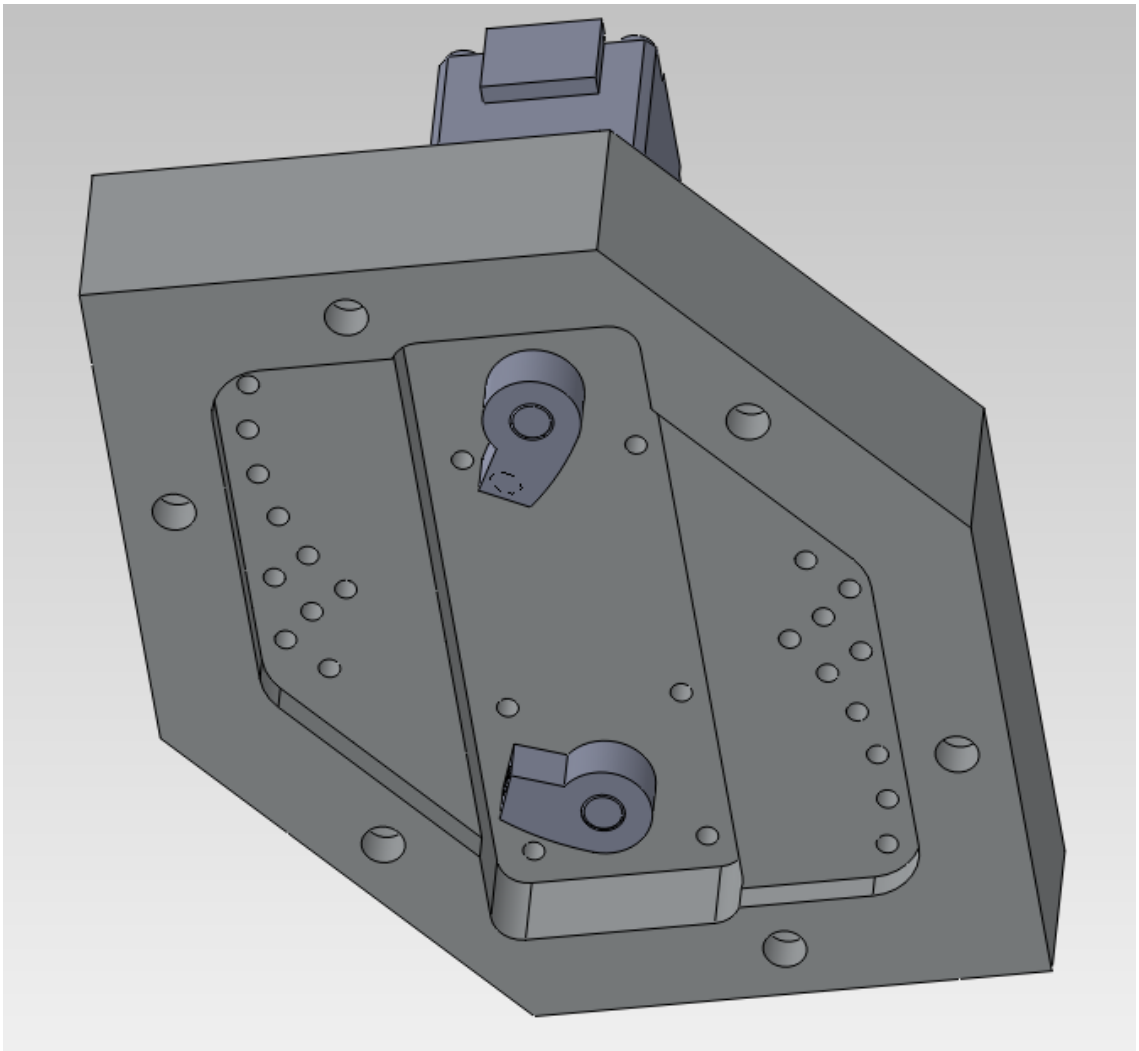

Figure 8: Stepper motors with attached motor arms

Next, the motor controller was programed with the correct sequence for the machine to unlock. This was accomplished following AllMotion's programming guide [21]; a copy of the code is listed in Appendix A.2. In addition to the program that operated the device, during the unlock sequence there is a program run before each test to ensure the device begins in the same configuration each time. This program is run before every test, and the motors turn at low power until they hit the swinging arm's pin where they stop due to the low power setting, ensuring that the motors always start at the same position.

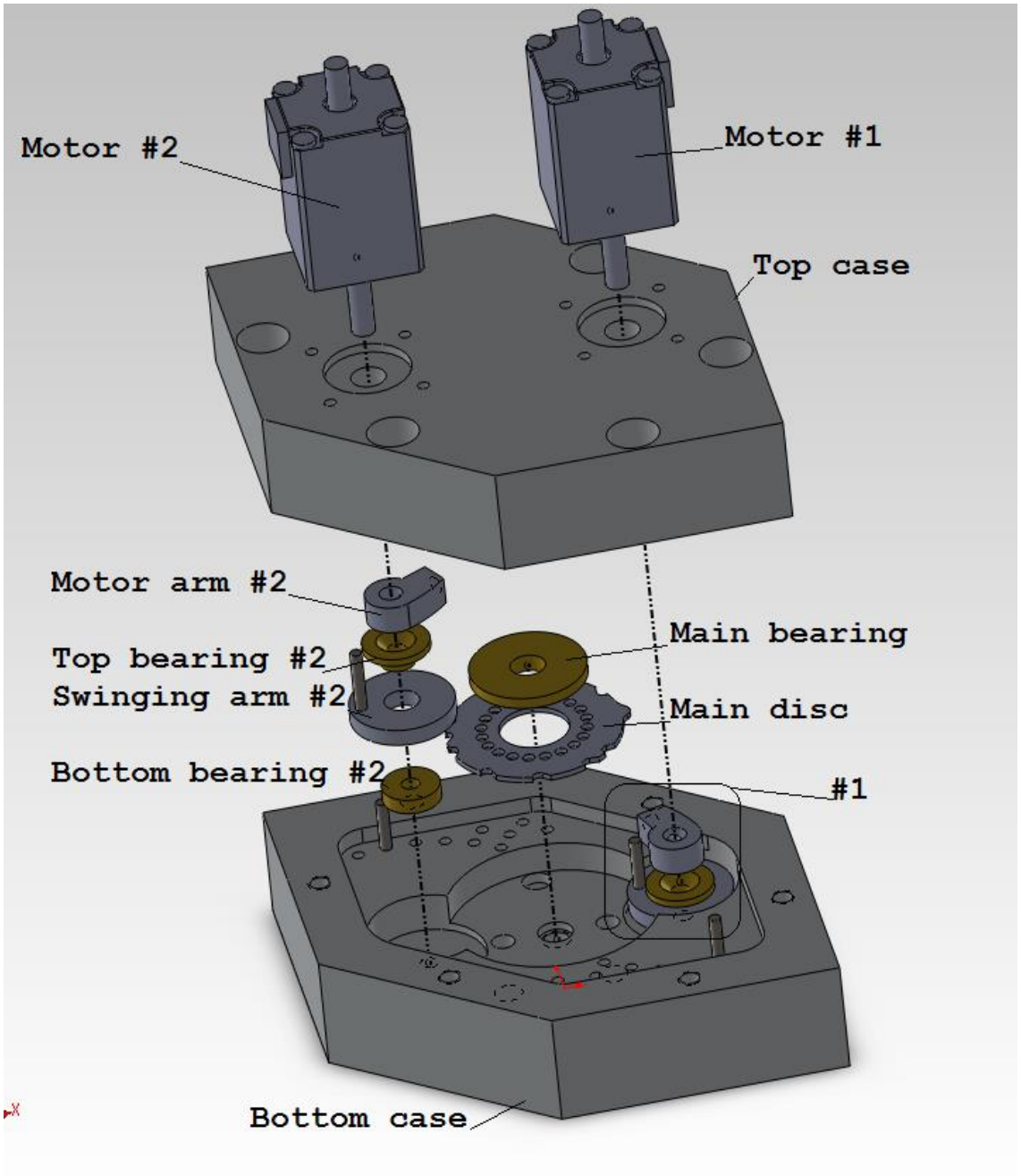

Figure 9: Final design

2.1.3. Test Stand

A base was designed to mount the device to during testing. This base was designed to leave as much of the device's surface open as possible to refrain from restricting any sensor mounting locations. One goal of the test stand design is to create the base to be ten times the mass of the device. The test stand is placed on an optics table for testing. The test stand and final testing configuration can be seen below in Figure 10.

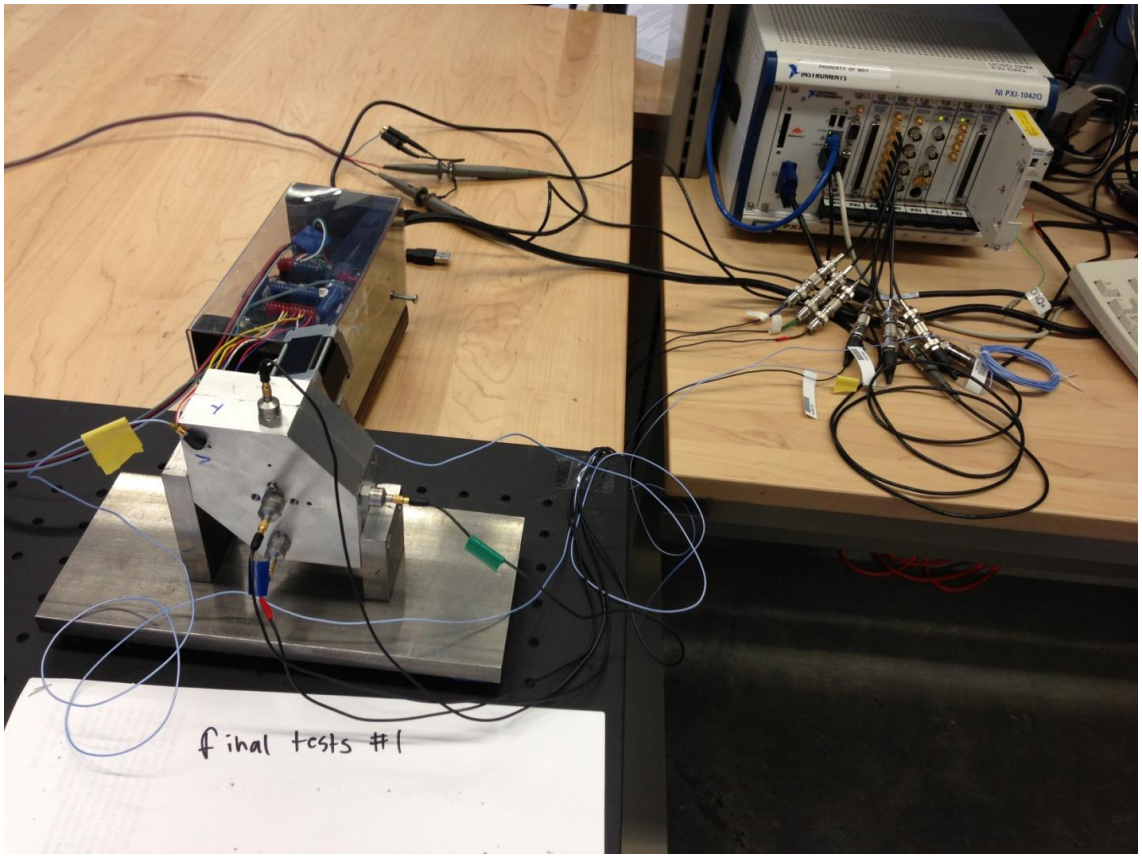

Figure 10: Test stand

2.2. Designed Damage Types

Data was collected from damaged versions of the device to be used later in damage detection analysis. To do this, the potential damage types of this device were evaluated. Artificial damage types were created for this project, but during initial testing and adjustment of the device a few new damage types were discovered and are listed below in Table 1. If the damage was severe enough to cause a failure to close the switch, then this damage type can already be detected by monitoring whether the switch closes or not, so the only damaged scenarios considered were ones that did not cause failure. 18 damage types were chosen and are described in the three categories, bearing, spring, and central damage listed below. Damage to the housing and electrical failures were not addressed in this study.

Table 1: Damage types

Damage types	
originally targeted	broken spring
	weak spring
	moderately worn swinging arm bearing
	severely worn swinging arm bearing
	moderately worn main wheel bearing
	severely worn main wheel bearing
	no lubrication
discovered	main wheel over rotation
	loose bearing screw
	deformed spring
	binding/tolerance stack up

2.2.1.1. Spring Damage

The first spring damage type considered was the loss of spring stiffness. This was simulated by replacing the healthy springs with identical length springs that had a lower spring

constant value. Damage types one and two, respectively, consist of replacing the spring on side one and two with a spring whose constant lowered from 1.16 lbs. /in to .84 lbs. /in. The next damage type considered was discovered by accident; while assembling the machine, one of the motors arms was installed incorrectly and caught on the spring while operating and deformed the spring. The machine still operated correctly with the deformed spring, so this spring was saved and installed on side one and two respectively for damage types three and four. The deformed spring is pictured below in Figure 11.

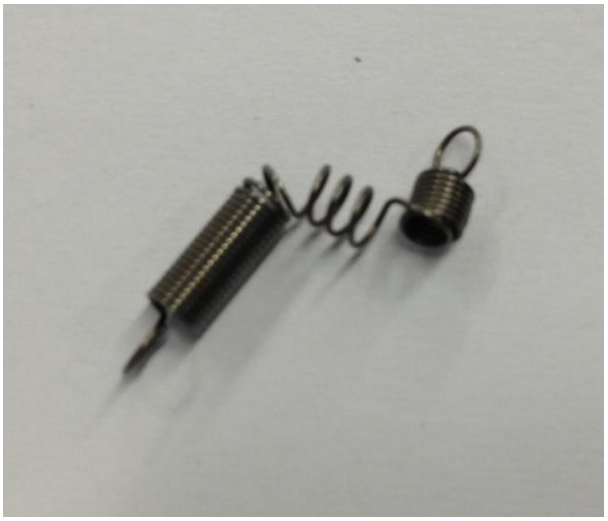

Figure 11: Highly deformed spring

The final spring damage type was designed to simulate a spring that has broken in half or fallen off its mount, which is simulated by simply failing to install a spring. Damage type 17 and 18 consist of the lack of spring installation on arm one for type five and on arm two for type six.

2.2.1.2. Bearing Damage

Bearing damage was chosen as the next category of damage. The device had three bearings installed, one for each side and one for the main wheel. Worn bearings were

simulated by manufacturing bearings that were undersized. Differing severities of bearing damage were implemented. A 3% +/- .5% reduction in bearing diameter was chosen as the least severe bearing damage type and a 7% +/- .5% reduction in bearing diameter was chosen as the next step of severity. This simulated bearing damage for both severities was applied to all three bearings individually, creating six damage types that are listed below in Table 2.

Table 2: Damage types 5-10

Damage Type	Description	Severity
5	Worn Main Bearing	3% Reduction in diameter
6	Worn Main Bearing	7% Reduction in diameter
7	Worn Side #1 Bearing	3% Reduction in diameter
8	Worn Side #2 Bearing	3% Reduction in diameter
9	Worn Side #1 Bearing	7% Reduction in diameter
10	Worn Side #2 Bearing	7% Reduction in diameter

The last two bearing damage types were naturally occurring. It was found that the swinging arm bearing mounting screws would become loose if the device was operated for a long period of time and no thread locker was installed on the screw. This was used as a damage type, where type 13 included side one's bearing screw installed with minimal torque applied and damage type 14 is the same as type 13, but for side two.

2.2.1.3. Central Damage

The final category of damage consists of damage that affects both sides of the machine while operating. The first damage type considered was over rotation of the main

wheel, caused by a lack of force from the ball and spring plunger, which was simulated by adjusting the installation position of the ball and spring plungers. The optimal ball and plunger force was found to occur when the plungers were installed in their threaded holes until they bottomed out, and then backed off by 45 degrees. Damage type eleven consisted of the plungers installed and backed off by 180 degrees. Damage type twelve consisted of the ball and spring plungers being removed, this damage type would primarily result in a failure to close the switch, but some events were able to be recorded prior to failure.

The device's moving parts were lubricated with a film of light weight oil; therefore damage type 15 was designed to simulate a lack of lubrication. The device was taken apart and the oil was removed, then it was reassembled dry. The final central damage type, type 16, was discovered during initial testing of the device. If the swinging arm and main wheel were both installed at a position that minimized their distance apart then the swinging arm would bind with the main wheel causing either their motion to stop thus a failure to complete a cycle or it would slow the device down. This damage type was a result of a tolerance stack up, and it was recreated by assembling the machine in the same way.

Chapter 3
Testing

3.1. Introduction

Department of Energy (DOE) workers have long been able to detect and remove defective units simply by listening to them operate and recognizing inconsistencies. However, without an automated process in place to accurately monitor vibrations, it has not been a reliable process. This process spurred the thought behind this project: to observe the vibrations coming from the machine and numerically quantify when these vibrations have changed enough to indicate damage. These vibrations can be observed with a microphone or an accelerometer. For this project accelerometers were chosen to measure the vibrations because they avoid the metal to air interface that distorts the signal when vibrations convert from a mechanical vibration to an audible pressure wave that's measured with a microphone.

For testing purposes, the safety mechanism under consideration was mounted to a frame. A "test" consisted of the machine completing one cycle composed of 20 events. To measure the vibrations during the test, the accelerometers are mounted to the aluminum housing of the device. Measuring and converting the output of these accelerometers requires

an appropriate data acquisition system. The details of this equipment their selection process and the testing procedure are described in detail below.

3.2. Data Acquisition

3.2.1. Sensors

Two categories of accelerometers exist, AC and DC response accelerometers. AC-response accelerometers are suitable for measuring dynamic events [22] and were chosen for this project as opposed to DC-response accelerometers that are primarily used for measuring static forces. A piezoelectric type AC-response accelerometer was chosen for its high frequency operating range [23] and availability.

The frequency range and sensitivity of the accelerometers were chosen based on experimental data coming from initial testing. These initial tests were run with the frequency sampling rate set at 330 KHZ using an accelerometer with a frequency range up to 60KHZ to measure the vibrations from the operating machine. Frequency analysis showed frequency content up to 50KHZ which narrowed the required frequency operating range of the accelerometers.

Sensors with a sensitivity of 100mv/g were used initially, but the data from these sensors, when located closest to the machine's parts, showed saturation (depicted below in Figure 12). Sensors with a sensitivity of 10mv/g were then selected to eliminate saturation because the lower sensitivity correlates to a larger measurement range.

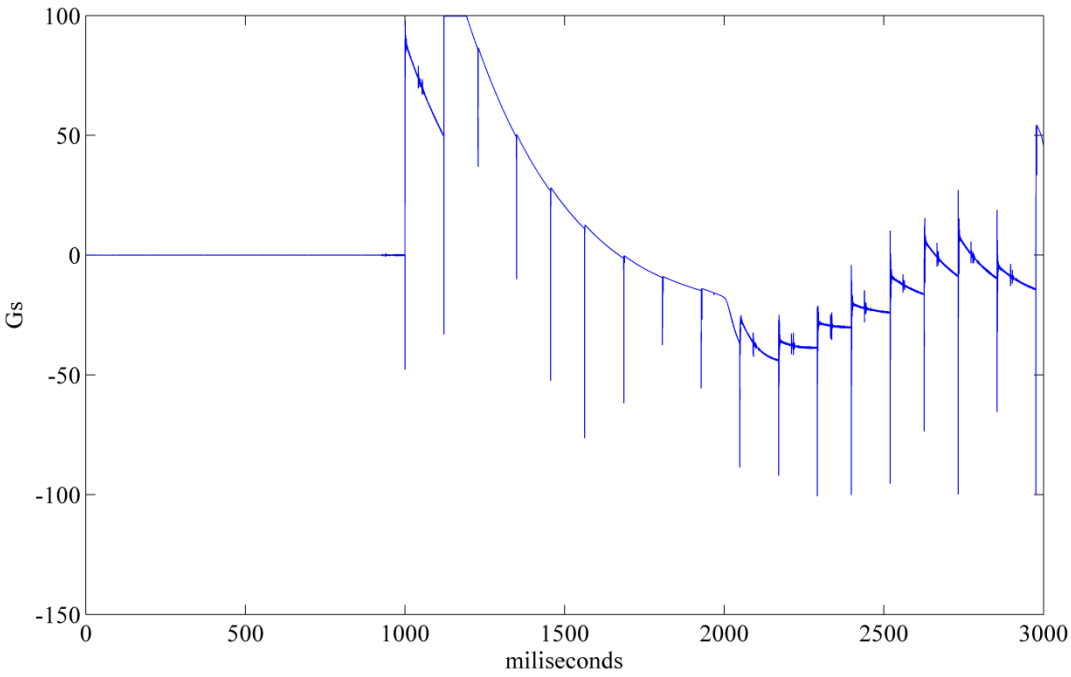

Figure 12: Data saturation, accelerometer sensitivity=100mv/g

As a result of the initial testing, six accelerometers were chosen; four with a high frequency range of 5-60KHZ and two with a lower frequency range of 1-20KHZ. Two Tektronix model P2200 voltage probes were attached to the motor controller to measure part of the system input, namely motor operation; this information is helpful in numerically separating the data. The motor controller is programed to output voltage signals that correlate with each motor turning on and off. These voltage signals are used to separate the data into appropriate events. Model specific information about all sensors used is listed below in Table 3.

Table 3: Sensor information

Device	channel	Brand	Model	Serial #	Description	Frequency Range (±3 dB)	Sensitivity
Accelerometer	1	PCB	352A60	SN150359	6gm, ceramic shear ICP®	5 to 60k Hz	10 mV/g
Accelerometer	2	PCB	352A60	SN146955	6gm, ceramic shear ICP®	5 to 60k Hz	10 mV/g
Accelerometer	3	PCB	352A60	SN146956	6gm, ceramic shear ICP®	5 to 60k Hz	10 mV/g
Accelerometer	4	PCB	352A60	SN150361	6gm, ceramic shear ICP®	5 to 60k Hz	10 mV/g
Accelerometer	5	PCB	352C22	SN76082	Miniature (0.5 gm), ceramic shear ICP®	1 to 20k Hz	10 mV/g
Accelerometer	6	PCB	352C22	SN76081	Miniature (0.5 gm), ceramic shear ICP®	1 to 20k Hz	10 mV/g
Voltage probe	7	Tektronix	P2200				
Voltage probe	8	Tektronix	P2200				

3.2.2. Data Acquisition System

A data acquisition system (DAQ) is required to operate the accelerometers used in this study. Since the highest rated accelerometers used have a frequency range up to 60KHZ, it is necessary to sample at twice that rate, or 120KHZ. However, it was difficult to find a DAQ that was capable of measuring eight channels at a sampling rate of up to 120KHZ simultaneously. A DAQ that is capable of measuring 8 channels at 100KHZ, with a built in anti-aliasing filter was available. It was decided that the sampling rate of 100KHZ was suitable because its nyquist frequency of 50KHZ should cover the majority of the frequency content emitted from the machine. In fact, damage was able to be detected by observing frequency content up to 50KHZ proving this DAQ to be a suitable choice. DAQ information is listed below.

National instruments: NI PXI-1042Q

Card slot #3: NI PXI-4472B card

3.2.3. Software

A program called SeDaq was run on MatLab to control the DAQ and convert the recorded output to MatLab files. The program SeDaq outputs the data, and other pertinent information, in these six sub structures:

Table 4: SeDaq output structure

NAME	DESCRIPTION
Fs	sampling frequency
ChannelColumns	the titles for each column of data within channels
Channels	information about each channel ie. Measurement type and measurement device
Warnings	equals a value of 1 if the accelerometer readings were clipped
Samples	total number of samples for that test
Data	the accelerometer readings converted from millivolts to gravitational acceleration (g)

3.3. Initial Testing

During the initial testing phase, information critical to operating the machine was gathered as well as information critical for subsequent testing. The dynamics of the machine were also adjusted for it to run properly at any speed. The ball and spring force, spring pin location, spring k value and the motor's speed and acceleration were all tuned. It was determined that running the machine at 80% of its full speed required 2 ball and spring plungers, both backed 45 degrees off from the fully tightened position, and ¾" long springs with a spring constant of 1.16 lbs./in with its mounting location installed in the pin hole that is second from the center of the main wheel. The motor controller was then adjusted and programmed. The machines combination code was programmed as a sequence of either motor completing a full revolution. To ensure the motor arms would start at the same location every time a test was run, a sequence was programmed that places the arms in the same location.

This sequence was run every time the device was powered on. This sequence spun the motors a full revolution at the lowest power setting, resulting in the motor arms connecting with the swinging arms and stopping, not having enough power to move them. Finally, the initial testing phase also included tests conducted without certain parts to help determine the makeup of the signal.

3.3.1. Sensor Locations

To observe the effect of sensor location on damage detection and to accomplish damage localization, an array of sensors were strategically placed around the outside case of the machine. The accelerometers chosen, by design, were only able to pick up accelerations in one axis. Two PCB 352A60 sensors were stud-mounted in the Z-axis as close as possible to the internal components - one near the main wheel and another near swinging arm #2. This is accomplished by mounting the sensor on the same threaded hole that the wheel and arm are mounted to, putting the sensor in line with the wheel and arm's spinning axis (Z-axis) as pictured below in Figure 13.

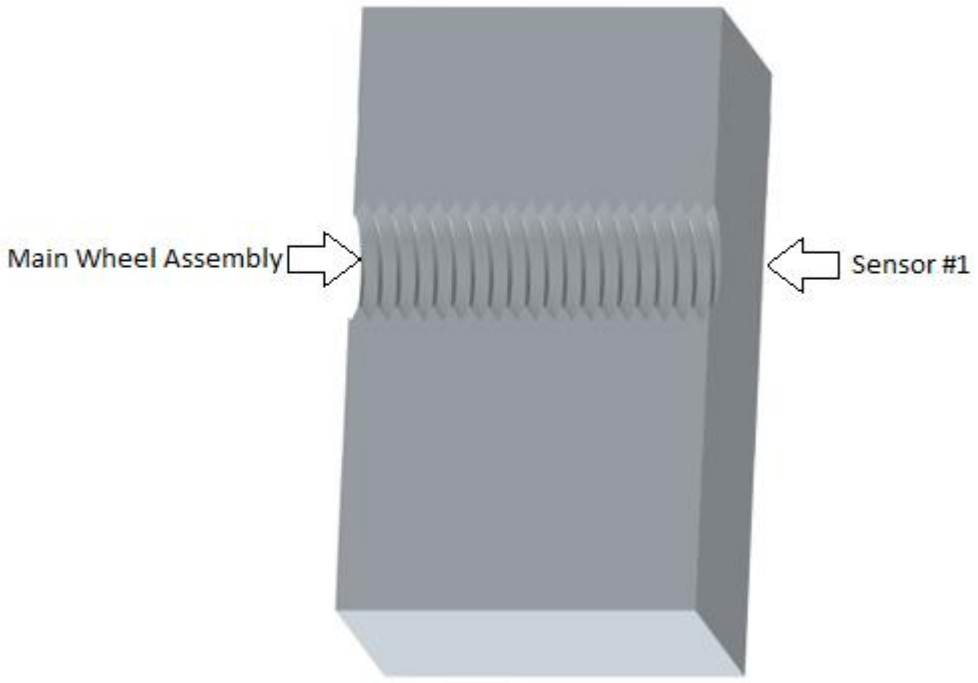

Figure 13: In-line sensor mount

Two more PCB 352A60 sensors were stud-mounted in the rotating plane of the main wheel and the swinging arms. Sensor #4 was mounted in the X-axis in line with the swinging arms on side one and the sensor #2 was mounted in the Y-axis in line with the main code wheel. Both these are pictured below in Figure 14.

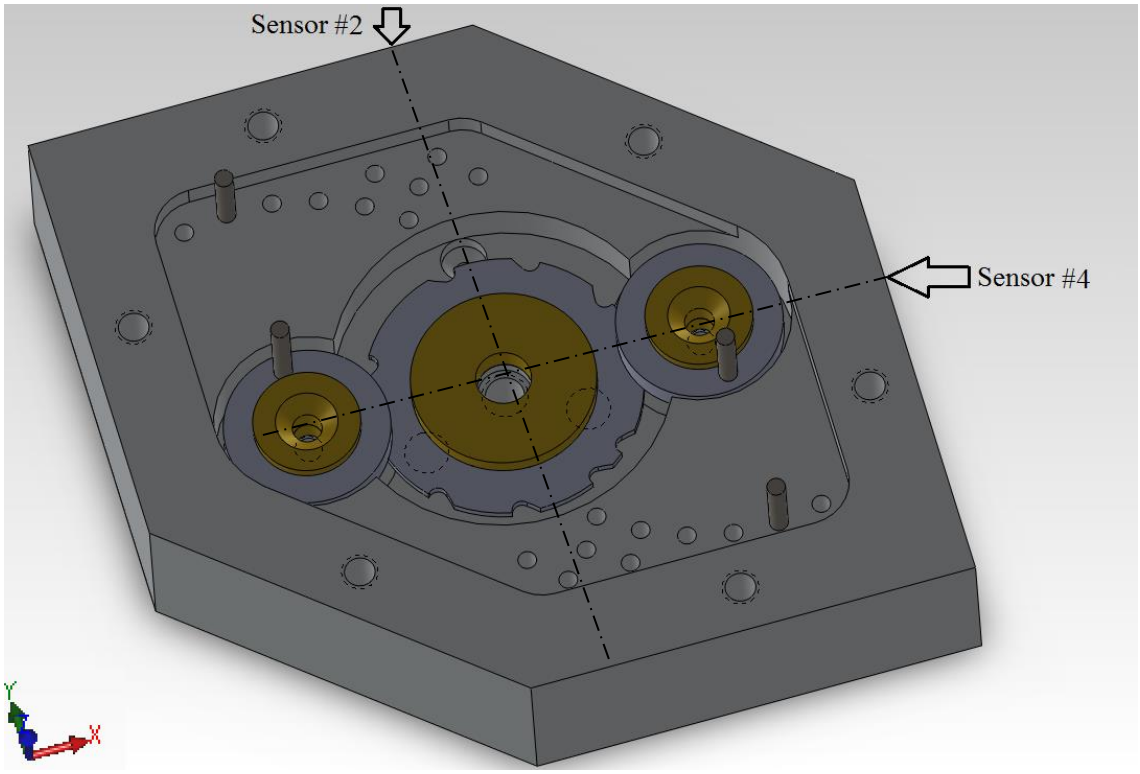

Figure 14: Sensor #2 and #4 mounting location

The final two accelerometers did not have stud mounts and were superglued to the structure. Sensor #6 PCB 352C22 accelerometer was superglued to the top case directly in between the two motors. The location for sensor #5 was not geometrically significant but was determined through trial and error; during testing it was found that an extra impulse could be seen in the recorded time series only when sensors were placed in the corner of the structure. In Figure 15 (below), this location is outlined in red.

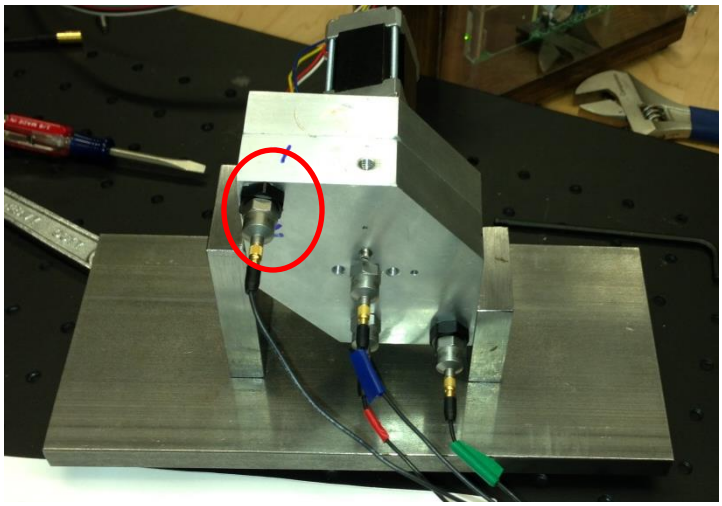

Figure 15: Aditional sensor location trials

The extra information recorded from sensors in this location could be due to resonances and the shape of the structure, although a detailed finite element model was not constructed to predict structural resonances. Sensors were mounted on the corner of the machine in all three axes, as shown in Figure 16 (below), to better understand this signal but the Z-axis was the only one to show the extra impulse. Sensor #5 accelerometer PCB 352A60 was super glued in this location.

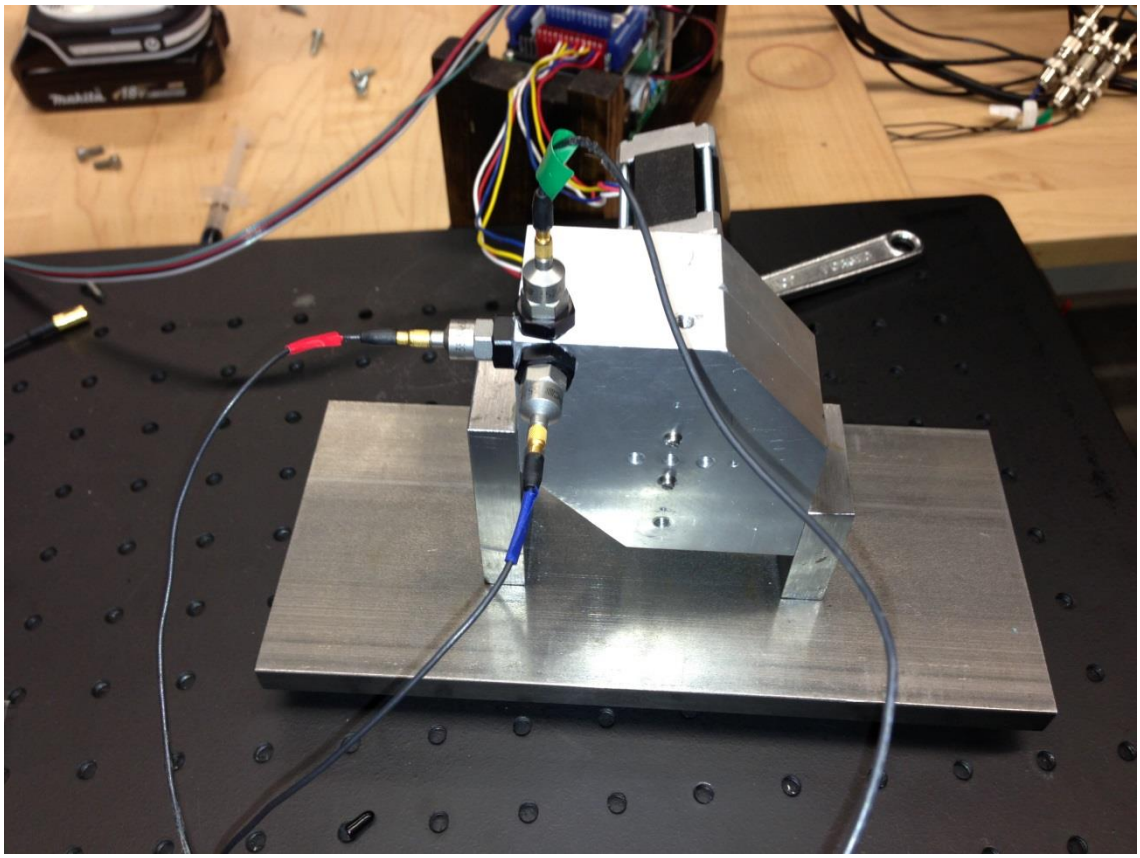

Figure 16: Aditional sensor location trials (2)

3.4. Testing Procedure

The testing phase is broken into three categories damaged tests, un-damaged ("baseline") tests, and training tests –Baseline tests consist of 80 events of the machine run in the un-damaged condition. These 80 events are separated from the training events (which are also run in the un-damaged condition) so that damage detection analysis can be performed on both damaged and un-damaged tests. A separate larger set of training tests were required for damage detection analysis as they give a representation of the baseline un-damaged condition. In total, 700 training events were recorded. The damaged tests consist of disassembling the

machine and reassembling it in the intentional damaged condition. There are 80 events run for each damage type, totaling 1,440 recorded damaged events.

The initial testing results show that variance due to assembling the machine is high. For some damage types the assembly variance is as high as the variance due to damage. This could make damage detection very difficult due to the generation of substantial false positives. To avoid this, the baseline set of data that damage detection models are trained on needs to incorporate the full range of assembly variance recorded. Doing so helps the damage detection model differentiate damage from assembly variance.

The final test procedure is designed with assembly variance in mind and introduces the most assembly variance possible. The mechanism was disassembled and reassembled prior to testing for every un-damaged test. This was accomplished by alternating damaged tests between each undamaged test.

Chapter 4
Feature Extraction and Discrimination

4.1. Background

After testing was completed, the accelerometer readings were processed and analyzed in the coding software MatLab. MatLab allowed numerical processes to be iteratively processed and provided several options to visualize the results. In the program, algorithms were written to process and detect damage. These algorithms were run in a reasonable amount of time (under 30 minutes) on a standard laptop with an Intel core i7 processor. This is significant as it shows that this process could be implemented remotely on portable devices.

4.2. Basic Data Interpretation

Basic information was extracted from the recorded data to form a strategy to analyze and detect damage. This basic information includes video footage which was synchronized with the acceleration time history to determine the cause of the accelerations in the data. Then, basic statistics were drawn from the data to differentiate events. This basic data interpretation is described in the sections below

.

Initial observations of the acceleration time series for all completed cycles showed that there were 20 impulses whose amplitudes were an order of magnitude higher than any other events in the acceleration data. The observed impulses in the acceleration time history are from high energy impacts in each event. Each cycle must complete 20 events to unlock the mechanical combination lock; each event has a high energy impact resulting from the swinging arm connecting with the main code wheel. Each impact can be seen in Figure 17 below.

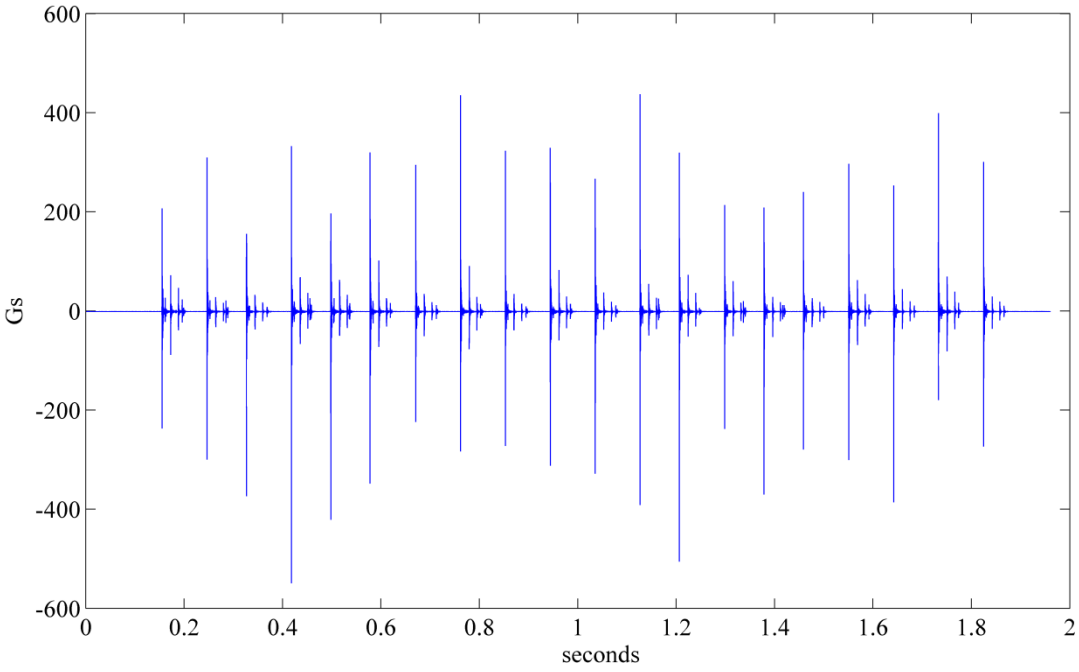

Figure 17: Acceleration time history: one cycle (20 events)

4.2.1. Video Analysis

During video analysis, the top half of the device was removed and the motors were mounted on a frame which allowed the internal components to be viewed during operation. Video footage was shot with a Go-Pro hero2 video camera and recorded with a frame rate of 120 frames per second. The footage was synchronized with the acceleration time history by

lining up the first impulse in the acceleration time history with the first contact of the swinging arm and main code wheel seen in the video. The subsequent 19 major impulses in the data lined up with the next 19 impacts seen in the video, verifying that the 20 high energy impulses seen in the acceleration time history are caused by the swinging arm impacting the main code wheel.

Another important observation from the video analysis was that each main impulse had a couple smaller impulses that followed. The repetition is pictured below, with each impulse being numbered for reference later in this thesis.

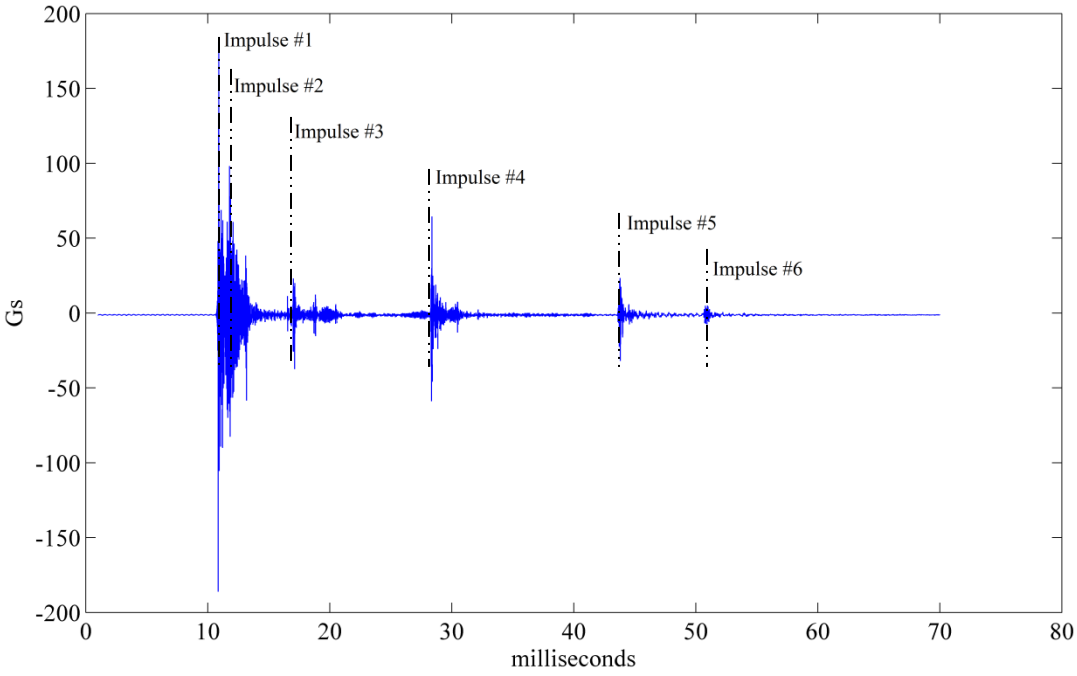

Figure 18: Basic event with labeled impulses

The video footage was useful in identifying the cause of the secondary impulses as well as confirming the main impulse's cause. The first secondary impulse, impulse #2, occurs in the middle of the main impulse. This occurred at the same time as the main code wheel finished its 18 degree rotation and was moving into its next spot. At the end of this move the

ball and spring plunger lodge into the next hole on the main code, resulting in a clicking sound. The remaining impulses in the data do not occur until the moment the motor arm reconnects with the swinging arm. A figure of the acceleration time series for the first event is shown below, along with labels for the cause of each impulse.

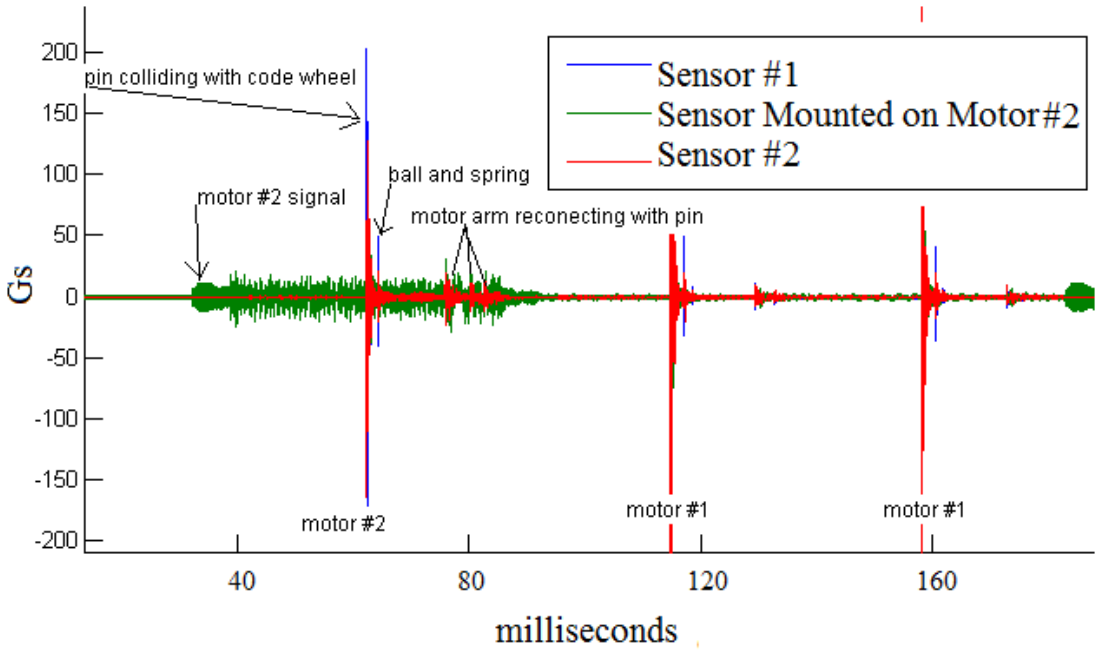

Figure 19: Video analysis results

4.2.2. Data Division

Each event, and all the impulses contained in it, is caused by the parts in either side one or side two of the device (swinging arm, bearing…). The 20 events can be separated, using the combination code that unlocks the machine, into 10 events correlated to side one and 10 events correlated to side two's operation. Separating side one from side two will enable localization of damage. If only side one's events show a change from baseline data, it can be assumed that only side one has damage present and vice versa for side two. If a central component were damaged like the code wheel, then both side one and side two would show

deviation from baseline signals. To be able to numerically differentiate these events certain features were drawn from the data. Channels seven and eight recorded voltage signals, these signals went from eight volts to zero volts when motors one or two, respectively, were powering the device. These channels were used to numerically differentiate the exact beginning and end of all 20 events. A figure with these divisions for each event and correct labels for side one or two is pictured below.

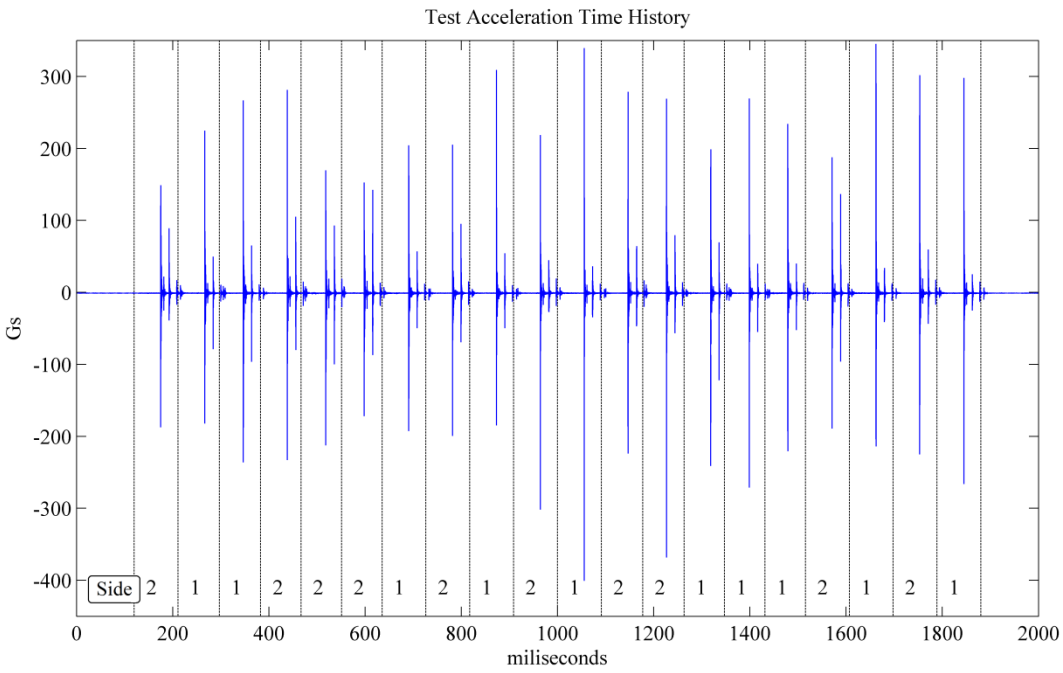

Figure 20: Divided and labeled events from test acceleration time history

The data, along with its additional information, was put into a structured array in MatLab to make coding for iterative calculations less complex. All acceleration time series and information pertaining to it were stored in two structured arrays: one for all undamaged tests and another for all damaged tests. These structured arrays and their sub-categories are pictured below.

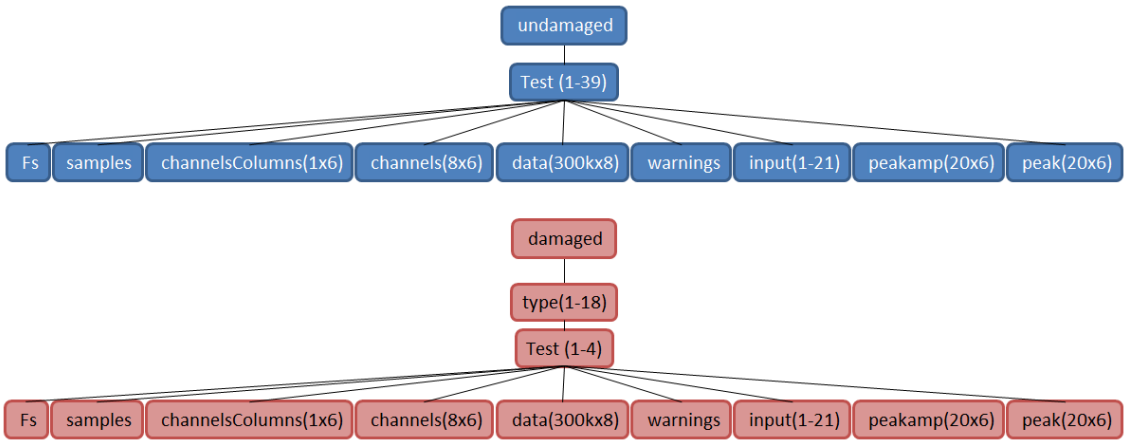

Figure 21: Data structure

4.3. Mahalanobis Distance

The strategy to detect damage consists of reducing the data set down to a manageable number of parameters for both the undamaged sets of data and the damaged sets. These parameters will then be compared to the baseline set of undamaged parameters to see if they differ. The strategy chosen to measure similarity, and thus determine if the device is damaged or un-damaged, is Mahalanobis distance. Mahalanobis distance is an intelligent distance measure and can give a single scalar metric that tells how far one parameter is from a group of multiple parameters [24,27]. It is an intelligent measure of distance because it measures more than the Euclidian distance (the standard distance measure) and it accounts for correlations in the data sets by using covariance as weights. An example of how Mahalanobis distance is useful compared to standard distance measures is pictured below.

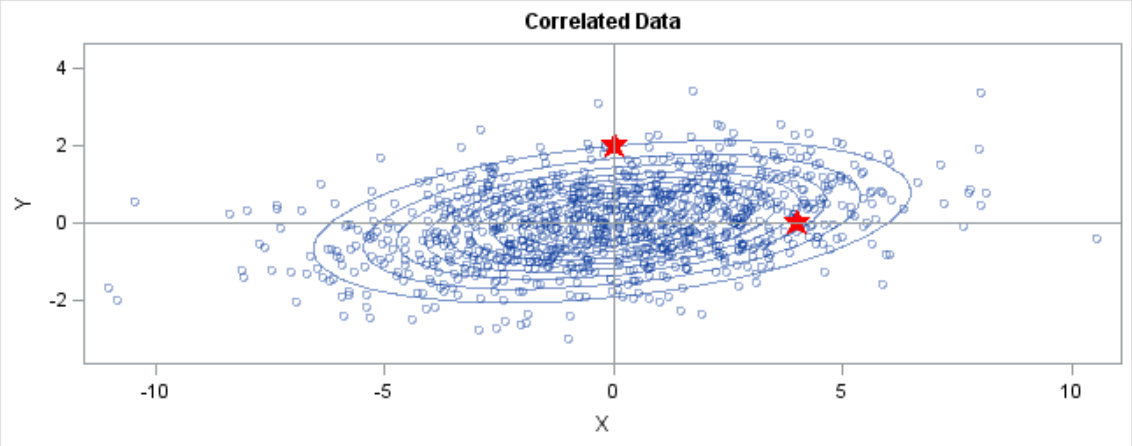

Figure 22 : Mahalanobis distance example [25]

The ellipses in the figure represent prediction ellipses for the bivariate normal distribution that generated the data with the inner most ellipse being the 10% prediction ellipse and the outermost the 90% prediction ellipse. The star on the right in the figure has a higher Euclidean distance but a lower Mahalanobis distance and the star on top has a higher Mahalanobis distance and is the data point that is furthest from the set of data in terms of its distribution. The resulting distance measure from Mahalnobis distance is scale invariant. Mahalanobis distance *d* is given by the following equation:

Equation 1: Mahalnobis Distance [26]

$$d^2 = (\mathrm{Y} - \mu) * \sum^{-1} * (\mathrm{Y} - \mu)^T$$

This is the distance of each data point Y from the baseline data X where μ and ∑ are the sample mean and covariance of the data in X.

4.4. Peak amplitude

4.4.1. Introduction

The first feature considered was peak amplitude analysis, given that it clearly should be influenced by most of the failure modes that modify the internal dynamics. This single feature, peak amplitude, was drawn from only Impulse #1 in each event because this impulse contained the most energy and was repeated consistently across all baseline and training events. More advantages of choosing impulse #1 are described in more detail in section 4.5.2. Using the defined beginning and end points in each event, the MatLab function max() was used to find the absolute value of the magnitude of impulse #1 and its location. Both peak amplitude and location values were stored in the data structure. This peak amplitude analysis was performed on the 700 training events, the 80 undamaged events, and for the set of 80 damaged events for all 18 damage types.

To unlock the machine, 20 events had to be completed. The 20 events were a repetition of the same action, so each event was considered its own test. However, when comparing events to the training data, the events were treated individually. Each damaged and un-damaged event's peak amplitude value was compared to the set of peak amplitude values from its respective events in the training data. These peak amplitude values were compared to the training data using Mahalanobis distance. Each Mahalanobis distance gave a single value that represented how similar that event was to the training data. The prior knowledge of whether the event was from a damaged machine or un-damaged machine and the calculated Mahalanobis distance was then used in a receiver operating characteristic (ROC) analysis to gauge the peak amplitude analysis's effectiveness at classifying the machine as damaged or undamaged.

4.4.2. Receiver Operating Characteristic (ROC)

An ROC curve compares the true positive rate (TPR) to the false positive rate (FPR) of damage detection as the detection threshold (classification boundary) is varied [28]. For an ROC curve the threshold must be moved from a minimum point that classifies all tests as un-damaged to a maximum point that classifies all tests as damaged or vice versa. At each point that the threshold is placed, all parameters for undamaged and damaged events are classified as either a true positive or false positive detection based off the value for the threshold at that point and the prior knowledge of whether the event was from an un-damaged machine or a damaged machine. A figure depicting a detection threshold choice is shown below.

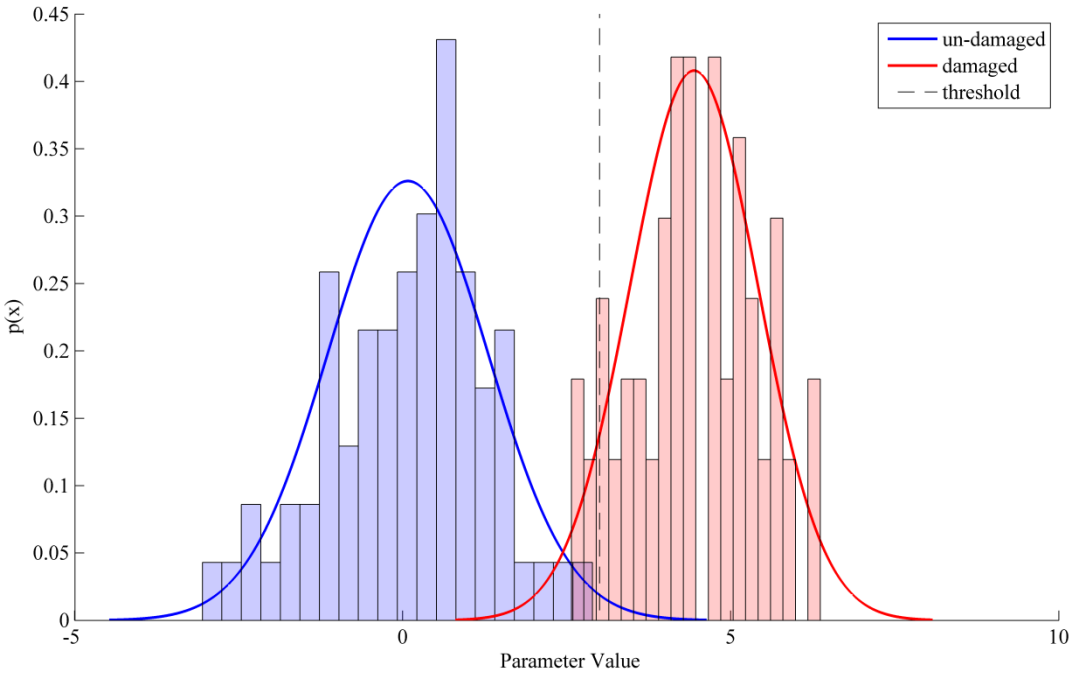

Figure 23: Histogram of damaged features and un-damaged features

All true positive detections and false positive detections are summed and put into a fraction to determine the FPR and TPR for each threshold value. This FPR and TPR is recorded for all threshold values then plotted to form the ROC curve. This curve enables the

performance of a damage detection analysis to be visualized [29]. This curve can then be used to directly compare damage detection analysis to see which analysis has the best performance. The curve also provides information that, when combined with the preference to minimize false positive or false negative indications, aides in determining the optimal damage detection threshold value. Depicted below are three examples of ROC curves and threshold selections.

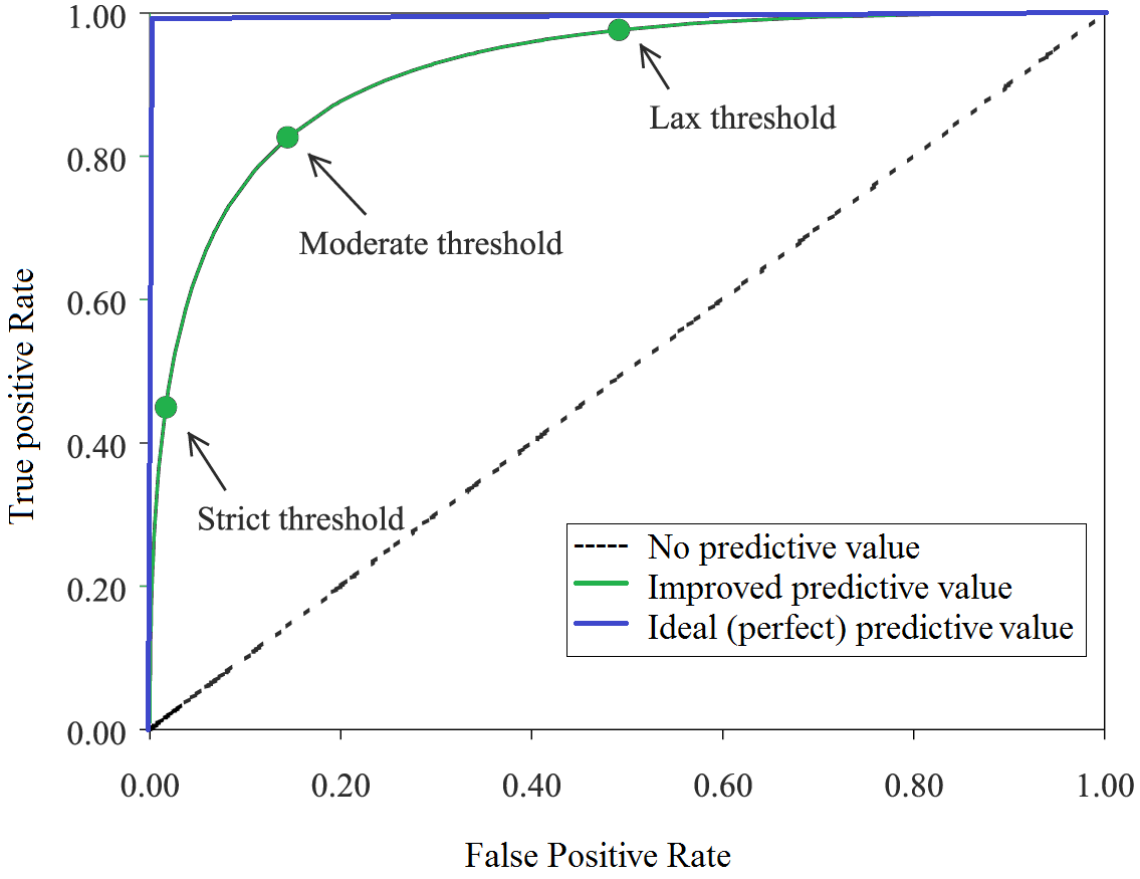

Figure 24: ROC curve examples [30]

4.4.3. Discussion & Results

ROC curves were made for each damage type's six sensors, and each sensor's signal was split into side one and side two events as described previously. Peak amplitude analysis performed the best at detecting damage types 1 – reduced spring constant on arm #1, 2 –

reduced spring constant on arm #2, 3 – broken spring on arm #1, 4 – broken spring on arm #2, 17 – no spring on arm #1, 18 – no spring on arm #2. The ROC curves for theses damage types are shown below in Figure 25

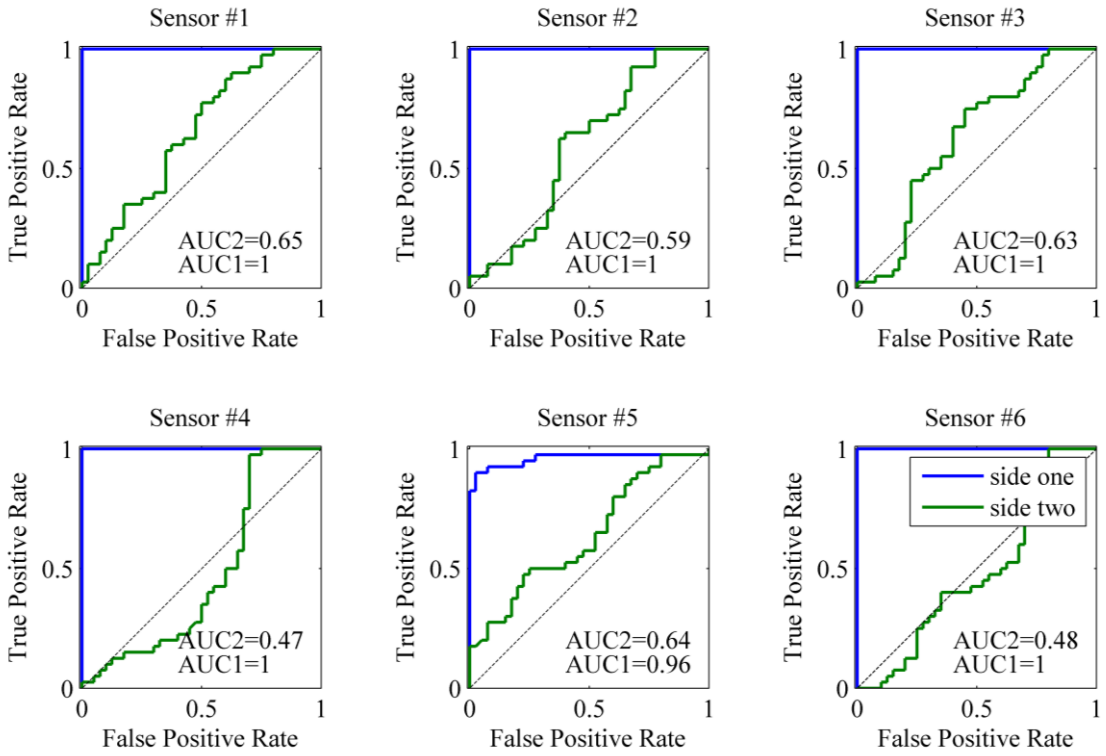

Figure 25: Peak amplitude roc curves damage type #17 - no spring on arm #1

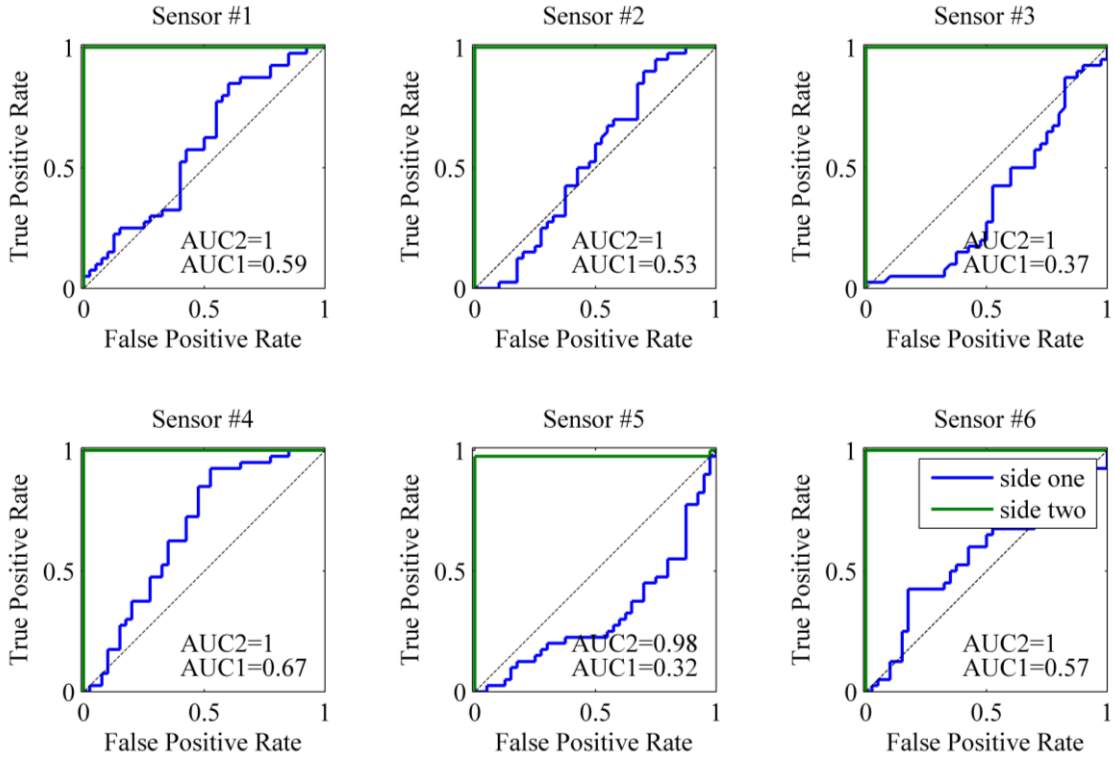

Figure 26: Peak amplitude roc curves damage type #18 -no spring on arm #2

Damage types 17 and 18, pictured above, were the most easily detected damage types. They consisted of removing the spring to simulate a spring breaking in half or falling out of position. Peak amplitude analysis predicted the damage almost perfectly regardless of sensor location and was able to localize the damage to which side it occurred on.

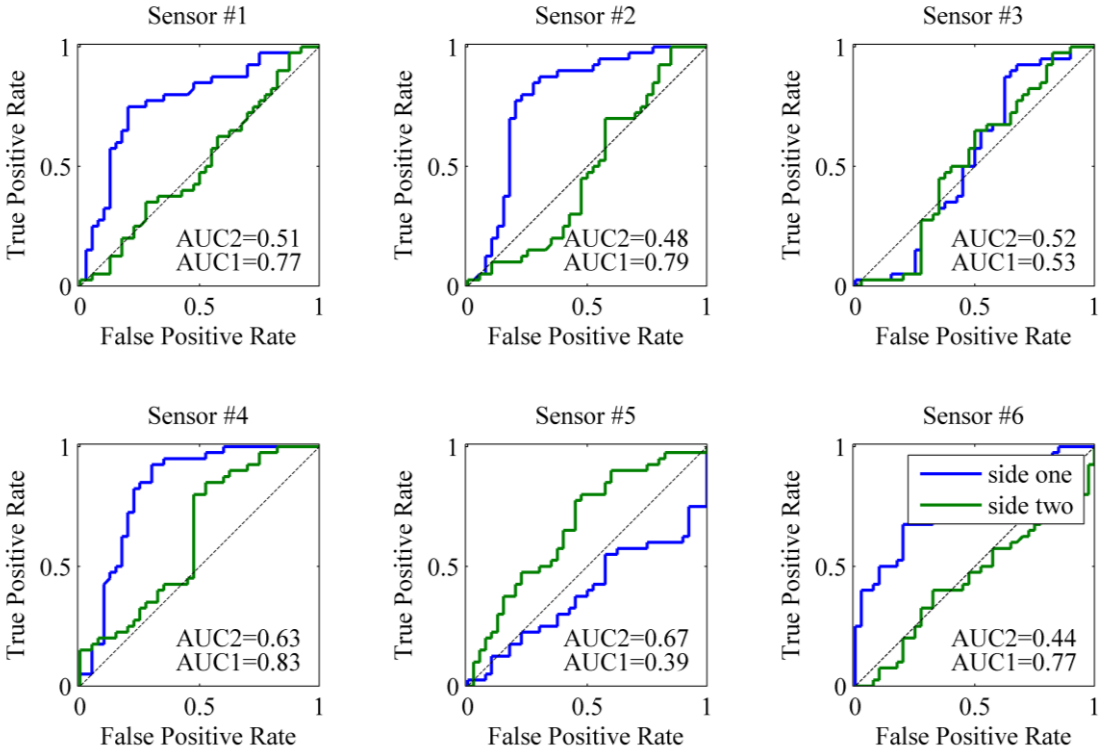

Figure 27: Peak amplitude roc curves damage type #1 - k=.84 on arm #1

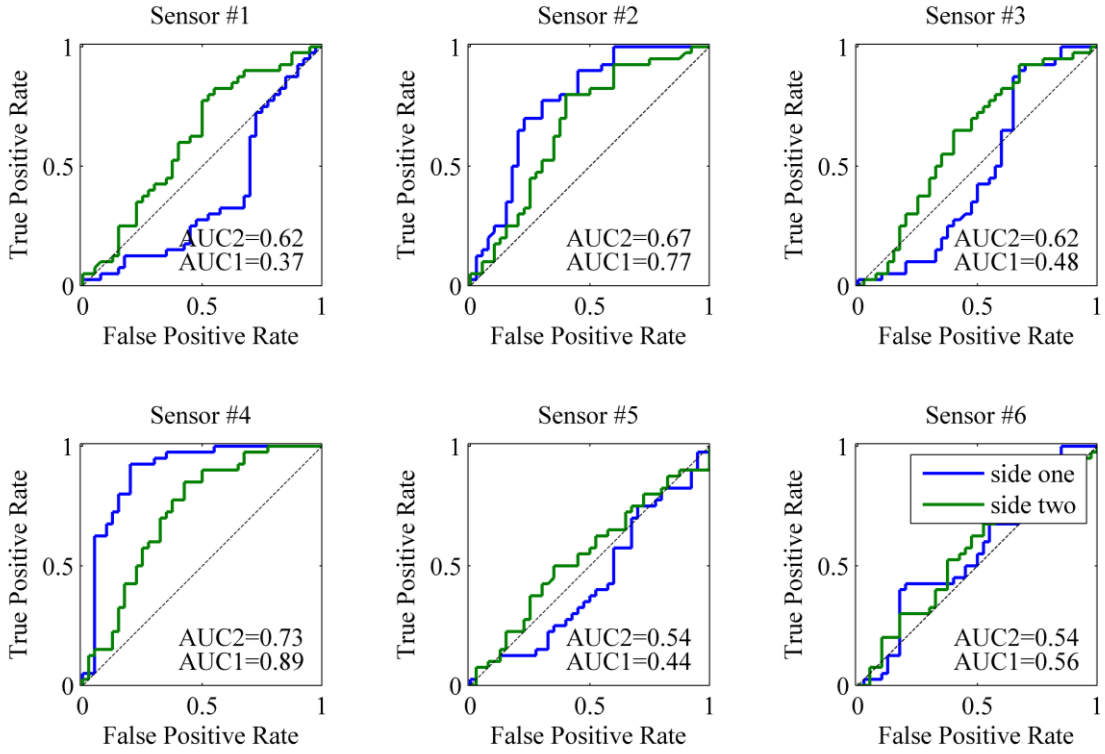

Figure 28: Peak amplitude roc curves damage type #2 - k=.84 on arm #2

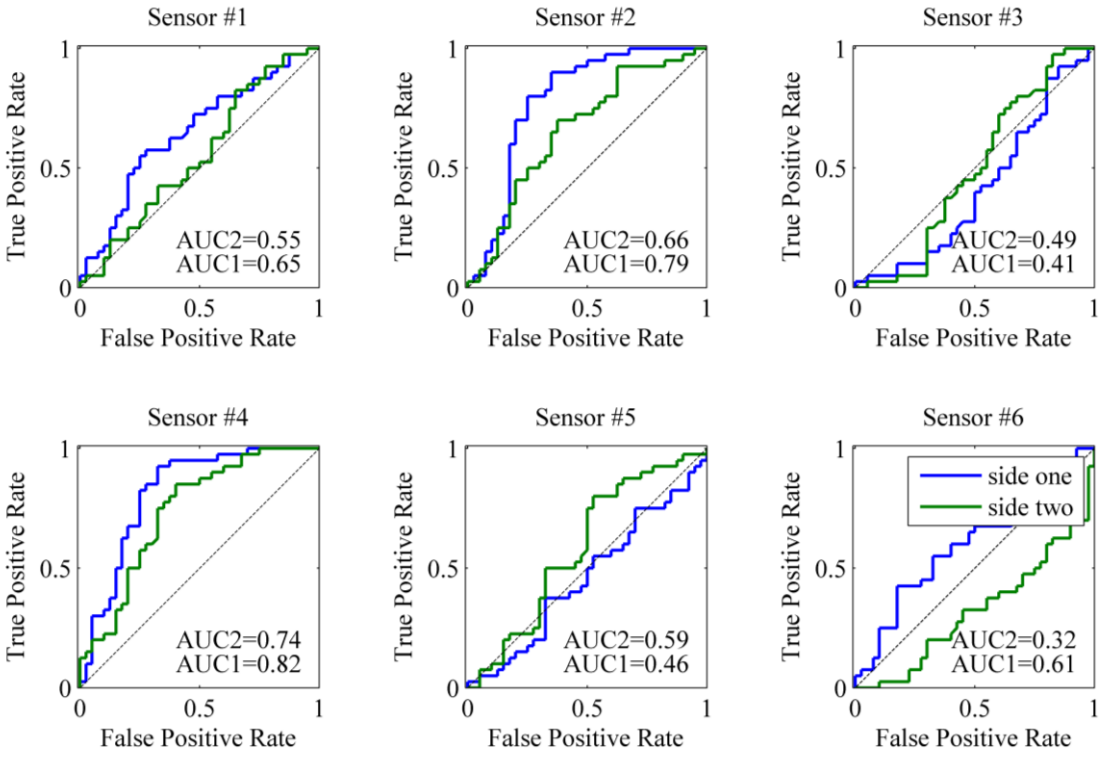

Figure 29: Peak amplitude roc curves damage type #3 -broken spring on arm #1

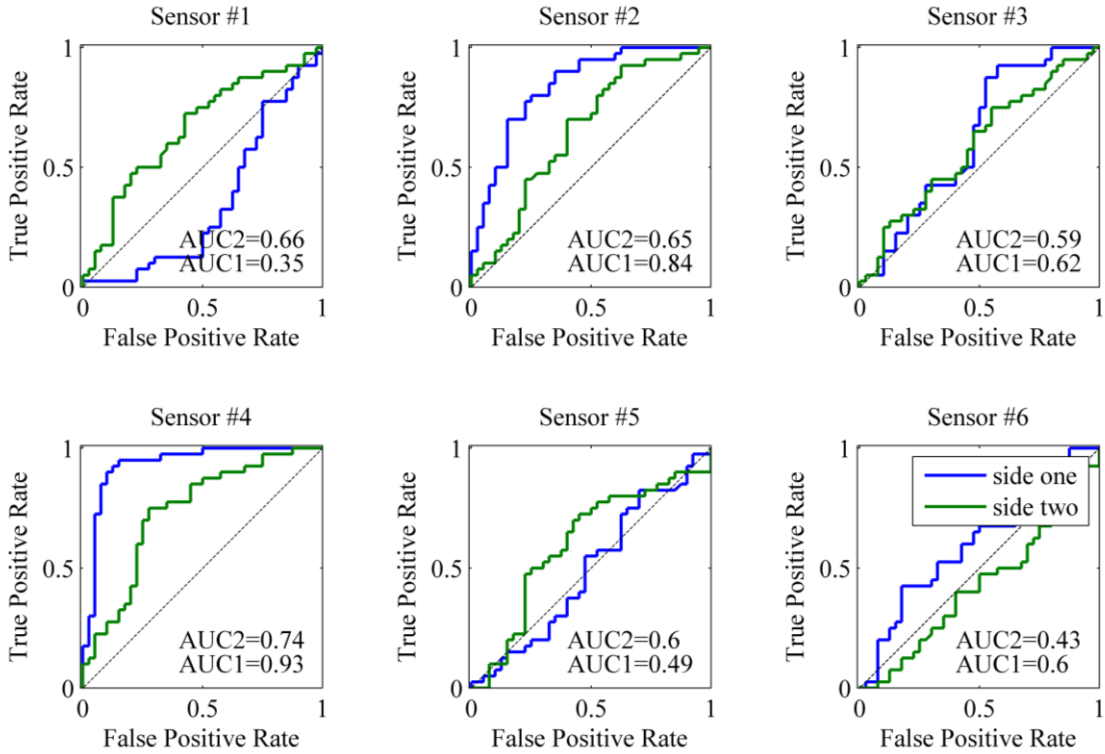

Figure 30: Peak amplitude roc curves damage type #4 - broken spring on arm #2

Damage types 1-4 consisted of replacing the spring with one that has a slightly different spring constant (a detuned stiffness). These damage types ended up being the most difficult to detect; only two sensors were capable of detecting the damage. Sensor #2 and #4 were the only sensors in the same plane as the rotating components in the machine, this location enabled the sensors to pick the energy from the impulses in a path that is more direct than the other sensor locations. Peak amplitude is directly correlated to the impulse energy, explaining why peak amplitude analysis performed well with sensor locations #2 and #4.

When the springs were removed from the component, the events correlated to the side without the spring had perfect damage detection while the other side had no predictive value of damage. However when the damage severity was decreased to damage types 1-4, both sides of the machine were able to predict damage regardless of what side was damaged. These damage types, 1-4, were the hardest damage types to detect and their results did not follow clear explainable trends as closely as the rest of the results did.

4.5. Autoregressive Model

4.5.1. Introduction

An Autoregressive (AR) model predicts the current value in a time series as a weighted linear combination of some number *n* previous points making an AR model rely on only output data from a system. An AR model is well suited for this project as only the output, not the input, from the safety mechanism under consideration is fully known. The weights (θ) for the previous *n* time points are the parameters for the AR model with an order of *n*. The equation for the time series prediction using linear regression is shown below in Equation 2.

Equation 2: Linear Regression Model [31]

$$\hat{x}(t|\theta) = \varphi^T(t)\theta + e(t|\theta) \qquad t = 1 \dots N,$$

Where φ is the regression vector, $\varphi = [-x(t-1) - x(t-2) \dots - x(t-n)]$, and e is the prediction error. The solution to this model is over-constrained and has multiple possible solutions. The optimal solution was defined as one that minimizes prediction error e in the least squares sense. This is done by setting the prediction error term equal to zero and solving with a matrix inversion as shown in Equation (3) below:

Equation 3: Matrix based solver for AR model

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 \\ x_2 & x_3 & x_4 & x_5 & x_6 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{N-5} & x_{N-4} & x_{N-3} & x_{N-2} & x_{N-1} \end{bmatrix} \begin{Bmatrix} \theta_5 \\ \theta_4 \\ \vdots \\ \theta_1 \end{Bmatrix} = \begin{Bmatrix} x_6 \\ x_7 \\ \vdots \\ x_N \end{Bmatrix}$$

The parameters θcan be compared to detect changes in the system or the parameters can be used in the original formula (equation) to form the prediction $\hat{x}(t|\theta)$ which can then be subtracted from the actual time series to get a prediction error vector.

4.5.2. Model Analysis

AR model damage detection analysis was performed for each event on impulse #1 only. If the model was created off the entire event's time series, then most of the model would be representing white noise and the same model would also have to represent multiple impulses that are all different. These conditions would make a model's predictive performance very low as information from each impulse would average out with each other and would be lost. Therefore, the model was created using the time series from only one of the impulses contained in the event. Impulse #1 was chosen first for a variety of reasons, with the two

primary ones being reliable (consistent) observability and signal-to-noise ratio. The first impulse contained the most energy in the signal which increased the chance that the damage type would be excited and output a detectable signal. The first impulse also showed the least amount of variance across all the training data, suggesting that a reliable model could be made based on impulse #1. Modeling other impulse numbers and comparing their performance is discussed later in the thesis.

AR models of order n were created for all training, un-damaged, and damaged event's first impulse. Their parameters (1:*n*) and error vectors were stored for comparison. Each event's individual parameters had a Mahalanobis distance calculated from the set of training data for the same number *n* parameter with the exception of the first parameter which was equal to one for all models. The average Mahalanobis distance was taken across all parameters for each event. This average Mahalanobis distance was then used in ROC analysis. A similar approach was also taken using the model error as opposed to model parameters. The root mean square value (RMS) was calculated for each event's error vector and then used in ROC analysis.

The ability of Mahalanobis distance and ROC analysis to provide a scale invariant result, when comparing damage detection performance, was taken advantage of to find the optimal AR model order for damage detection. Damage detection performance was defined as the area under the ROC curve. The best ROC curve for all 18 damage types was averaged together to obtain a single value used to compare performance. The performance of the AR model peaked at a model order of 28 and became unstable at model orders higher than this, as can be seen in the figure below. A model order of 28 was chosen to perform the final ROC analysis.

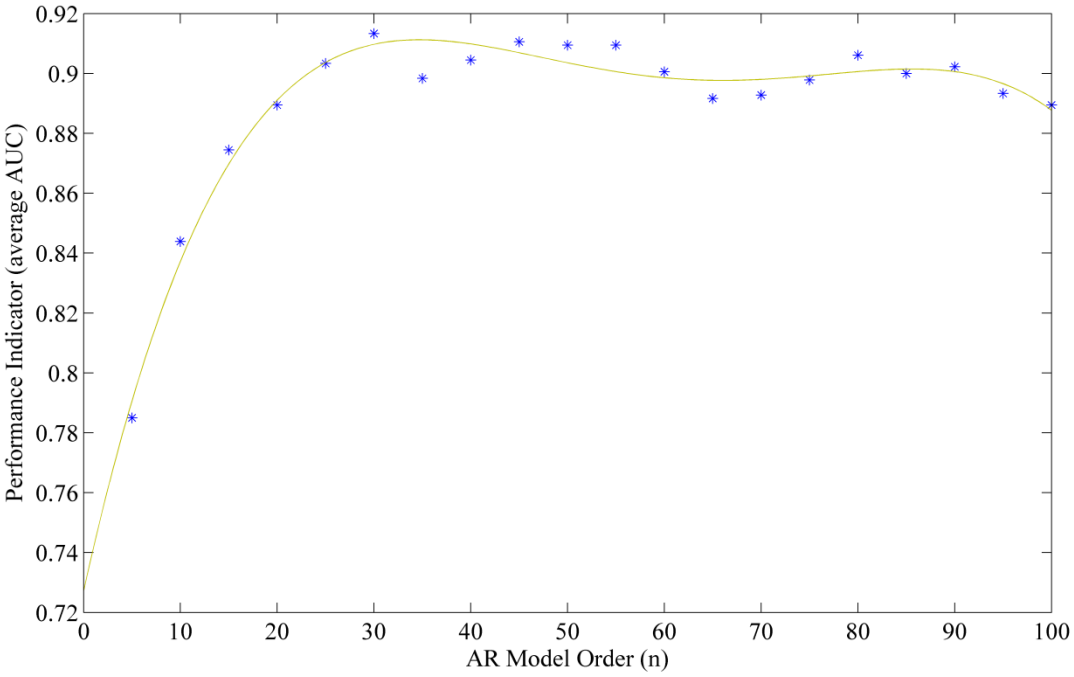

Figure 31: AR model performance vs. model order n

4.5.3. Results

Comparing the area under the curve (AUC) for the ROC curves revealed that comparing variations in parameter values to detect damage instead of error vector variation produced better damage detection performance across all damage types and sensors. Comparing the ROC curves for all damage detection analysis revealed that AR parametric analysis performed the best at detecting any damage related to the bearings. Damage types 5- Worn main bearing 11.5mm, 6- Worn main bearing 11mm, 7- Worn bearing 5.75 mm on arm #1, 8- Worn bearing 5.75 mm on arm #2, 9- Worn bearing 5.5mm on arm #1, 10- Worn bearing 5.5mm on arm #2, 13- Loose bearing screw on arm #1, and 14- Loose bearing screw on arm #2 were all predicted best by AR parametric analysis. Their ROC curves for all sensors and both sides are pictured below.

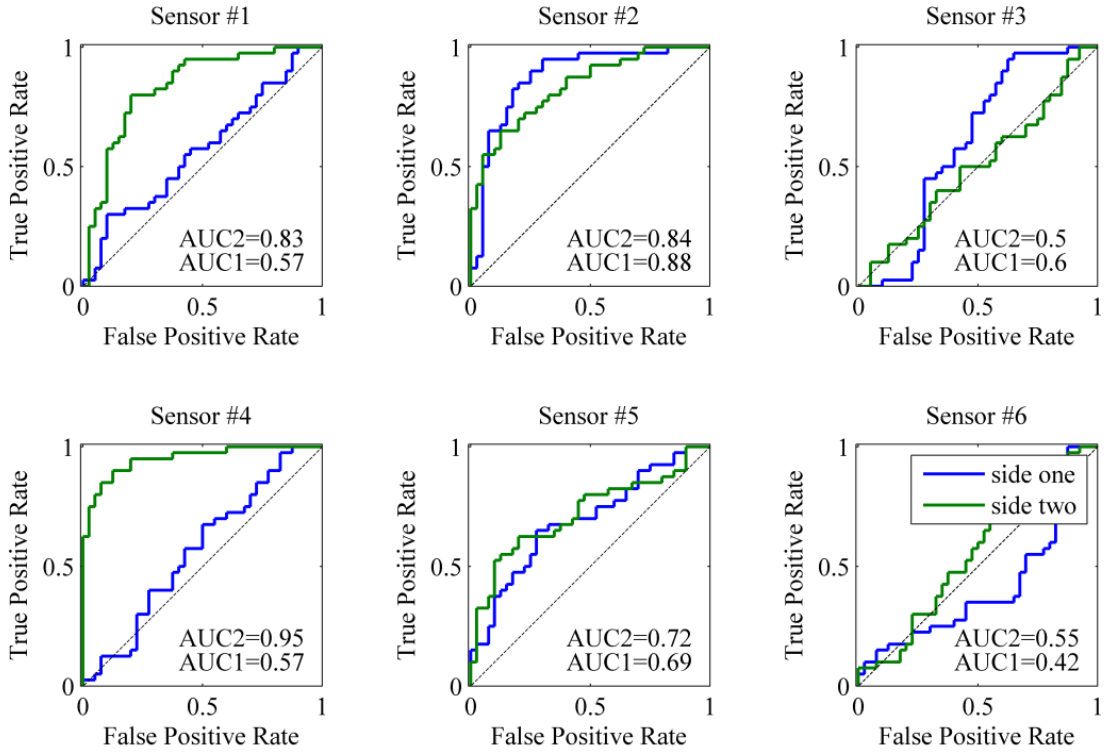

Figure 32: AR parameter analysis ROC curves, damage type 5 worn main bearing

Damage type 5 (pictured above) has minor damage to the main code wheel's bearing and damage type 6 (pictured below) is a more severe case of the same damage. Both figures show that sensor #2 is in the best position to detect this damage while sensor locations #3, #5 and #6 have little predictive value and sensor #4 is biased towards side two's events.

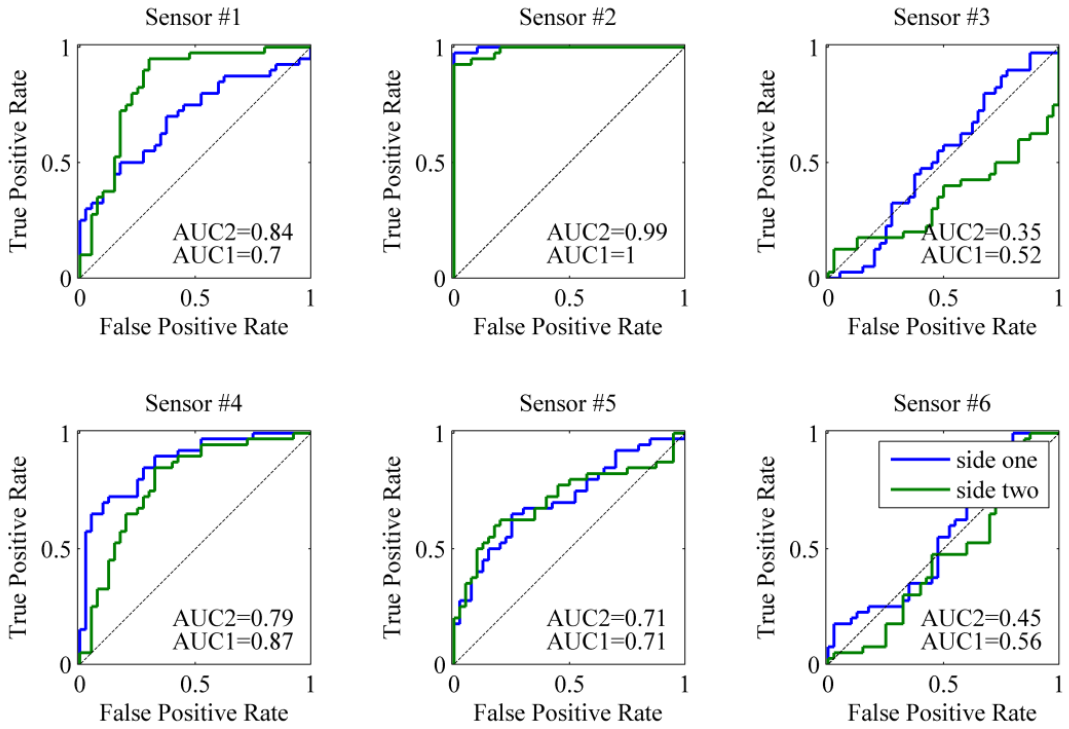

Figure 33: AR parameter analysis ROC curves, damage type 6 worn main bearing 11 mm

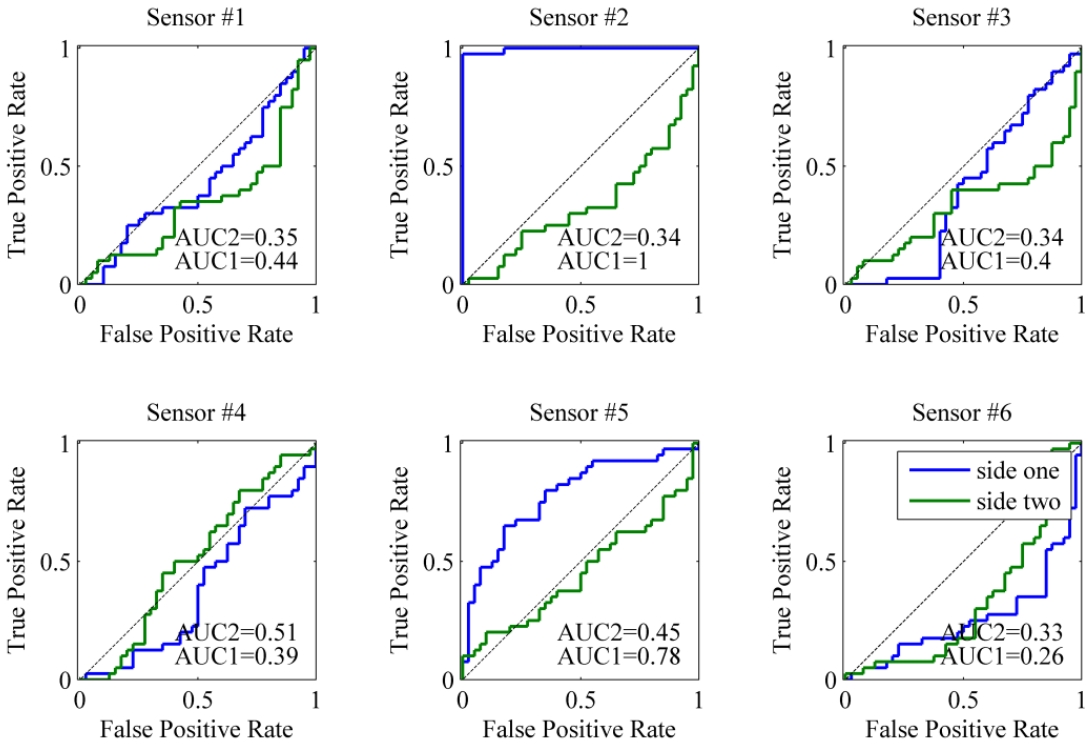

Figure 34: AR parameter analysis ROC curves, damage type 7 worn bearing 5.75 mm on arm #1

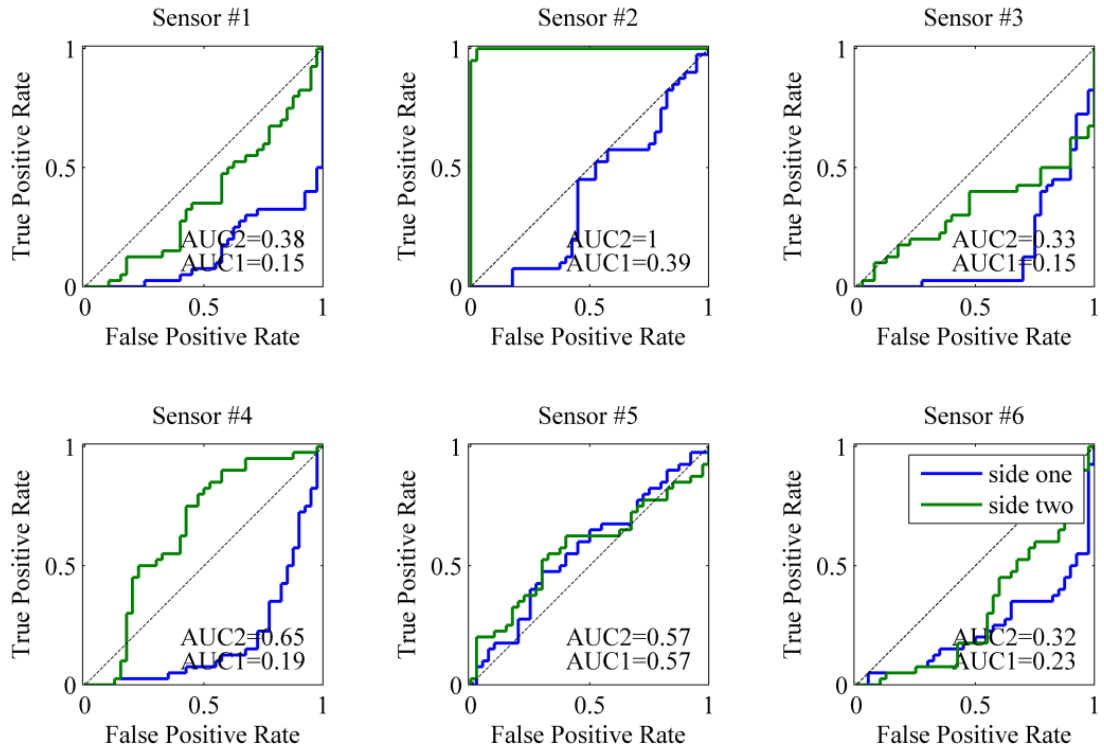

Figure 35: AR parameter analysis ROC curves, damage type 8 worn bearing 5.75 mm on arm #2

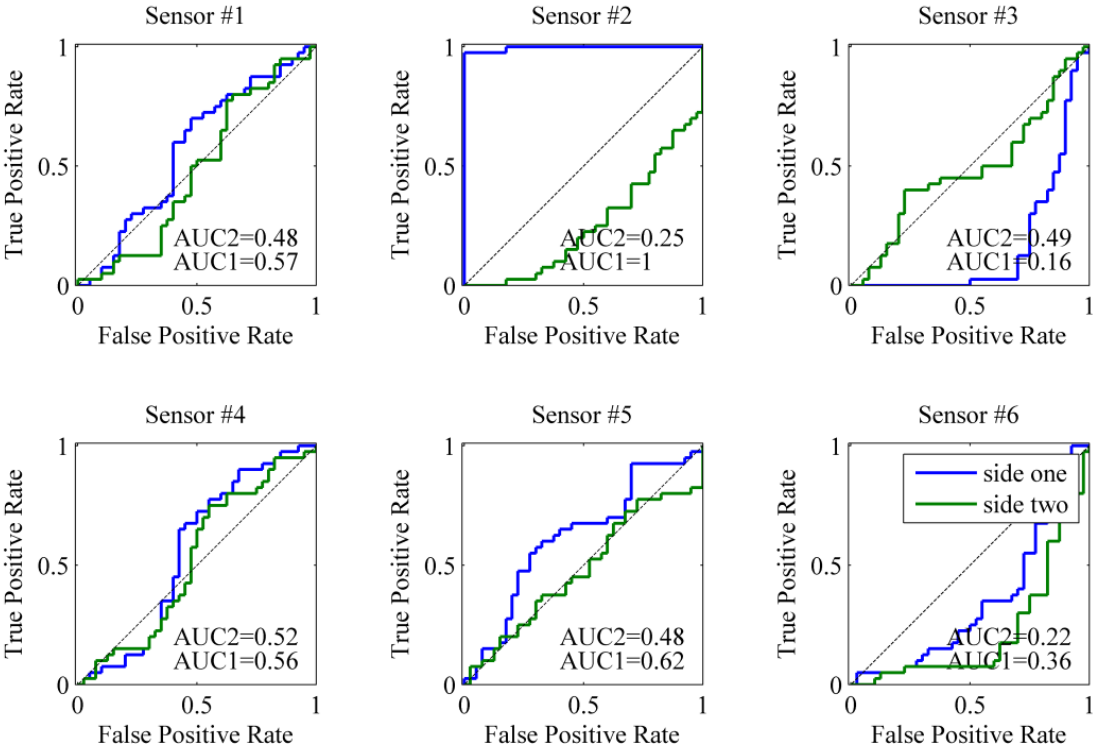

Figure 36: AR parameter analysis ROC curves, damage type 9 worn bearing 5.5 mm on arm #1

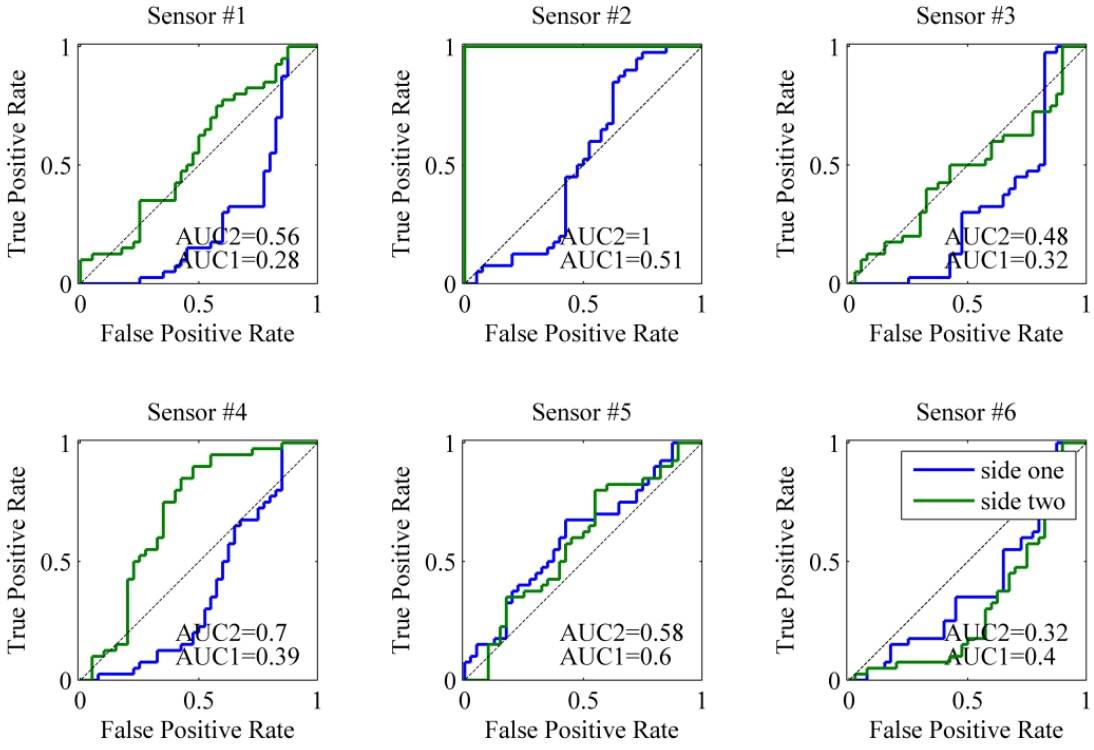

Figure 37: AR parameter analysis ROC curves, damage type 10 worn bearing 5.5 mm on arm #2

The previous four figures, AR ROC curves for damage types 7-10 (swinging arm bearing damage), show similar results to the ROC curves for the main bearing damage. Sensor two is in the best location to detect the damage. When the damage moves from one side of the machine to the other, that side's ROC curve detects the damage, as damage severity increases the damage detection performance indicator increases as well. These continuing trends help verify the validity of AR parametric analysis.

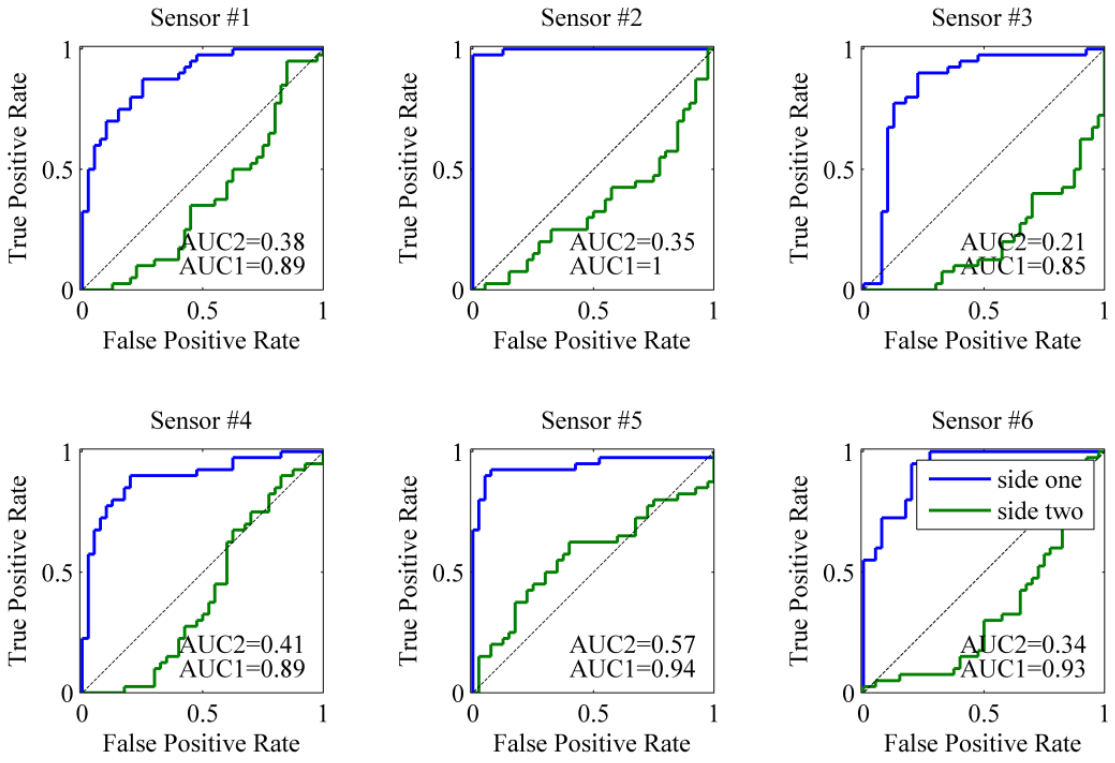

Figure 38: AR parameter analysis ROC curves, damage type 13 loose bearing screw on arm #1

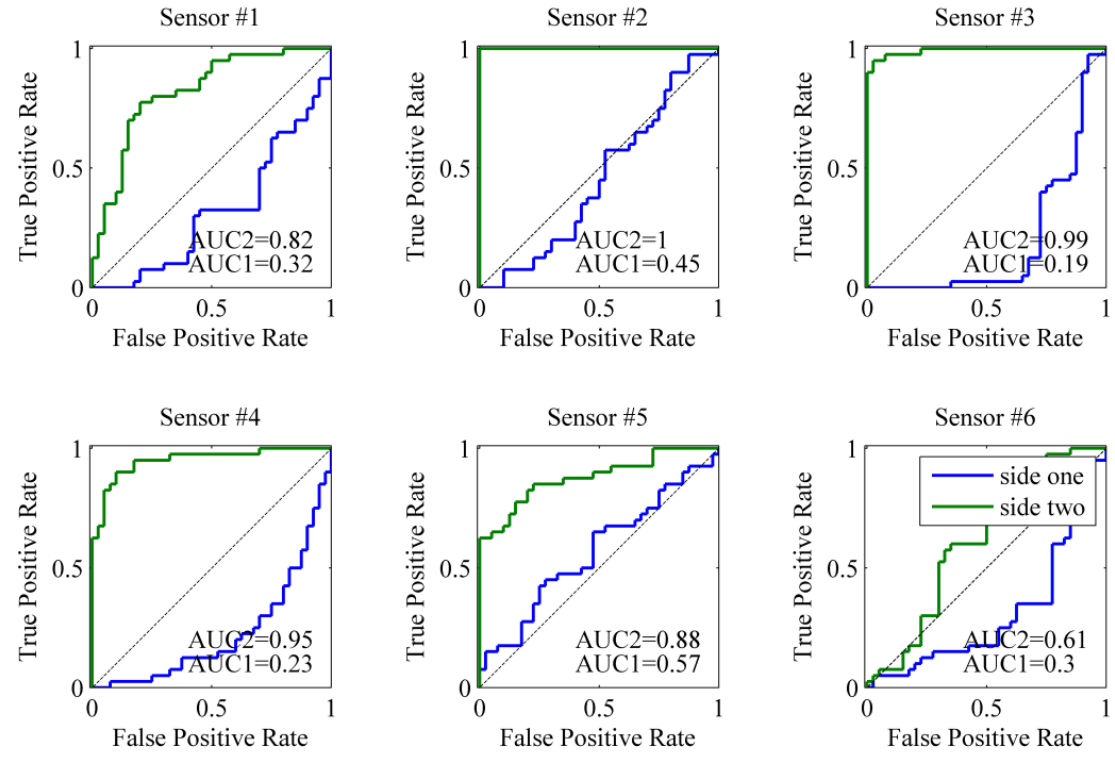

Figure 39: AR parameter analysis ROC curves, damage type 14 loose bearing screw on arm #2

The ROC curves for damage types 13 and 14 show that AR parametric damage detection can sense a loose bearing screw ideally in sensor location #2 and very well at all other sensor locations.

4.6. Realization Algorithm (RA)

4.6.1. Motivation

Generally, pulsed based realization algorithms are based off an unreliable biased estimate of the impulse response coefficients {g(k)} [32]. This makes the results of the model less desirable, as they are an estimation that relies on biased data. This flaw in pulsed-based realization algorithms is not applicable for this project as the data from the safety mechanism under consideration is an impulse. Impulse #1 is caused by a short duration, high energy impact in the machine which creates broad-band excitation, and the measured impulse. Accurate estimates of the impulse response coefficients can be pulled directly from the output data of impulse #1 making the pulse-based realization algorithm suitable for this project. To help further validate the choice of a RA, several common parametric models were created from the data in this project and compared.

All common parametric models under consideration are the pulse-based realization algorithm, box jenkins, ARMAX, ARX, finite impulse response, and output error. These models were compared using frequency analysis to an empirical transfer function estimate (ETFE) to gauge performance. An ETFE model is an appropriate model to gauge performance since it accounts for multiple frequency inputs and has modest data reduction (modest loss of information) when creating the model [31] thus creating an accurate representation of the frequency content in the transfer function of the system. However, an ETFE model is not a preferable model to use for damage detection in this thesis as it is non-parametric and the

variance in an ETFE model cannot be eliminated. When comparing models the effect of changing model order should be taken into account, in this analysis all models were created for different model orders ranging from two to thirty. When comparing bode plots of these models the pulse- based realization algorithm matched the resonance poles indicated by the ETFE model closest across all model orders for frequencies higher than 10,000 Hz. A bode plot for all models with an order of *n*=14 is shown below.

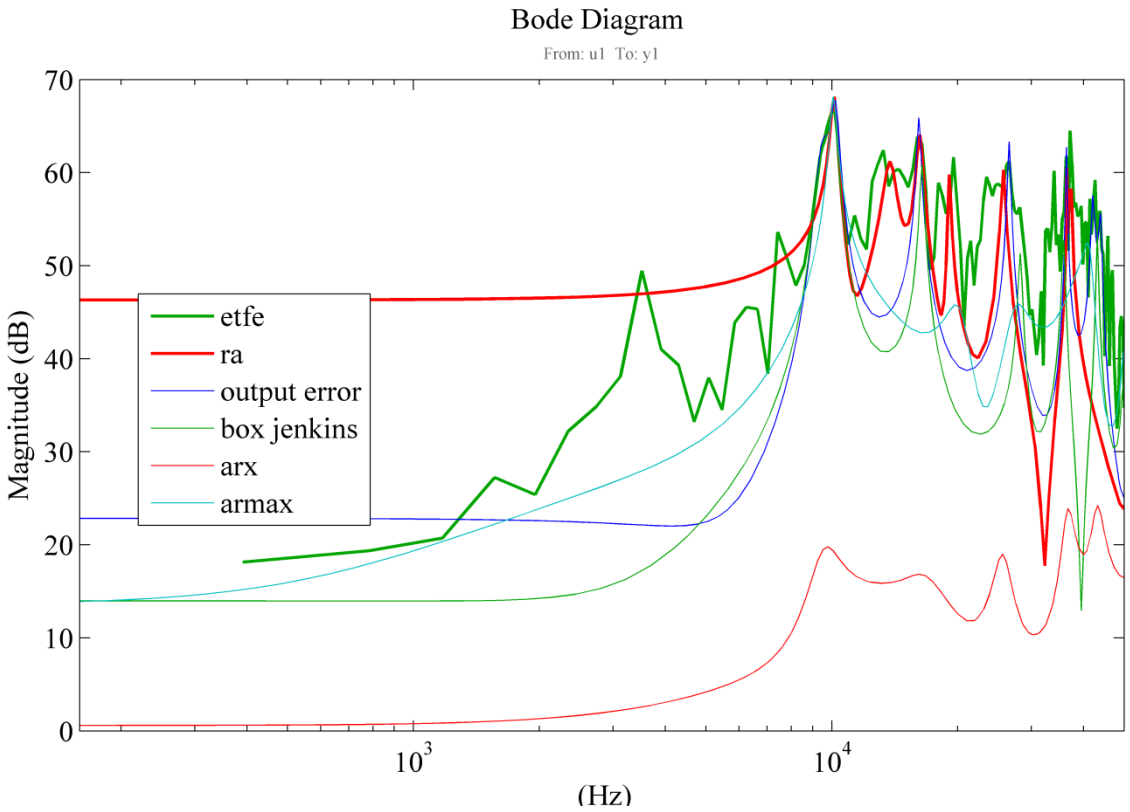

Figure 40: Bode diagrams all models, n=14

Comparing the model's bode diagrams revealed that low frequencies (<10,000 Hz) were not modeled accurately, this was attributed to the fact that the input was not an ideal impulse and possibly output mainly high frequency content.

4.6.2. Pulse-Based Realization Algorithm

The pulse-based realization algorithm, specifically Ho-Kalman's Algorithm and Kung's algorithm were used to obtain the steady state model from the impulse response coefficients. These algorithms and their equations were referenced from [32]. Ho-Kalman's Algorithm states that the impulse response coefficients {g(k)} can be directly placed into a Hankel matrix and a shifted Hankel matrix, shown below.

Equation 4: Hankel matrices

$$H := \begin{bmatrix} g(1) & g(2) & \cdots & g(N_2) \\ g(2) & g(3) & \cdots & g(N_2+1) \\ \vdots & \vdots & \vdots & \vdots \\ g(N_1) & g(N_1+1) & \cdots & g(N_1+N_2-1) \end{bmatrix} \overrightarrow{H} := \begin{bmatrix} g(2) & g(3) & \cdots & g(N_2+1) \\ g(3) & g(4) & \cdots & g(N_2+2) \\ \vdots & \vdots & \vdots & \vdots \\ g(N_1+1) & g(N_1+2) & \cdots & g(N_1+N_2) \end{bmatrix}$$

Where $N_1 + N_2 \leq N$ and $n \leq N_1 \leq N_2$. The Hankel matrix H can be decomposed into $H = H_1 H_2$where H_1 & H_2 are full column and row rank observability and controllability matrices. This full row rank decomposition can be computed using Kung's algorithm via singular value decomposition $H = U\Sigma V^T$. The diagonal entries of Σ are known as the singular values. The singular values give insight to the order of the system. The number of non-zero singular values indicates the order n of the system; however if noise is present in the measurements of the impulse response coefficients, then the singular values can be plotted to see the order of the system. Pictured below is a plot of the singular values obtained from one of the training events.

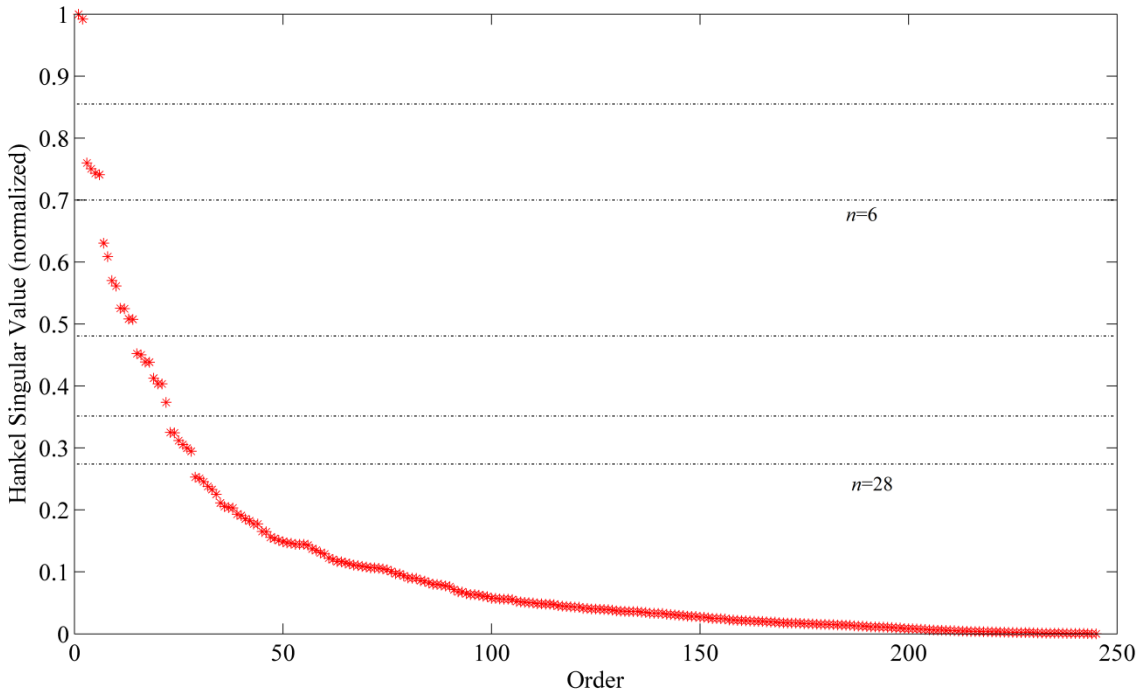

Figure 41: Singular value plot with potential model order lines

The noise in the system creates what appears to be a higher order system when observing the singular values. In this case it is difficult to tell exactly which singular values are due to noise and which are due to actual orders of the system. Using prior knowledge of the mechanical system and by observing the trend of the singular values, five possible different values for *n* were drawn above. Further analysis on model order is discussed later in this thesis. The Hankel matrix is then further decomposed to only include the entries relevant to the chosen model order *n*.

Equation 5: Henkel Singular Value Decompisition

$$H = U\Sigma V^T = [U_n \quad U_s]\begin{bmatrix}\Sigma_n & 0\\ 0 & \Sigma_s\end{bmatrix}\begin{bmatrix}V^T{}_n\\ V^T{}_s\end{bmatrix}$$

Kung's algorithm next lists the formulas necessary to move from this singular value decomposition to the estimated steady state matrices $\hat{A}$,$\hat{B}$, $\hat{C}$, and $\hat{D}$. These formulas are listed below, where $\hat{A}$ $\hat{B}$ $\hat{C}$ and $\hat{D}$ are the estimated steady state matrices.

Equation 6: Kung's algorithm

$$H_1 = U_n \sum_n{}^{1/2} \qquad H_1{}^{\dagger} = \sum_n{}^{-1/2} U_n{}^{T}$$

$$H_2 = \sum_n{}^{1/2} V_n{}^{T} \qquad H_2{}^{\dagger} = V_n \sum_n{}^{-1/2}$$

Equation 7: Kung's algorithm

$$\hat{D} = g(0), \hat{C} = H_1(1,:), \hat{B} = H_2(:,1) \; and \; \hat{A} = H_1{}^{\dagger} \vec{H} H_2{}^{\dagger}$$

The information pertaining to these algorithms, as well as their formulas, was referenced from [32].

4.6.3. Model Analysis

The realization algorithm was performed on the training data first with a choice of model order n=6 to form the estimated steady state model. This model order was chosen based off the second line drawn in Figure 41. To gauge how well the steady state model estimated the true system a simulated output was plotted on top of the measured output. A bode diagram of the model was also plotted on top of the Empirical Transfer Function Estimate (ETFE) bode diagram of the measured data, both figures are below.

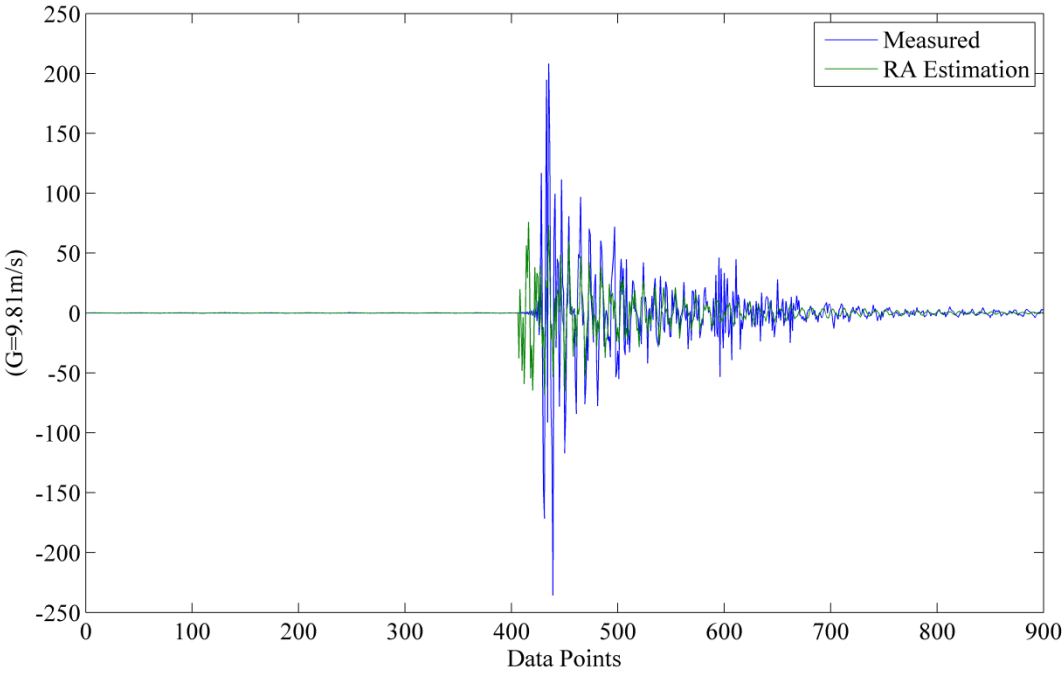

Figure 42: Measured vs. simulated time series, realization algorithm model order n=6

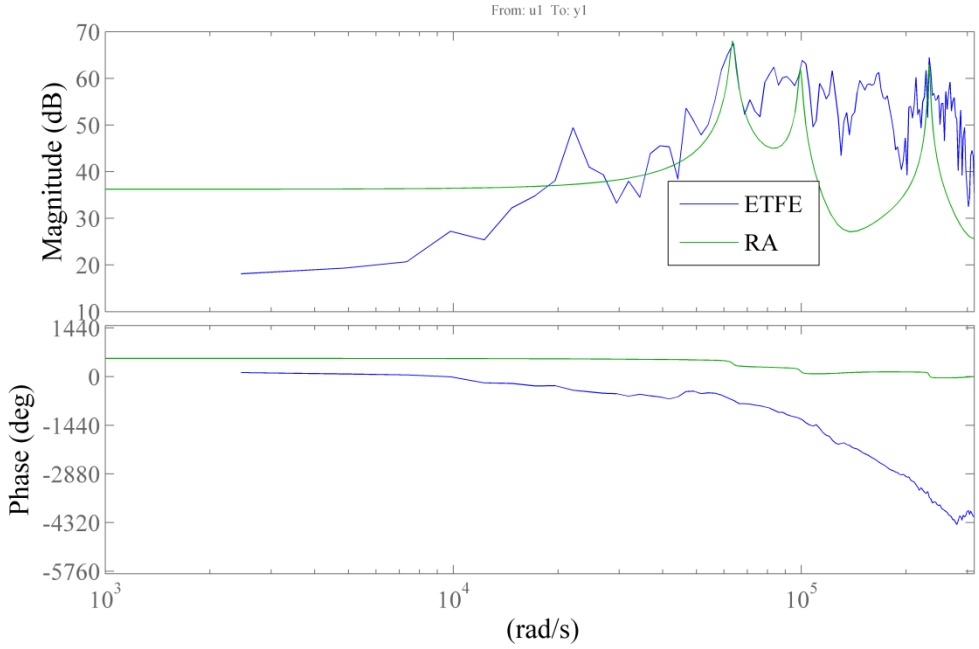

Figure 43: Bode diagram realization algorithm model order n=6

Using a model order of 6 to estimate the state space model produced a model that did not simulate the output or follow the dynamics of the system well. The system order had to be reconsidered, looking at the plot of sigma values and potential model orders, a higher model order was chosen. A model order of 28 proved to simulate the output and model system dynamics best. The simulated output and bode plot for model order n=28 are shown below in Figure 44 and Figure 45.

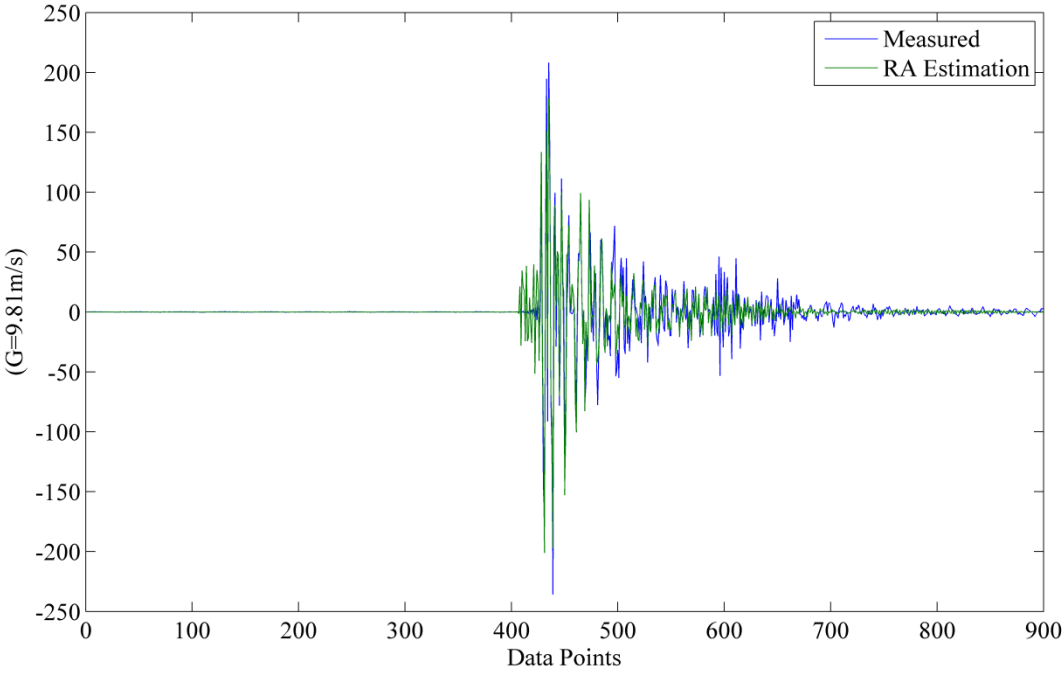

Figure 44: Measured vs. simulated time series, realization algorithm model order n=28

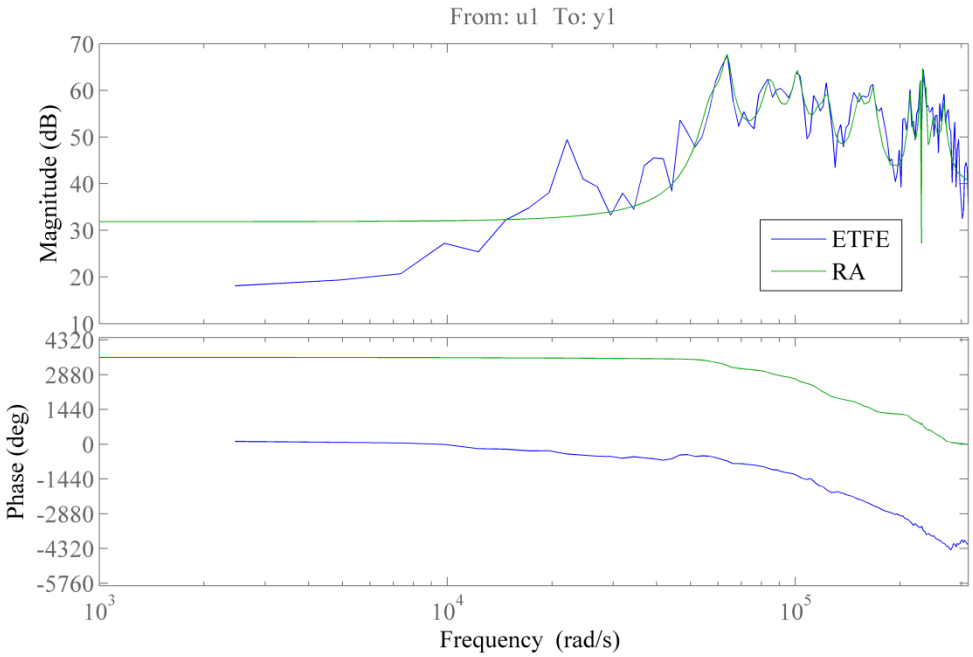

Figure 45: Bode diagram realization algorithm model order n=28

After a satisfactory model was created for the training data, the Realization Algorithm was then repeated creating models for all damaged, un-damaged, and remaining training data. Certain properties were drawn from the model and were compared using the same damage detection strategy described in section 4.6.2. The properties drawn from the estimated steady state model for use in damage detection comparison were the system's natural frequencies and dampening factors, singular values and the singular values normalized. The damage detection performance indicator AUC was used to select the optimal property to compare for damage detection. Singular values from the singular value decomposition proved to detect damage the best across all damage types, regardless of model order. The singular values detected damage the best, when compared to other properties in the RA model as they contained the largest amount and least processed information about the system . The RA was completed using data

for impulse #1 in each event. The same analysis was then repeated for impulse #3; however no improvements in damage detection were shown.

4.6.4. Discussion and Results

The complete RA damage detection analysis was repeated for a set of model orders ranging from 2 to 200. The performance indicator was recorded for each model order and for each damage type, and is displayed below in Figure 46. This figure shows damage detection performance as model order changes for all 18 damage types. It shows that damage types 11, 12 and 15-18 are detected very well with the RA model regardless of its order.

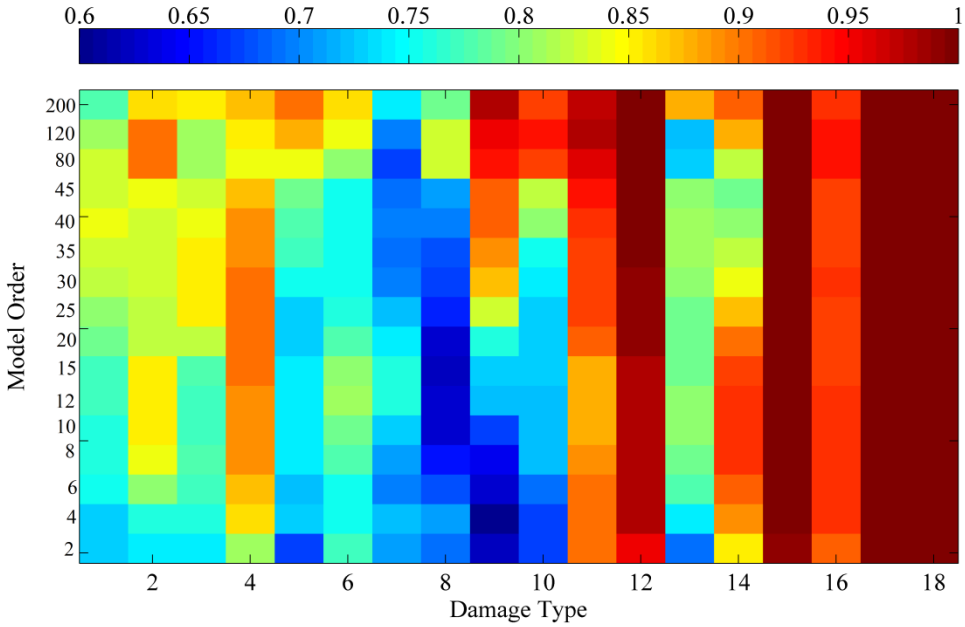

Figure 46: Performance indicator (AUC) for RA models

The results from Figure 46 were compared to damage detection results from AR and Peak amplitude analysis. RA performed best of the three analyses at predicting damage types

11 - ball and spring force reduced, 12- no ball and spring, 15- No lubrication, and 16- tolerance stack up. The resulting ROC curves for these damage types are shown below.

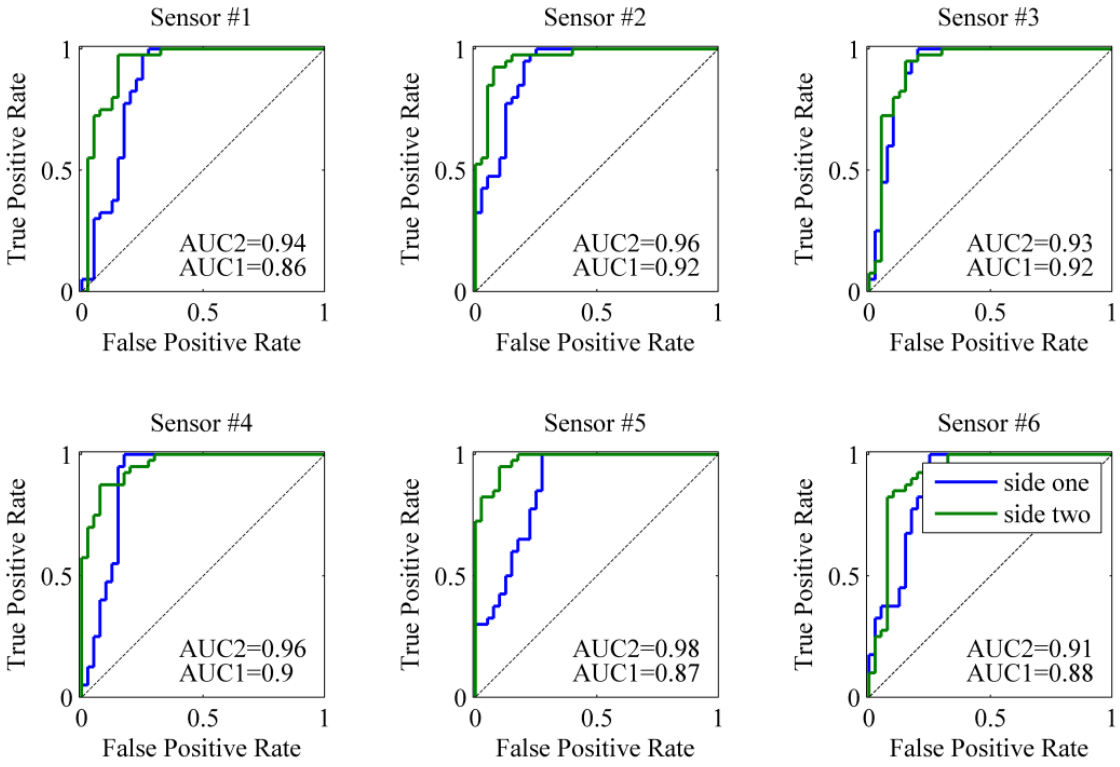

Figure 47: RA singular value ROC curves, damage type 11 ball and spring force reduction

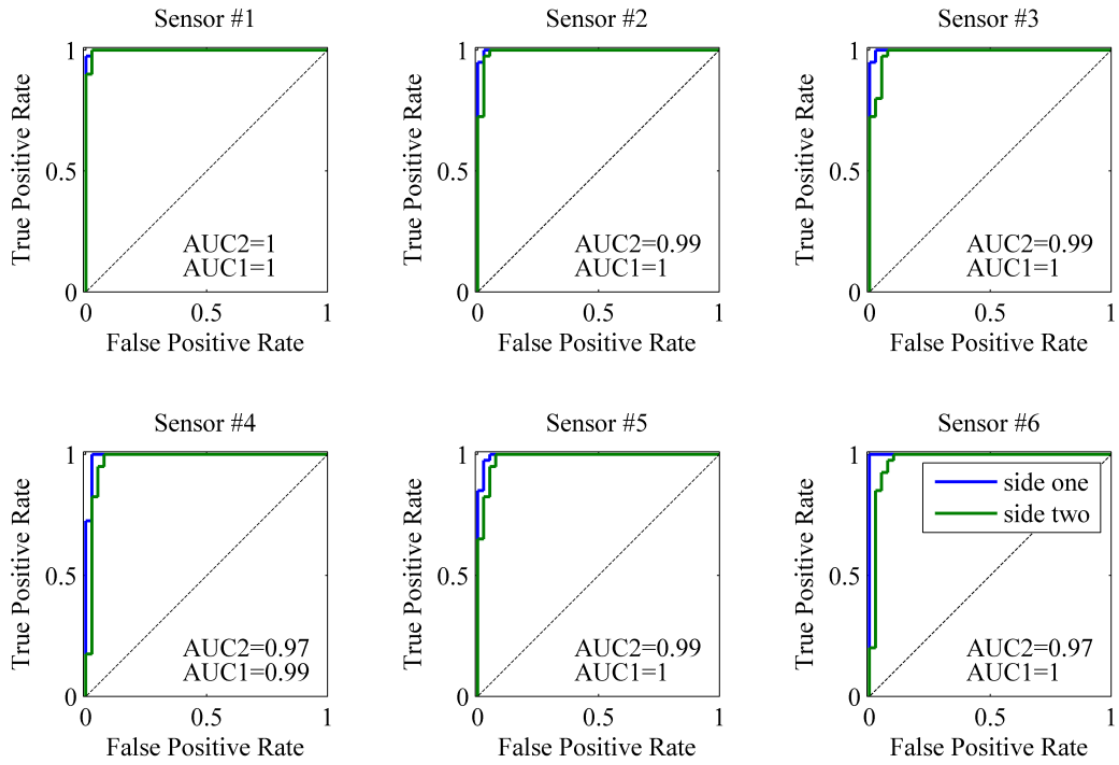

Figure 48: RA singular value roc curves, damage type 12 ball and spring removed

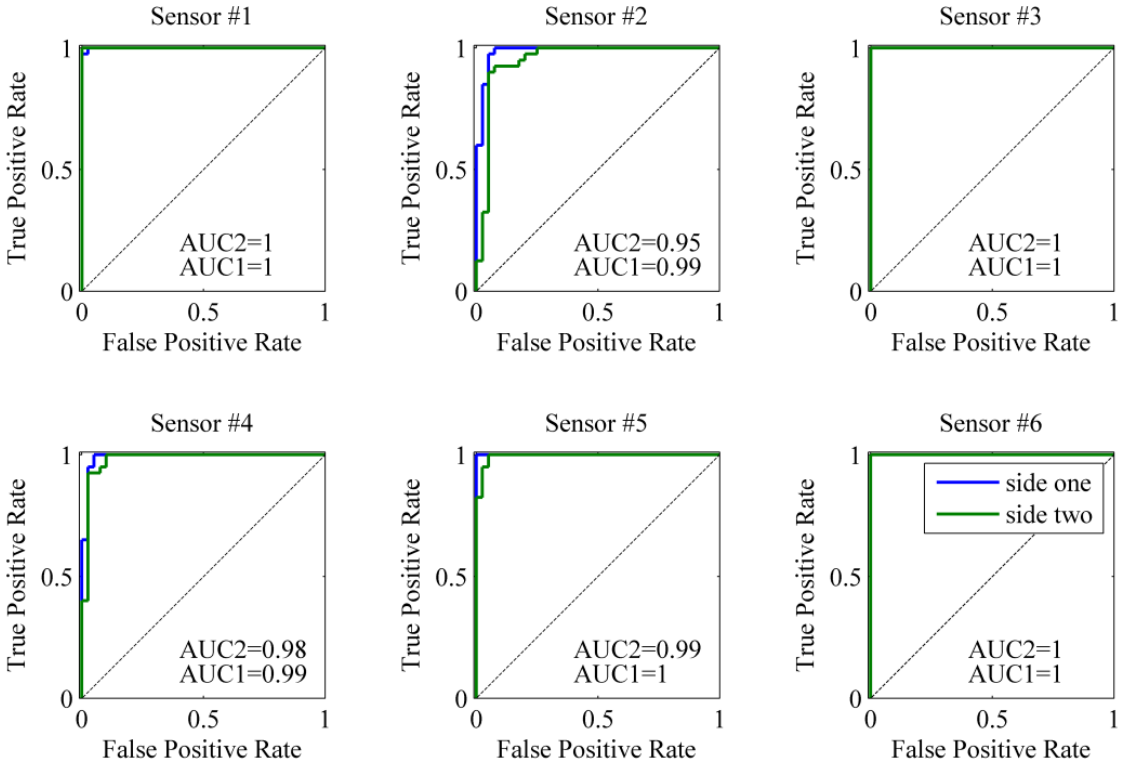

Figure 49: RA singular value ROC curves, damage type 15 no lubrication

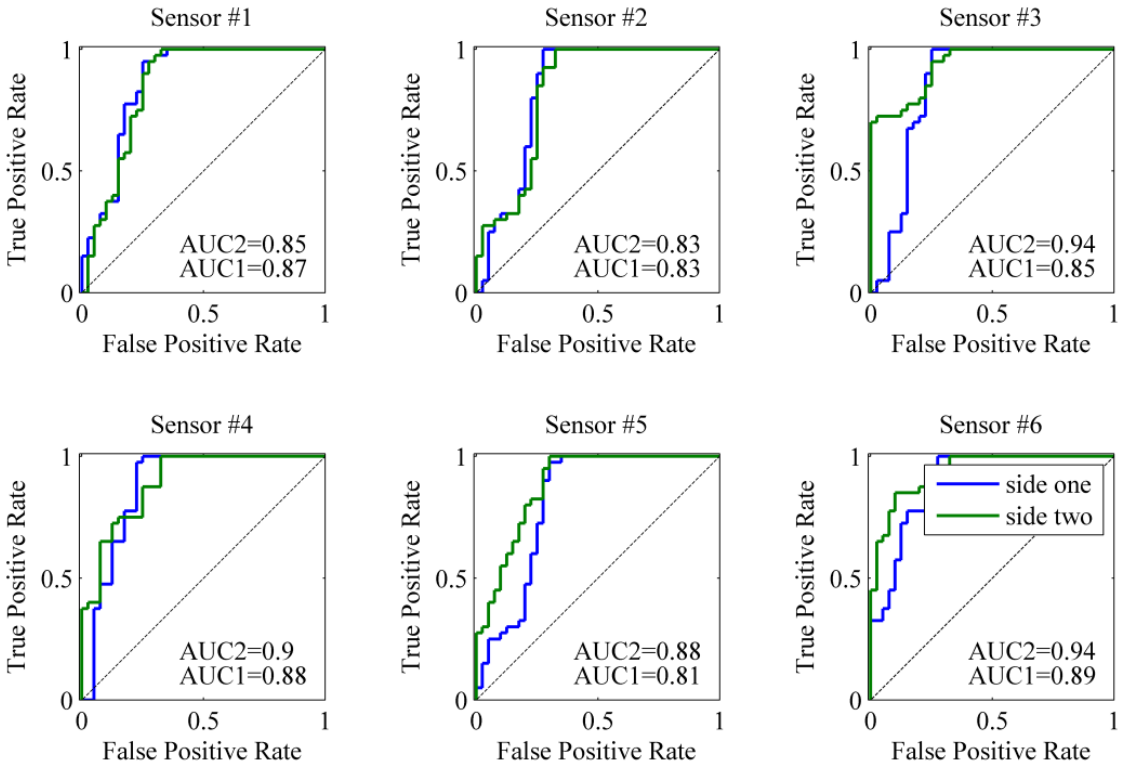

Figure 50: RA singular value, ROC curves, damage type 16-tolerance stack up

Chapter 5
Conclusions, Recommendations, and Future Work

5.1. Summary

Implementing SHM on the designed safety mechanism is quite promising at detecting damage. SHM enabled inspection to occur within the sealed environment of the previously un-inspectable mechanism. All 18 damage types attempted were detected, and while there are plenty more damage types not tested for, it is reasonable to assume that SHM is a viable solution for detecting the majority of damage types this mechanism could encounter.

Disassembling and reassembling the machine in its un-damaged state caused variation in the extracted features that was as high as the variation caused by installing damaged parts, complicating damage detection. This problem was overcome by collecting a large sample of baseline events that were used to train the damage detection analyses to differentiate variation caused by assembly from variation caused by damage. Using the intelligent distance measure, Mahalanobis Distance, improved this differentiation.

The final ROC curves showed that there was a range of thresholds that could be chosen to sufficiently detect damage for all 18 damage types. This project did not address

picking a specific threshold as threshold selection is a regulatory/surety matter established by the DOE client depending on whether minimizing false positive or false negatives are more important. After the weighting criterion is determined they will then have to select a threshold for each of the three analyses suggested in this project, namely RA, Peak amplitude and AR.

Splitting the recorded data by events associated with either side of the machine allowed the damage to be characterized as being present in side one, side two or centrally located (localization as opposed to just detection). This result is seen in the ROC curves for damage types 5-18. If the damage is present on side one of the machine, then the best damage detector is the ROC curve for side one and vice versa. If the damage is present in the main wheel then damage is detected in both side one and side two's ROC curves.

Sensor location analysis showed that sensors in location number two and four performed the best when comparing AUC as a performance metric. This result can be seen below in Figure 51. These two sensors were mounted in the same plane as all the rotating components of the machine. The figure below shows the damage detection performance versus sensor #.

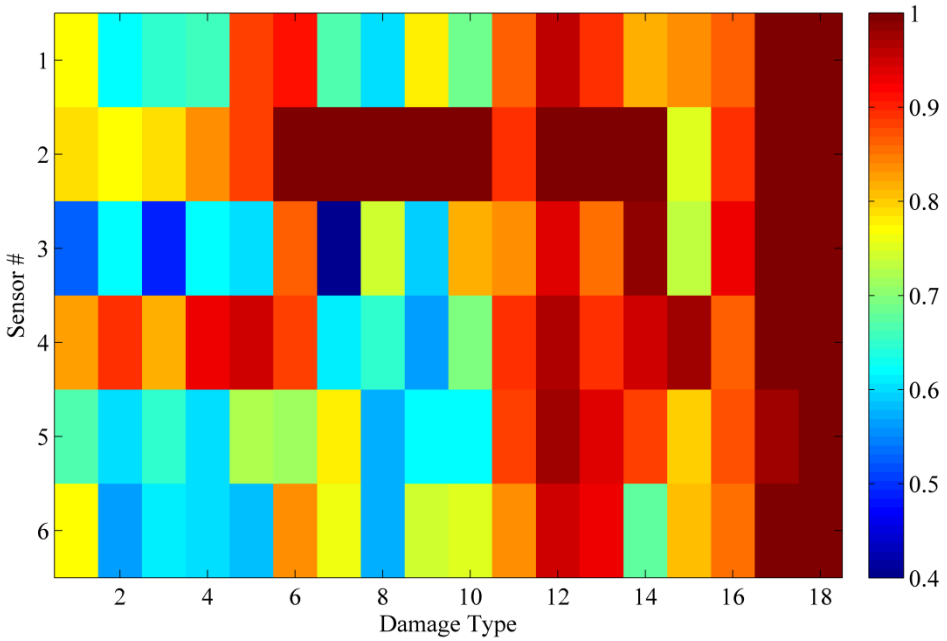

Figure 51: AUC sensor # vs. damage type

The damage detection results for the deformed spring and the spring with the lowered spring constant (types 1-4) were the damage types least sensitive to detection. These results are an area of concern not only for their low sensitivity but also because their location could not successfully be characterized. Only two sensors were able to detect any change in the signal from the baseline data for these damage types.

5.2. Recommendations and Future Work

Any potential owners of this application should pursue implementing SHM on their devices as it will enable them to improve their service life predictions and save costs associated with disassembly. Only two sensors would be required mounted in locations similar to sensors #2 and #4 attached with the same method. If the number of sensors is an issue it would be recommended to only use one sensor, sensor #2. The features peak amplitude, auto-

regressive model parameters and pulse based realization algorithm model parameters all should be extracted. The same analysis described in chapter four can be used to find the range of thresholds necessary to detect damage. The tradeoff of minimizing false positives or false negatives will need to then be considered to pick out the specific threshold values used in the final damage detection code for each feature type.

Implementing SHM for this application could be improved upon in future work in a variety of ways. Damage types 1-4 provided the least sensitive damage detection and did not follow characterization patterns that all other damage types did. Another analysis method like frequency analysis could be tried to see if it can sense the change in resonant frequencies caused by the change in spring constant associated with damage types 1-4 more accurately. An additional improvement would be to have multiple devices manufactured to ensure that variability associated with manufacturing can be accounted for in the damage detection process as in this thesis only one device was manufactured and tested. More advanced damage characterizations can be accomplished with further research of thresholds. Current analysis shows that damage can be localized as occurring in one of three quadrants in the device- the two sides and center. It is suspected that specific damage types can be characterized by observing the way features vary.

Appendix A

A.1. Motor Controller Code

EZ commander: (motor control code – input)

/1s0A0m1,1L3000,3000V6000,6000P3500,3500e1R<CR>

/1s1m70,70L9000,9000V40000,40000H02e2R<CR>

/1s2J2P,3200J1P6400J2P,9600J1P3200J2P,3200J1P3200J2P,3200J1P3200J2P,6400J1P9600J

2P,3200J1P3200J2P,3200J1P3200J0H02e2R<CR>

A.2. Figures

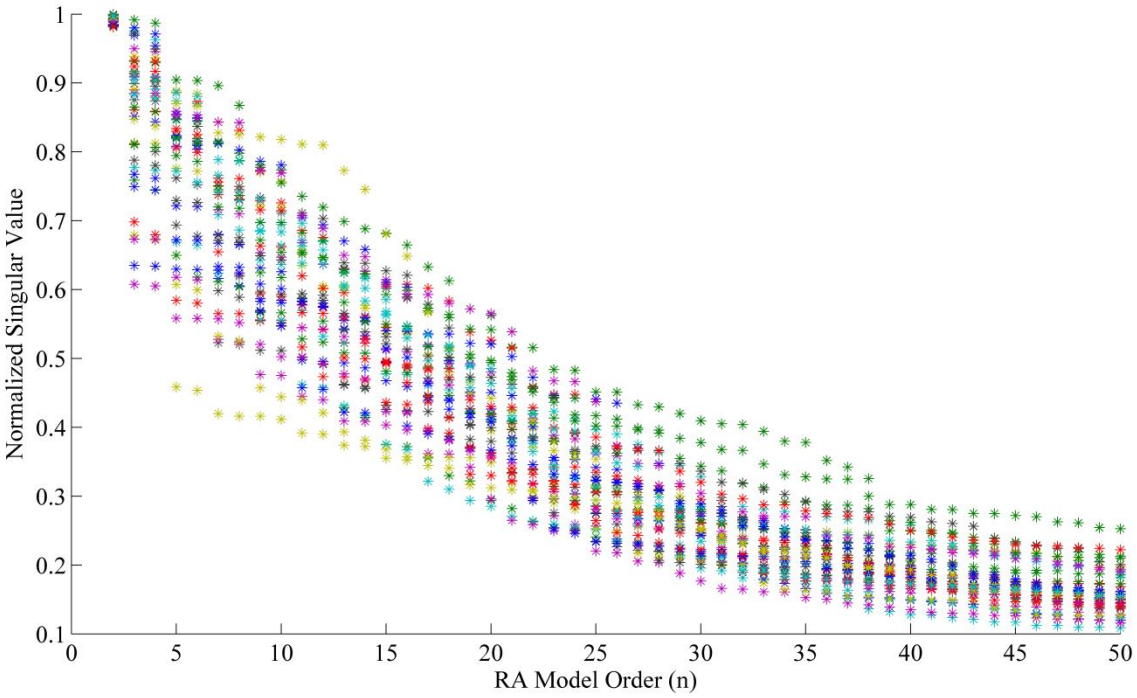

Figure 52: Singular Value Variance all 35 training events sensor 1 side 1 event 1

References

1. Chang, Fu-Kuo. Structural health monitoring 2000. CRC, 1999.

2. Bengtsson, Marcus. "Condition based maintenance system technology–Where is development heading." CONDITION BASED MAINTENANCE SYSTEMS–AN INVESTIGATION OF TECHNICAL CONSTITUENTS AND ORGANIZATIONAL ASPECTS (2004): 55.

3. Farrar CR, Worden K. An introduction to structural health monitoring. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2007;365(1851):303–315.

4. Farrar CR, Doebling SW, Nix DA. Vibration–based structural damage identification. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences. 2001 Jan 15;359(1778):131-149.

5. Fugate, Michael L., Hoon Sohn, and Charles R. Farrar. "Vibration-based damage detection using statistical process control." Mechanical Systems and Signal Processing 15.4 (2001): 707-721.

6. Shull (Ed.), P., (2002). *Nondestructive evaluation: theory, techniques and applications*, Marcel Dekker, New York.

7. M. D. Todd and E. B. Flynn, "A Bayesian Experimental Design Approach for Structural Health Monitoring," 14th International Symposium on Dynamic Problems of Mechanics (DINAME 2011), Sao Sebastio, Brazil, March 14-18, 2011.

8. Doebling, Scott W., et al. Damage identification and health monitoring of structural and mechanical systems from changes in their vibration characteristics: a literature review.

No. LA--13070-MS. Los Alamos National Lab., NM (United States), 1996.

9. Roebben, Gert, et al. "Impulse excitation apparatus to measure resonant frequencies, elastic moduli, and internal friction at room and high temperature." Review of scientific instruments 68.12 (1997): 4511-4515.

10. Guo, Wei-Hua, Wei-Jun Li, and Yong-Zhen Huang. "Computation of resonant frequencies and quality factors of cavities by FDTD technique and Padé approximation." Microwave and Wireless Components Letters, IEEE 11.5 (2001): 223-225.

11. Scott A. Ouellette ; David D. L. Mascarenas ; Michael D. Todd; Corrosion-enabled powering approach for structural health monitoring sensor networks. Proc. SPIE 7288, Active and Passive Smart Structures and Integrated Systems 2009, 728821 (April 06, 2009); doi:10.1117/12.815397.

12. M. Todd, D. Mascarenas, E. Flynn, T. Rosing, B. Lee, D. Musiani, S. Dasgupta, S. Kpotufe, D. Hsu, R. Gupta, G. Park, T. Overly, M. Nothnagel, C. Farrar, "A different approach to sensor networking for SHM: Remote powering and interrogation with unmanned aerial vehicles", Workshop on Structural Health Monitoring, 2007

13. Farrar, Charles R., and Hoon Sohn. "Pattern recognition for structural health monitoring." Workshop on Mitigation of Earthquake Disaster by Advanced Technologies. 2000.

14. McFadden, P. D., and M. M. Toozhy. "Application of synchronous averaging to vibration monitoring of rolling element bearings." Mechanical Systems and Signal Processing 14.6 (2000): 891-906.

15. Tandon, N., and A. Choudhury. "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings." Tribology International 32.8 (1999): 469-480.

16. Orhan, Sadettin, Nizami Aktürk, and Veli Celik. "Vibration monitoring for defect diagnosis of rolling element bearings as a predictive maintenance tool: Comprehensive case studies." Ndt & E International 39.4 (2006): 293-298.

17. Heng, R. B. W., and M. J. M. Nor. "Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition." Applied Acoustics 53.1 (1998): 211-226.

18. Powell, George. "Exploded view of bearing" Photo. bonesbearings.com @2013. 20 Feb. 2013 < http://bonesbearings.com/support/maintenance/>.

19. "Ball and spring plunger" Photo. springplungers.com @2006 19 Feb 2013 <http://www.springplungers.com/images/detail/spring-ball-plungers/standard-ball-plungers-without-locking-element-la.jpg>

20. *Stepper Motor Basics.* (2001) Retrieved June 28th 2012, from <http://www.solarbotics.net/library/pdflib/pdf/motorbas.pdf>

21. AllMotion. (May 2011) Programming instructions for Models EZHR17EN,

EZHR23ENHC, and EZ4AXIS. Retrieved Sep. 2012 from <http://www.allmotion.com/PDF_Datasheets/Command_Set_EZHR17EN.pdf>

22. Chu, Anthony. (2012 June 12). Choosing the Right Accelerometer. Engineers Circle. retrieved from <http://www.meas-spec.com/downloads/Choosing_the_Right_Type_of_Accelerometers.pdf>

23. -Anon (Nov. 1999). *Vibration Sensors*. SKF Condition Monitoring. Retrieved from < http://www.skf.com/files/260860.pdf>

24. Xiang, Shiming, Feiping Nie, and Changshui Zhang. "Learning a Mahalanobis distance metric for data clustering and classification." Pattern Recognition 41.12 (2008): 3600-3612.

25. Wicklin, Rick. "Mahalanobis distance" Photo. blogs.sas.com 15 Feb. 2012. 20 Feb 2013 <http://blogs.sas.com/content/iml/files/2012/02/mahal.png>

26. mahal. (2012). In *Mathworks Help online.* Retrieved Oct. 15th 2012 from <http://www.mathworks.com/help/stats/mahal.html>

27. De Maesschalck, Roy, Delphine Jouan-Rimbaud, and D. L. Massart. "The mahalanobis distance." *Chemometrics and Intelligent Laboratory Systems* 50.1 (2000): 1-18.

28. Tom Fawcett, An introduction to ROC analysis, Pattern Recognition Letters, Volume 27, Issue 8, June 2006, Pages 861-874, ISSN 0167-8655, 10.1016/j.patrec.2005.10.010. (http://www.sciencedirect.com/science/article/pii/S016786550500303X) Keywords: ROC analysis; Classifier evaluation; Evaluation metrics

29. Hanley, James A., and Barbara J. McNeil. "A method of comparing the areas under receiver operating characteristic curves derived from the same cases."*Radiology* 148.3 (1983): 839-843.

30. Braga, Ana Cristina, and Pedro Oliveira. "Diagnostic analysis based on ROC curves: theory and applications in medicine." International Journal of Health Care Quality Assurance 16.4 (2003): 191-198.

31. Ljung, L. *System identification: theory for the user*. Upper Saddle River, NJ: Prentice-Hall. 1999 Print.

32. De Callafon, R., Moaveni, B., Conte, J., He, X., and Udd, E. (2008). ”General Realization Algorithm for Modal Identification of Linear Dynamic Systems.” J. Eng. Mech., 134(9), 712–722. doi: 10.1061/(ASCE)0733-9399(2008)134:9(712)