**2012 Summer Research Project Report**

**Early Gearbox Fault Detection via Auto-Regressive Models in the Time Domain constructed from Vibrational Data**  (Draft Version)

**Abstract:**

Vibration-based testing and diagnostic techniques have been widely applied to detect and prevent the degradation and deterioration of continuously functioning gearboxes. The typical set up of this technique involves the collection of vibration signal data over discrete time periods obtained from an array of sensors mounted onto the shell of the gearbox. During this time frame, the gearbox is subjected to load conditions exceeding the load-bearing capacities of the machine.  
  
The vibration data collected is used to create an early fault detection scheme for the machine. In order to process this data, one must first identify the correct meshing frequency pertaining to the rotating gears. After that, the technique of time-synchronous-averaging is used to reduce noise and produce a cleaner signal.  
  
Once the signal is processed, an auto-regressive model is constructed from the data collected from the healthy state. Subtracting the AR model prediction values from the healthy state observations we obtain residual values of the vibration signal. A standard 3-sigma control chart based on the residual data in the healthy state is set up and used to monitor the behavior of the machine leading to early detection of failure.    
  
The result of this research is beneficial in many industries including companies which operate machines with a critical function that must comply to a guaranteed runtime prior to failure. Such industries include, aerospace, manufacturing and military.

**Introduction: (To Be Completed)**

**Data Collection / Determining sampling frequency**

During the data retrieval process, raw vibrational signals is directly imported from file A02 into MATLAB. Before the technique of time synchronous averaging can be applied to clean the signal, we must identify the time intervals where synchronous cyclic repetition occur.

Before the collected data can be interpreted and analyzed, the correct [K value, # of data points pertaining to one cycle of the target gear] must be obtained in order for the data to be processed correctly. The [K value] is a property of the mechanical configuration of the gearbox, and is used to isolate the target gear for failure analysis.

To identify the desired [K value] of the target gear we use the following equation (reference derivation).

[Equation 1] K = (fs/fm)\*Nt

After obtaining the K value of the gearbox configuration, the collected raw data must be sectioned into partitions in each the size of the amount of data prescribed by the meshing frequency. We begin by assuming that the natural frequency of the target gear should be embedded in the raw signal. The technique of time synchronous averaging is then applied to remove the white noise caused by gear shaft imbalance signals, and load variation signatures etc. (ref Yang).

As evidenced from the raw signal file (1)

**TSA, using peak-to-peak intervals**

Due to the fact that the gear configuration of the gearbox in in contact with one another, the varying frequencies excluding the white noise contribution would be considered to be an effect of the varying harmonics within the gearbox, the dominant frequency being the meshing frequency of the gear (875 Hz). Therefore, we can expect to observe periodic behavior within each moment triggered data file.

The intervals of periodic behavior mark the beginning of repetition within the signal, where we can subsequently use time synchronous averaging across each repetition to produce a cleaner signal with reduced noise.

In order to find the areas where cyclic repetition occurs, we must first obtain the peak to peak intervals within each data file collected we use a peak-to-peak detection within MATLAB. Furthermore, we take into account the nature of the data files, which consist of 10 second windows provided by moment triggered mechanisms built into the sensors. Therefore, we expect each 10 second window to contain synchronous data.

However, when performing the peak to peak algorithm independently on each 10 second window, the intervals where peak to peak detection occur are very inconsistent. This is due to the fact that white noise produces randomly distributed amplitudes within the signal.

To produce a cleaner peak to peak detection signal, with consistent intervals, we take the averaged signature values over 60 files, and then run a peak to peak algorithm on it. The results produce a more refined and precise interval locations, with less variance. Using these location points we initiate the time synchronous averaging algorithm.

**Determine RMS of TSA, explain significance of RMS**

Time synchronous averaging (TSA) involves averaging the periodic data over the range corresponding to the number of data points representing one cycle of the target gear [K]. Each interval pertaining to one cycle corresponds to K = 1051 data points, initiate at the time locations provided by the peak to peak algorithm. This ensures that each packet of time data is in sync with the next period. Subsequently for each 10 second window, we average periodic intervals corresponding to K = 1051 data points, at approximately 190 synchronous starting points.

Once the TSA model is produced, each 10 second window is subsequently summarized into a collection of K (1050.8 rounded to 1051) data points corresponding to the number of data points relative to each cycle of the target gear. Notice that each 10 second window consisting of 200,000 data points is compressed into one packet consisting of K = 1051 data points. Thereafter, each compressed data packed produced by TSA is plotted together continuously for the purpose of comparison (note that the actual data collected by the 10 second time frame is not continuous in time).

**RMS**

For purposes of simplification, we obtain the Root Mean Square (RMS) value of each TSA packet. Where RMS is the measure of varying vibrational quantity indicative of the energy of the vibration in each target gear cycle. As we see that the RMS values increases, the process become increasingly out of control.

**Obtain AIC lag**

The next step is to produce an Auto-Regressive model to simulate the future performance of the gearbox in the healthy state. We use the AR model to model the RMS values corresponding the TSA packets to predict the future performance of the gearbox.

In order to select an optimal AR model, we produce candidate AR models from lag order 1 to 15.

We first assume that the data collected at RMS files 1 to 40 functioning gearbox is stationary model in the healthy state. Next the optimal order (or lag value) must be obtained to create the most suitable Auto-Regressive model. We use the Akaike Information Criterion to obtain the optimal lag order corresponding to the minimum descriptive length/information loss of the RMS data. Using the System Identification Toolbox in MALAB, We obtain the lag order 1 as the optimal AR model with the lowest AIC value. However, we also see that the highest lag order, arbitrarily chosen to be 15, produces the AR model which fits the testing data best. Keeping in mind these lag orders, we can use it to generate an AR model.

**Construct AR Model**

Using the aforementioned lag order to construct an auto-regressive model (lag 1 & 15), we utilize the “aryule” function in MATLAB. The “aryule” function produces the coefficients corresponding to the lag order of the AR model. Subsequently the AR Model is plotted in comparison with the actual RS values.

**Obtain Residuals, and set up 3 sigma bounds.**

The RMS values of the observed data A02 is subsequently subtracted from the AR model prediction of the data to obtain the error of the prediction, defined as the residual value. We stipulate that the first 40 residual values pertain to the gearbox running in healthy state, which we denote as the training data. And thus, the files from 41 to 136 corresponding to the gearbox testing data, is utilized to monitor the condition of the gearbox.

Calculating the standard deviation of the residual data in the training phase we obtain the 3 sigma bounds which correspond to the entire run time. Evaluating the values of the RMS residual data, we can determine whether the process is in control or not. Residual points successive of that point fall outside of the control bounds, indicating that maintenance of the gearbox should be performed in order to have the machine persist operation in the healthy state.

Observing the lag order 1 residuals, we can see that file #77 corresponding to file 328 (time stamp value 1997/11/24  10:00:21 ) .

Observing the lag order 15 residuals, we can see that file #72 corresponding to file 323 (time stamp value 1997/11/24  07:30:21 ).

Effectively, these are the times where the machine should undergo maintenance in order to ensure future stability.

Failure locations:

@ Lag 1:

**File #77:** A0200328.010

**Duration:** 62.0 hrs

@ Lag 15:

**File #72: ‘**A0200323.01’

**Duration:** 59.5 hrs