The Application of Time-Frequency Methods to HUMS

Anna H. Pryor
NASA Ames Research Center
Moffett Field, CA 94035
M.S. 269-3
apryor@mail.arc.nasa.gov

Marianne Mosher
NASA Ames Research Center
Moffett Field, CA 94035
M.S. 269-3
mmosher@mail.arc.nasa.gov

David G. Lewicki
U.S. Army Research Laboratory
NASA Glenn Research Center
Cleveland, OH
david.g.lewicki@grc.nasa.gov

Abstract

This paper reports the study of four time-frequency transforms applied to vibration signals and presents a new metric for comparing them for fault detection. The four methods to be described and compared are the Short Time Frequency Transform (STFT), the Choi-Williams Distribution (WV-CW), the Continuous Wavelet Transform (CWT) and the Discrete Wavelet Transform (DWT). Vibration data of bevel gear tooth fatigue cracks, under a variety of operating load levels, are analyzed using these methods. The new metric for automatic fault detection is developed and can be produced from any systematic numerical representation of the vibration signals. This new metric reveals indications of gear damage with all of the methods on this data set. Analysis with the CWT detects mechanical problems with the test rig not found with the other transforms. The WV-CW and CWT use considerably more resources than the STFT and the DWT. More testing of the new metric is needed to determine its value for automatic fault detection and to develop methods of setting the threshold for the metric.

Introduction

Many metrics based on frequency analysis are currently used on vibration data to detect faults from gear boxes [1]. Although these methods have been shown to find faults [2], traditional methods such as spectrum analysis, waterfall plot, cepstrum analysis, and matched filtering [3], were developed for use on stationary data. However, many systems are not stationary. For instance, in Health and Usage Monitoring Systems (HUMS) [4], the vibrational frequencies being measured by the accelerometers can change rapidly in time, especially if a fault has occurred. Many types of gear damage produce localized changes in the signal so that the signal is no longer stationary on the time-scale of the gear tooth meshing. The signal near the meshing with a defective tooth may vary considerably from the rest of the signal. If the situation is critical, it is important to determine the severity so that corrective actions can be taken.

In standard Fourier analysis, for example, a signal is decomposed into individual frequencies. Unfortunately, there is no way to determine at what point in time each of those frequencies has occurred. However, there are signal processing methods which give local information about both time and frequency. These methods localize faults in time and therefore may be more sensitive to early changes in the signal due to impending faults.

Many time-frequency (TF) methods have been applied to the detection of faults in gears. Wang has claimed the Short Time Fourier Transform (STFT) to be a powerful tool in detecting local gear damage at an early state [5]. In periodic data, strong harmonics may obscure small transient events. According to one study, STFTs will perform better here than conventional methods [6]. Others have investigated STFT methods for the early detection of faults in gears [7, 8]. With the Wigner-Ville Distribution (WV-CW), local tooth faults in spur gears have been found with this method [1]. Other researchers have been inspired by WV-CW and have modified it to create their own methods [9, 10, 6]. The Continuous Wavelet Transform (CWT) is becoming studied more and its usefulness increases. In one study, the phase map of the wavelet transform was found to have distinctive features near a tooth that was cracked [11]. In another study, satisfactory results were obtained from a de-noising method based on wavelet analysis used for the diagnosis of mechanical vibration signals [12]. Many others use the CWT as applied to the detection of faults in gears [13, 5, 14, 15, 16, 17, 18]. The Discrete Wavelet Transform (DWT) or Multi Resolution Analysis has also had some good results with gear faults. In one study, when the data was preprocessed using the DWT and then fed into a multilayer neural network, gear faults were successfully found and classified into different groups [19]. In another study, the DWT was used to create a residual error. The probability density function of the residual error was then expanded into a Hermite polyno-

Presented at the American Helicopter Society's 57th Annual Forum, Washington D.C., May 9-11, 2001. Copyright ©2001 by the American Helicopter Society. All rights reserved.

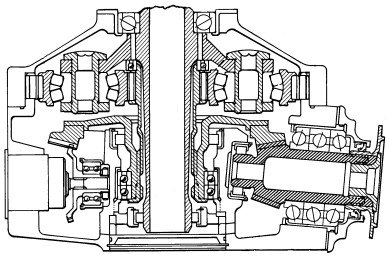

Figure 1: OH-58A Helicopter Main-Rotor Transmission.

mial and the coefficients were used as a feature vector to estimate early fatigue cracks in gears [3]. Other researchers have had both mixed and good results using the DWT in gear fault detection [20, 4].

It is extremely difficult to use a TF method as it stands alone for HUMS analysis. The dimensionality is simply too large. In addition, there is no obvious fault detection criteria, other than human observation, without some dimensionality reduction. Some authors have reduced the dimensionality without creating a metric [12, 8]. Other researchers have reduced the input into a neural net and used it as a classifier [19, 1]. Because of the way the nets were trained, the nets were over-trained on a few specific examples of faults and therefore may not find faults that differ from any in the training set. Finally, some authors have both reduced the dimensionality and created metrics [13, 3, 21, 4, 6, 7]. For example, in [6], they use a generalized and improved variation of Choi-Williams called RID. They analyzed a spectrogram with singular value decomposition. Then they looked at scatter plots of the first two principal components. Williams also examined a metric produced by projecting onto what Williams calls the “zero-subspace” and what the current authors interpret as being the noise subspace for the normal and faulted data. A squared distance measure between the projection of faulty and normal signals was then studied.

These four methods will be compared by first analyzing them on a set of experimental data known as the Multiple-Seeded Fault Data. Then the dimensionality of their results will be reduced by using singular value decomposition (SVD). Finally all four methods will be compared with each other using a simple one-dimensional metric developed by the authors. The following are included: the assumptions, resources, resolution, speed of calculation and results of analyses.

Description of Test and Analysis

Test

The vibration data reported in the current work were acquired from tests performed at the NASA Glenn Research Center. The tests were performed on an OH-58A helicopter main-rotor transmission (Fig. 1) in the Glenn 500-hp Helicopter Transmission Test Facility as seen in Fig. 2. The objective of the tests was to supply experimental data to study the detection of gear tooth fatigue cracks and their propagation rates under a variety of operating load levels. The gear under study is the spiral bevel pinion with 19 teeth (Fig. 3a). Nine notches were fabricated on

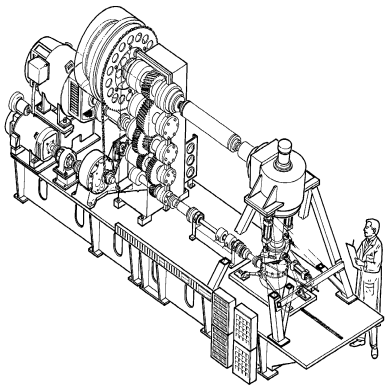

Figure 2: NASA Glenn 500-hp Helicopter Transmission Test Facility.

(a) normal teeth

(b) fractured teeth

Figure 3: OH-58A Spiral-Bevel Pinion

various teeth of the pinion. The notch sizes varied from 0.5 to 0.1" in length, and 0.015 to 0.045" in depth. The pinion was run 4.9 million cycles at 6060 rpm and torque levels from 2479 to 4649 in-lb (80, 100, 125 and 150% design load), see Table 1. After 1.9 million cycles at 4649 in-lb, five teeth fractured off the pinion as seen in Fig. 3b. More information about the transmission and test rig can be found in Lewicki and Coy [22].

During the tests, vibration data were periodically collected from the transmission. Five piezoelectric accelerometers were mounted on the OH-58A transmission housing as shown in Fig. 4. The accelerometers were light-weight, high-frequency, and had integral electronics. They had a nominal resonant frequency of 90 kHz and an output sensitivity of 10 mV/g. Measurements from accelerometer numbers 1, 2, 4 and 5 were used in the current study. Accelerometers 1 and 2 were closest to the pinion gear on the input shaft. Accelerometer number 1 is oriented radially to the input shaft and accelerometer number 2 is tangential to the input shaft. Accelerometer number 4 is about

Table 1: OH-58A Pinion Operating Conditions.

Data Records	1-43	44-126	127-209	210-449
Time (cycles/million)	1	1	1	1.9
Speed (rpm)	6060	6060	6060	6060
Torque (in-lb)	2479	3099	3874	4649
Torque (%)	80	100	125	150

45 deg away from the pinion oriented radial to the planetary gear. Accelerometer number 5 is by the output shaft about 90 deg from the input shaft oriented in the same direction as number 1. The synchronously averaged signal from accelerometer number 1 showed consistently higher amplitude than signals from the other accelerometers. The results presented in this paper will concentrate on measurements from accelerometer number 1. Because the pinion contained 19 teeth, the vibration signals from the healthy gear are expected to contain a waveform with 19 repetition per rotation and significant frequency components at 19 times the shaft rotation and its harmonics.

The vibration data were recorded on magnetic tape using high-frequency wideband group II electronics approximately every 180,000 cycles (every 30 minutes). The data were then post processed. Data for four of the accelerometers were interpolated to 1024 points per rotation of the pinion and then synchronously averaged. By synchronously averaging, the components of the signal not synchronous with the pinion rotation, both noise and signals from other hardware, are reduced. The vibration signal is relatively stationary on the time scale of the gear rotation because the test rig is operated under very tight conditions, so averaging on the time-scale of the gear rotation is reasonable. Local signal components due to gear damage will be preserved as long as the signal repeats each rotation of the gear.

Overview of Analysis

The current study started with the synchronously averaged time histories. The data were then decimated by an order of four to 256 points for the analyses in this paper. It was not practical to perform some analyses on the longer blocksizes. In addition, the higher frequencies were not needed to find changes in the data due to gear damage.

In this study, the term time-frequency will be used. The representations of the data are actually sampled at uniform phase spacing of the gear instead of uniform time. This keeps the signal in phase with the gear teeth even though the rpm may change. Also, some of the transforms do not transform into frequency, but something very similar to frequency.

The analysis in this study projects the synchronously averaged time history onto a time-frequency space. All time-frequency transforms in this study except the orthogonal wavelet transform, project the vibration signal onto a space of much higher dimension and thus produce a very redundant rep-

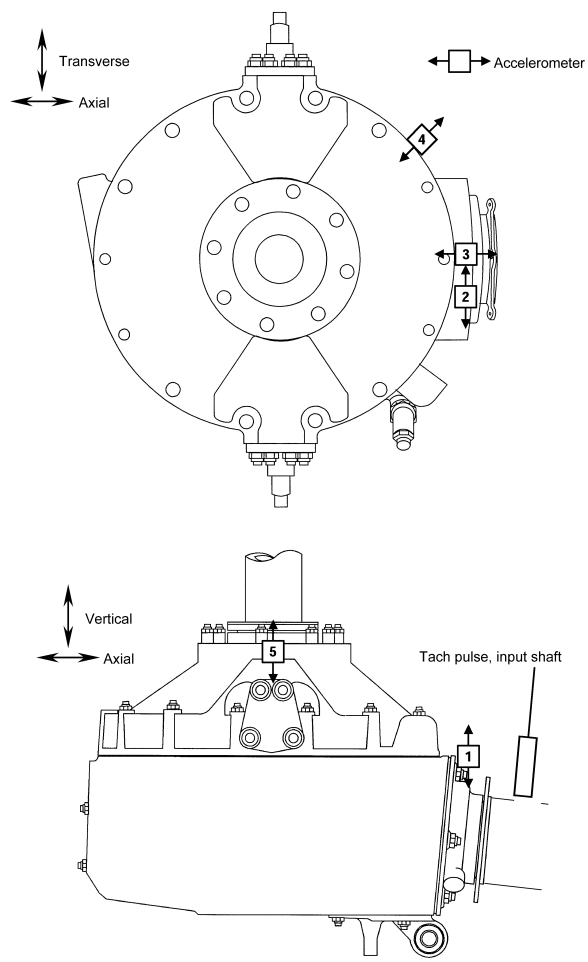

Figure 4: Accelerometer Locations on OH-58A Transmission.

resentation of the signal. The elements in the time-frequency space are signals with localized frequency and time content. Local features in the vibration signal will be well represented in the time-frequency space. A model will be produced by projecting the signal representation onto a very low-dimensional subspace. This low-dimensional subspace of the very large time-frequency space is optimized to represent the signal. A metric that determines faults will be produced based upon how closely the projection of the transformed signal onto the very small subspace matches the original signal. This metric is a measure of the distance of a signal from the model.

All of the computational analyses were done with the mathematical software package MATLAB.

Time-Frequency Methods

Formulas for continuous time transforms offer a more concise notation than those of discrete time. Although this paper contains continuous time formulas, the discrete time analogs were used for the computation. As far as a bit of notation, if $f(x)$ is a function, then $\overline{f}(x)$ is its complex conjugate.

Short Time Fourier Transform (STFT)

The Short Time Fourier Transform (STFT), also known as the Windowed Fourier Transform or spectrogram, is a development that extends standard Fourier Transform techniques to handle non-stationary data [23]. Fourier Transforms are applied to short windows of data. These windows are moved along the data and may overlap. This transforms one-dimensional data into two-dimensional data, one dimension for frequency and a second dimension for window location in the data. The STFT gives information for a fixed frequency and time resolution dependent on the window. An impulsive event appears in an STFT as increased levels for all frequencies at the time of the impulse. The limitation of the STFT is that signal components having a poor match to the fixed time and frequency resolution will be obscured.

Given a time interval $T > 0$, let $g(u)$ be a function that vanishes outside the interval $-T \leq$u≤ 0. The STFT of a function f(u) is defined to be

$$\widetilde{f}(\omega,t) = \int_{-\infty}^{\infty} \overline{g}(u-t)\, f(u) e^{-2\pi i\omega u} du. \tag{1}$$

The analyses in this paper use a Hanning window of 32 points in length with 24 points of overlap.

Choi-Williams Distribution

The Wigner-Ville Distribution (WV) is a nonlinear transform that maps one-dimensional data into two dimensions. One dimension has frequency-like characteristics and the other dimension has time-like characteristics. Integrating the WV along the frequency-like dimension produces the squared amplitude of the original signal. Integrating the WV along the time-like dimension produces the squared amplitude of the Fourier Transform of the original signal.

The Choi-Williams Distribution (WV-CW) is a modification of the Wigner-Ville Distribution[24, 25]. As one of the first time frequency methods to be used, this distribution has been widely studied. Given a function $f(u)$ the Wigner-Ville Distribution is defined as

$$W(t,\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \overline{f}(t - \frac{1}{2}u)\, f(t + \frac{1}{2}u) e^{-iu\omega} du. \tag{2}$$

Unfortunately, if the signal is even moderately complicated, there will be significant cross-terms which will make interpretation nearly impossible. The Choi-Williams Distribution uses a kernel (an additional multiplied term inside equation 2) which dampens out the cross-terms. While still present, the cross-terms are significantly dampened and no longer cause a problem in analyzing the signal. The kernel that Choi and Williams chose was

$$\phi(\theta,\tau) = e^{-\theta^2\tau^2/\sigma^2} \tag{3}$$

where the parameter σ controls the amount of attenuation. The amplitude of the cross-terms is directly proportional to σ. However, if the cross-terms are suppressed too much, then the auto-terms will lose resolution in the time frequency plane. The final Choi-Williams Distribution is defined as

$$W(t,\omega,\theta,\tau) =$$

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} \overline{f}(t - \frac{1}{2}u)\, f(t + \frac{1}{2}u) e^{-\theta^2\tau^2/\sigma^2} e^{-iu\omega} du \tag{4}$$

For this study $\sigma = .05$.

Continuous Wavelet Transform

The Continuous Wavelet Transform (CWT), is a time-frequency method that builds on the idea of the STFT. Whereas the STFT is limited in resolution because of its fixed window size, the CWT uses a variable window size with short windows for high frequencies and long windows for low frequencies. This allows the CWT to be a powerful tool in representing local features of a signal that other methods such as the STFT may miss entirely.

Given a mother wavelet, all other wavelets are made by dilating and/or translating the mother wavelet. These wavelets are moved along the data and overlap. This transforms the one-dimensional data into two dimensions. The first is for scale (the size of the wavelet) and the second is for the location in the data. For each wavelet, the scales can be converted back to a range of frequencies.

In continuous wavelet analysis [23], we begin with a complex-valued window function $\psi(t)$ called the mother wavelet. Given an arbitrary $p \geq 0$ and any real number $s \neq 0$, the mother wavelet is defined as

$$\psi_s(u) \equiv |s|^{-p} \psi(\frac{u}{s}). \tag{5}$$

In order to describe a given signal at a local time, the signal is compared to translated versions of ψ_s. Given an interval of

length T near $u = 0$, if $\psi(u)$ is supported (non-zero) on T, then near $u = 0$, $\psi_s(u)$ is supported on an interval of length $|s|T$ and the function

$$\psi_{s,t}(u) \equiv \psi_s(u-t) = |s|^{-p}\psi(\frac{u-t}{s}) \tag{6}$$

is supported on $|s|T$ near $u = t$. Assuming that the mother wavelet belongs to $L^2(\mathbf{R})$ then $\psi_{s,t}$ does also. If a given function (signal) $f(t)$ also belongs to $L^2(\mathbf{R})$, then the continuous wavelet transform (CWT) is defined as

$$\widetilde{f}(s,t) \equiv \int_{-\infty}^{\infty} \overline{\psi}_{s,t}(u) f(u) du. \tag{7}$$

For this study the seventh order Daubechies wavelet scaling filter, known as db7 was used on all of the data records.

Discrete Wavelet Transform

Given functions $f(x)$ and $\psi(x)$ such that $\int_{-\infty}^{\infty} \psi(t)dt = 0$, the discrete wavelet transform of $f(x)$ is defined to be [26]

$$\widetilde{f}(a,b) = \int_{-\infty}^{\infty} \psi(a^{-m}t - nb) f(t) dt \tag{8}$$

It is then divided into two cases. In the first case, called redundant discrete systems or frames, the dilation parameter, a, and the translation parameter, b, are discrete. In particular, for a, only powers of one fixed dilation parameter greater than one are used. This discrete system is not used in this paper and will not be discussed further. In the second case, called orthogonal, orthonormal or muti-resolution analysis, specific choices of a, b, and ψ are made so that the wavelets that are created form an orthogonal or orthonormal basis. For example, if $a = 2$ and $b = 1$, then there exists ψ such that

$$\psi_{m,n}(x) = 2^{-m/2}\psi(2^{-m}x - n) \tag{9}$$

form an orthonormal basis for $L^2(\mathbf{R})$.

The Mulit-Resolution Analysis (MRA) is a fast iterative algorithm which analyzes any signal. At each level of decomposition, a signal is high-pass filtered and down-sampled to produce one of the detail signals comprising the wavelet analysis. The same signal is also low-pass filtered and down-sampled to produce the next averaged signal to be used as part of the wavelet analysis. This whole process is repeated on the average signal enough times for the necessary analysis of the signal.

In this study, the seventh order Daubechies wavelet scaling filter, known as db7 was also used on the discrete wavelet transform. There were seven levels evaluated. This produces seven details and one average. The results are distinct from the CWT version of db7. One obvious reason being that the DWT db7 provides an orthonormal basis of wavelets so that there is no redundancy in the representation of the transform, unlike the CWT db7.

Modeling and Metric with SVD

By themselves, the time-frequency analyses alone are not sufficient for detecting faults. All of the time-frequency methods used in this paper, except the discrete wavelet transform, expand the dimensionality of the representation of the data. Although various researchers [1, 3, 4, 5, 7, 6, 15, 17] have found early indications of faults with time-frequency methods, manually inspecting the results of the transforms will not produce a viable method of detecting faults in a HUMS system. Fault detection can be simplified by projecting the transformed data onto a lower dimensional model subspace. A model will be made based upon a time-frequency analysis of a subset of the data called the training set. This technique produces a small number of coefficients that describe the data in the model subspace and a residual that is the difference between the data and the data projected onto the model subspace. The model coefficients and/or the residual can then be used for fault detection. In this work, we will use the ratio of the rms of the residual to the rms of the transform of the data as a metric for fault detection. The value of this metric ranges from 0 for data that is perfectly described by the model to 1 for data that is orthogonal to the model, meaning that it is totally outside the model. This normalized residual will be tracked and when the level increases there is indication of a change in the state of the gear outside the range of the training set.

The bases defining the model will be derived from data in the training set. By deriving the basis vectors for the model from the data rather than a standard expansion such as Fourier Series, data will fit better with fewer terms. This kind of modeling has been used to compress data for a variety of problems [27, 28, 29, 30, 31, 32, 33]. Several methods, Principal Component Analysis, Blind Signal Separation, Karhunen-Loève Transform and Singular Value Decomposition accomplish this modeling with the same mathematics. The modeling is described here within the framework of the Singular Value Decomposition.

For modeling, the time-frequency analysis of each data record is reshaped into a one-dimensional column vector and all of the vectors from the transformed data set are placed into a matrix X. The training set, Y is a subset of X. Y is an IxJ matrix of J data record transforms, I points long. Y is factored with a Singular Value Decomposition (SVD). The SVD produces a diagonal matrix of singular values, S, and two unitary matrices, U and V. These matrices are related by $Y = U * S * V'$. There are $K = \min(I, J)$ singular values. The columns of U form a basis for the column and left null spaces of Y. The columns of V form a basis for the row and right null spaces of Y. The singular values indicate the amount of each basis vector in the matrix Y. A subset W is formed from the columns of U depending on how many of the largest singular values in S are considered significant. The model consists of the projection of the data onto the subset basis, W. The coefficients, D, from this projection now form a compact description of the data analyzed with the time-frequency method,

$$D = W' * X \tag{10}$$

where W' is the Hermitian transpose of W. Each column in D contains the coefficients for the model of the data in the corresponding column of X.

A model representation of the data, M, can then be created

$$M = W * D. \tag{11}$$

Now that there is a model, define the residual, R, to be

$$R = X - M. \tag{12}$$

Each column of $R = (R_1, R_2, ...R_m)$ contains the residual for the corresponding column of the transformed data, $X = (X_1, X_2, ...X_m)$. Define a real-valued metric as the ratio of the rms of the residual to the rms of the transformed data,

$$r(l) = \frac{|R_l|}{|X_l|} \tag{13}$$

This metric measures how each data record differs from the model. High dimensional data have been compressed down to a single number for each data record.

For fault detection, the training set would contain a sample of data records from the gear running without faults. Measurements with no alterations to good gears were not available for this data set. In the current work, the training set contains data records from the earlier part of the test where the gear was expected to contain the least amount of damage. Models were first made using the first 126 data records for the training set. The results showed anomalies in some of the models for the points 27-32 and 44-53. A new training set was selected to contain half the data records from 1-26, 33-43 and 54-126. The resulting model coefficients and residuals for the training set were similar to those for the non-anomalous data in the first 126 data records and not in the training set; thus, giving confidence to the model.

The number of basis vectors or modes to use in the model must be determined for each model. If too few modes are used, the model will poorly represent the transformed vibration signal from the good gear. If too many modes are used, more computation will be required producing more coefficients than needed and the model may over fit the data. Several criteria are available. The number of modes can be chosen to account for a fixed amount of rms or variance in the training set [34]. This method requires selecting a cutoff criteria. The number of modes can be chosen by examining the basis vectors and choosing only those that look like a signal [33]. This method is labor intensive and requires selecting a cutoff criteria. Another way is that the number of modes can be selected by statistical hypothesis testing of the multiplicity of a noise eigenvalue in the singular values to distinguish between noise and signal [35, 36, 37]. The authors chose to use this last method which is commonly used in array signal processing for methods that require knowledge of the number of signals in the data. This method makes the assumption that data are the sum of an unknown number of stationary signals and ergodic Gaussian random noise.

When the individual data records, Y_l, in the matrix Y are random samples from a process containing j signals and random noise of amplitude σ, the first j of the J singular values are estimates of the signal strengths. The rms of the remaining singular values are an estimate of the noise strength, σ. For Gaussian random noise, the remaining lower level singular values come from a normal distribution with a mean of σ. A Minimum Description Length (MDL) is found by taking a maximum likelihood estimate of the parameters. Given a vector, s, of the K singular values, the estimate of the number of signals is the value of j that minimizes the MDL

$$MDL(j) = -L(j) + \frac{j(2D - j - 1)}{2} \ln(J). \tag{14}$$

where

$$L(j) = J(K - j) \ln \left[\frac{\left(\prod_{j+1}^{K} s^2 \right)^{\frac{1}{(K-j)}}}{\frac{1}{(K-j)} \sum_{j+1}^{K} s^2} \right] \tag{15}$$

In determining the number of degrees of freedom in the MDL calculation, the number of points in a data record, $D = 256$, was used, instead of $D = M$ for the data length. This was done because most time-frequency methods produce redundant information from the original shorter data record.

Results

Three data records were chosen to be representative of what occurred during the major points of the test. Although nine of the teeth had been notched in the beginning, it was assumed that for a period of time that the system would run in a less damaged state. Data record 25 was used to represent this period of time. In the middle of the test, it was assumed that some of the teeth would begin to break off or become severely damaged. Data record 250 was used to represent this middle part of the test. The most severe damage would occur near the end of the test. Data record 449 was used to represent this last part of the test.

The analyses were done on four accelerometers. The main discussion for each transform will cover accelerometer number 1 followed by a brief discussion of the other accelerometers.

Time History

There are a total of 449 data records with torque ranging from 80% to 150%. Figure 5 shows a sample of data records from different parts of the test recorded on accelerometer number 1.

In Fig. 5a and b, data record 25 and 100, torque 80% and 100%, the data clearly shows a waveform with a repetition of 19 throughout one rotation corresponding to the 19 teeth on the gear. In Fig. 5c, data record 150 with torque 125%, there is less clear definition of the signal between 0 and 0.2 of the rotation and between 0.7 and 1. This could be signals from adjacent teeth beginning to merge. In Fig. 5d, at data record 230 and 150% torque, at approximately 0.1, 0.78, and 0.9 of the rotation, the wavelength of the waveform appears to have doubled. In Fig. 5e, at data point 250 and 150% torque, the structure is now completely different and the amplitude has increased. By the final data record, Fig. 5f, data record 449, torque 150%, it appears that there is now a structure of approximately 9 per rotation with almost 6 times the original amplitude.

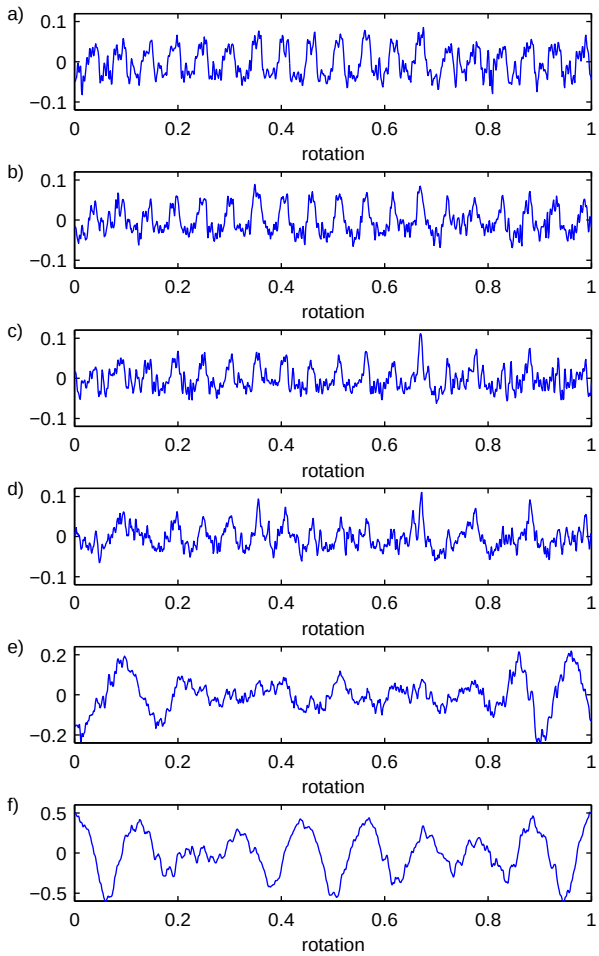

Figure 5: Time data for Accelerometer 1 a) data record 25, 80% torque, b) data record 100, 100% torque, c) data record 150, 125% torque, d) data record 230, 150% torque, e) data record 250, 150% torque, f) data record 449, 150% torque

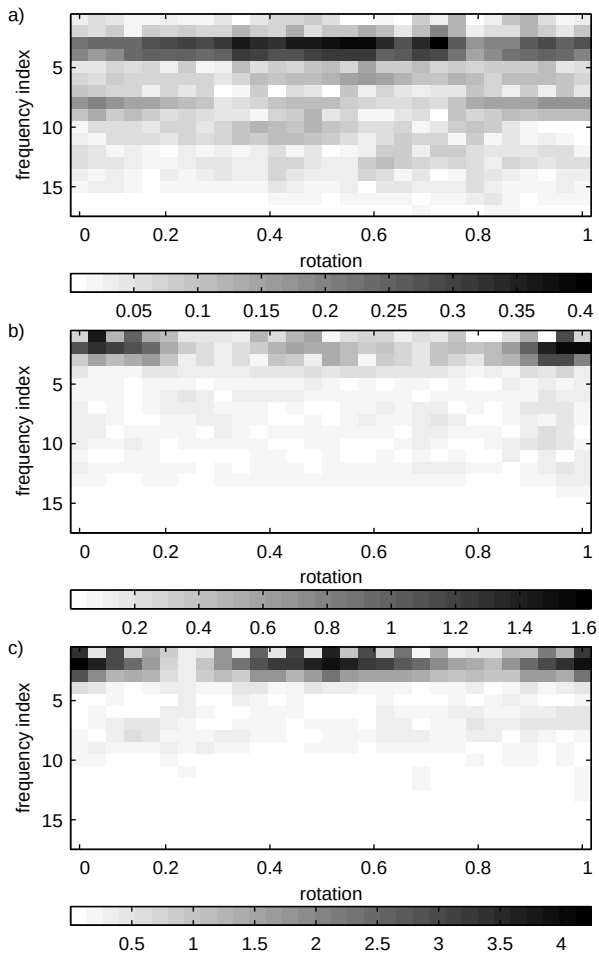

Figure 6: Short Time Fourier Transform for Accelerometer 1 a) data record 25, 80% torque, b) data record 250, 150% torque, c) data record 449, 150% torque

Short Time Fourier Transform

The time shift in the STFT varies along the x-axis over one rotation of the pinion. The y-axis represents each frequency with the lower frequencies at the top of the axis and the higher frequencies near the bottom. The block size represents the window size that was chosen. In this case there are 29 blocks that span the entire STFT. Unlike the CWT this is the smallest resolution because of the fixed window size. The smallest frequency is $\frac{1}{8}$ of the rotation or 2.375 times the gear mesh frequency, the gear mesh frequency will register at indices 3 and 4.

First, the discussion will be just about accelerometer 1. For the early data record, 25, the gear mesh signal at the 19 per rotation of data shows up in the analyses. Fig. 6a shows the amplitude of the STFT for data record number 25. Notice how the dark line at approximately frequency index 3, runs across almost the entire rotation indicating that the 19 per rotation is very dominant. Also, the second harmonic is apparent on this accelerometer. For the second data record, 250, the lower frequencies dominate as seen on Fig. 6b. The dark line has moved to a lower frequency and is no longer as solid, indicating that there may be problems with the gear. For the final data record, 449, the amplitude shown in the analyses is significantly larger.

Also, the dominant frequencies are lower than 19 per rotation and vary upon the location of the rotation. Fig. 6c shows the STFT for data record number 449. Now the line is once again more solid that in record number 250, but at a lower frequency, indicating that the gear no longer has 19 teeth.

Next compare the accelerometers. In the early data record, 25, all of the accelerometers have a strong component at 3 and 4, corresponding to the gear mesh frequency of 19 shaft harmonics. For the second data record, 250, the lower frequencies dominate on accelerometer 1. On accelerometer 2, the dominant frequency is still 19 shaft harmonics, and on the other accelerometers, 4 and 5, the dominant frequencies vary by location up to 19 per rotation. Finally, for data point, 449, all amplitudes have increased for the accelerometers. The dominant frequencies are all less than 19 per rotation and vary by location.

Choi-Williams Distribution

The time shift in the Choi-Williams Distribution varies along the x-axis over one rotation of the pinion. The resolution in the time shift is the same as the resolution in the data, 256 points per rotation. The y-axis represents each frequency with the lower frequencies at the top of the axis and the higher frequencies near the bottom. Since the Choi-Williams data is four times finer than the Fourier transform, the index for the frequency is at four times 19, or at 76.

The first discussion will be about accelerometer 1. For the early data record, 25, the gear mesh signal at the 19 per rotation shows up in the analyses. Also, the second and third harmonics are apparent. This can be seen in Fig. 7a, the WV-CW for data record 25 as a solid dark line at approximately 4 times frequency index equal to 76, or 19 per rotation. Vertical streaks spaced at 19 per rotation are visible throughout most of the "frequency" range for data record 25. For the middle data record, 250, a lower frequency dominates. If Fig. 7b is observed for data record 250, it is obvious that the dark line is now below 50. For the final point, 449, the amplitude is significantly larger. Notice, for example, on Fig. 7c for data record 449 that the amplitude range is now from -5 to over 15, a significant change from Fig. 7a and 7b. In addition, the dominant frequencies are lower than 19 per rotation.

Now all of the accelerometers will be discussed. In data record 25, all accelerometers showed the gear mesh signal at the 19 per rotation of the data. Accelerometers 1 and 5 also clearly show two harmonics. For the middle data record 250, the gear mesh frequency is picked up on accelerometers 2, 4 and 5. The transform from accelerometer number 1 picks it up very faintly. All of the accelerometers pick up the lower frequencies. At the final data record, accelerometers 2 and 5 have three prominent frequencies, including the original gear mesh signal. Transforms of both accelerometers 1 and 4 contain strong lines at 32 and 36, corresponding to 8 and 9 shaft harmonics.

Continuous Wavelet Transform

The time shift in the CWT varies along the x-axis over one rotation of the pinion. The resolution in the time shift is the same

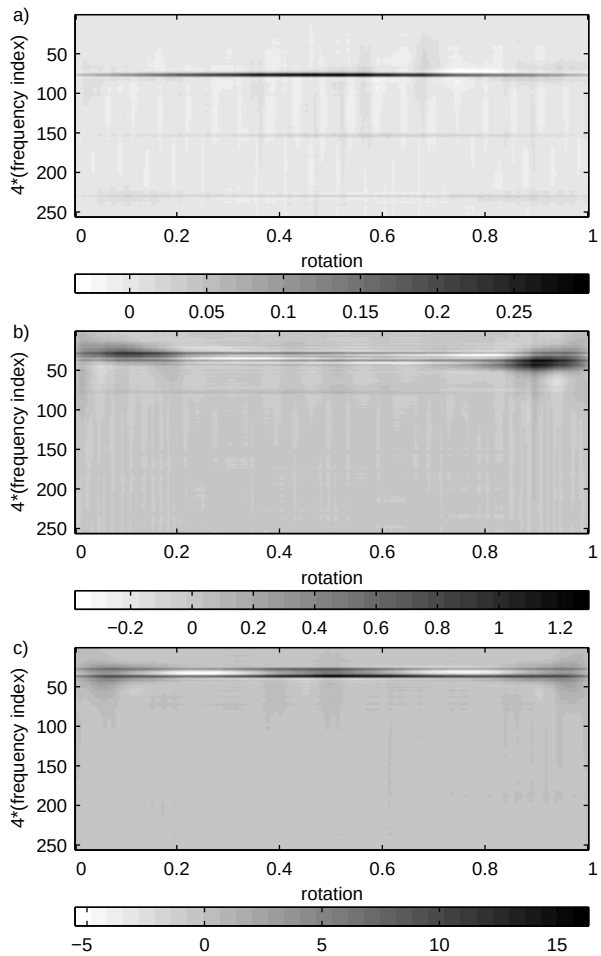

Figure 7: Choi-Williams Method for Accelerometer 1 a) data record 25, 80% torque, b) data record 250, 150% torque, c) data record 449, 150% torque

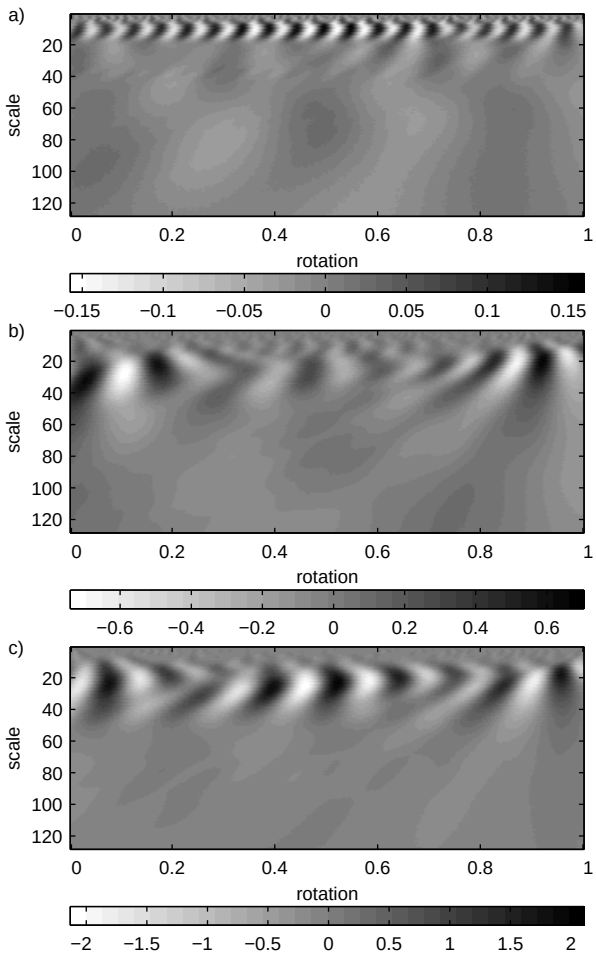

Figure 8: Continuous Wavelet Transform for Accelerometer 1 a) data record 25, 80% torque, b) data record 250, 150% torque, c) data record 449, 150% torque

as the resolution in the data, 256 points per rotation. The y-axis represents each scale of the wavelet. The scale 11 maximizes the gear mesh frequency. The value at each point of the image represents the transformed data at that scale. Notice, for example, how in Fig. 8a the amplitude goes up and down at 19 per rotation in the region around the scale that maximizes the gear mesh frequency. Also notice how at the edges of the image the amplitude is less than it is at the middle of the image. This corresponds to the time data in which the amplitude of the data also increases in the middle of the rotation.

The first accelerometer 1 will be discussed. For the early data record, 25, the gear mesh signal at 19 per rotation shows up in the analyses. Fig. 8a shows this quite clearly. Count the number of wave forms. For the middle data record, 250, the gear mesh at 19 per rotation is not readily apparent. But, structures showing lower frequency are there. These can be seen in Fig. 8b. They appear as larger waves at higher scales. For the final data record, 449, the dominant structures display frequencies lower than 19 per rotation. This indicates that there are no longer 19 working teeth, in fact, far fewer. In addition, the amplitude is significantly larger as can be seen in the scale bar in Fig. 8c for data record 449.

Now all of the accelerometers will be discussed. In data record 25, all accelerometers have a clear 19 per gear rotation waveform across the whole rotation except for accelerometer 5 which has a lowered amplitude for the last quarter rotation. In data record 250, accelerometer 1 has a lower frequency structure present while accelerometers 2, 4, and 5 have 19 waveforms still present, but with lower frequency waveforms occurring also. In data record 449, accelerometer 1 has a fairly clear 9 waveform while the other 3 accelerometers are mixed.

Discrete Wavelet Transform

The DWT is distinct from the other transforms in this paper. It does not expand the dimensionality of the representation of the data. In this case, where seven orders have been used, there are seven vectors returned which are referred to as details and one vector which is referred to as an average. In Fig. 9, the details with the highest frequencies are at the top of the plot. The average vector, which contains the lowest frequency, is at the bottom of the plot. Also in Fig. 9, the time shift in the DWT varies along the x-axis over one rotation of the pinion. The time resolution is finer in the details at lower scale and higher frequency near the top of the plots. The gear mesh frequency at 19 shaft orders does not match any of the scales. Signals at the gear mesh frequency are split mainly between scales 3 and 4.

The first part of the discussion will just include accelerometer 1. For the early data record, 25, the gear mesh signal at 19 per rotation shows up in the analyses as large details at scales 3 and 4. This can be seen in Fig. 9a for data record 25. At data record 250, details 4 and 5 are the largest. Refer to Fig. 9b for data record 250 for details. For data record 449, the amplitudes on all of the details and average are significantly larger. But, for details, 4 and 5, they are significantly larger. Fig. 9c shows this for data record 449.

Now all of the accelerometers will be included. For the early data record, 25, the gear mesh signal at 19 per rotation shows up in all of the analysis for all of the accelerometers. At data record 250, details 4 and 5 are large on all of the accelerometers. Finally, for data record 449, the amplitudes are significantly larger for all of the accelerometers.

SVD Models of Time-Frequency Analyses

SVD models were made for all four transforms of all four accelerometers. The training set consisted of half of the non-anomalous data for the two lowest torque levels for all 16 models. Results for accelerometer number 1, close to the pinion and approximately radial to the input shaft, follow: Figure 10 displays the normalized rms residual for all four time-frequency analyses. For the first torque level of 80% in data records 1-43, the residual level is at the low baseline level for all time frequency methods except the CWT. For data records 27-32 the residual increased almost 100% over baseline for the CWT. For the second torque level of 100% in data records 44-126, the residual levels are in two ranges, the low baseline and a higher level. For data records 44-53, the residuals for the STFT, WV-CW and DWT show a slight increase above the baseline, the residual for the CWT shows 100% increase. For the third torque

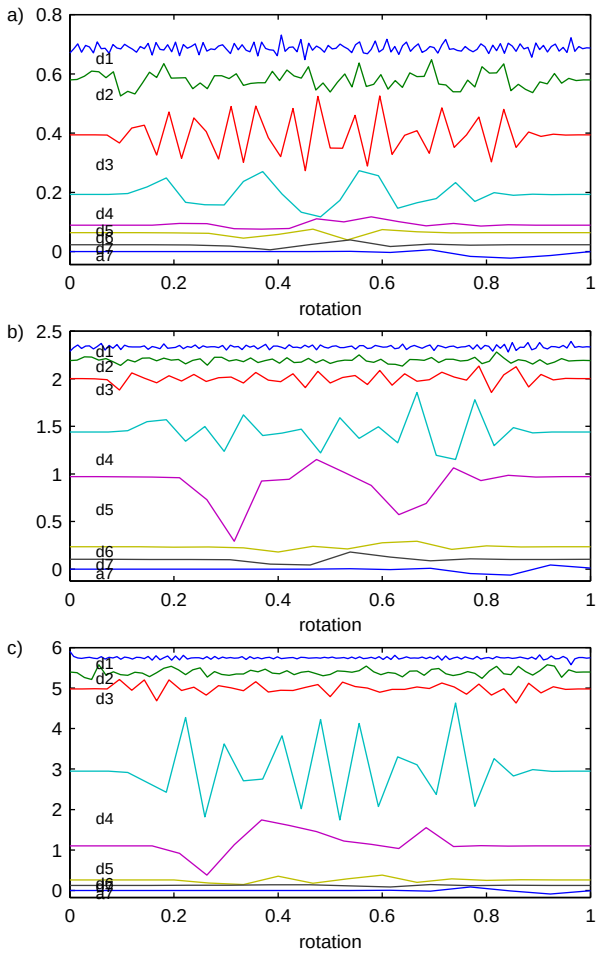

Figure 9: Discrete Wavelet Transform for Accelerometer 1 a) data record 25, 80% torque, b) data record 250, 150% torque, c) data record 449, 150% torque

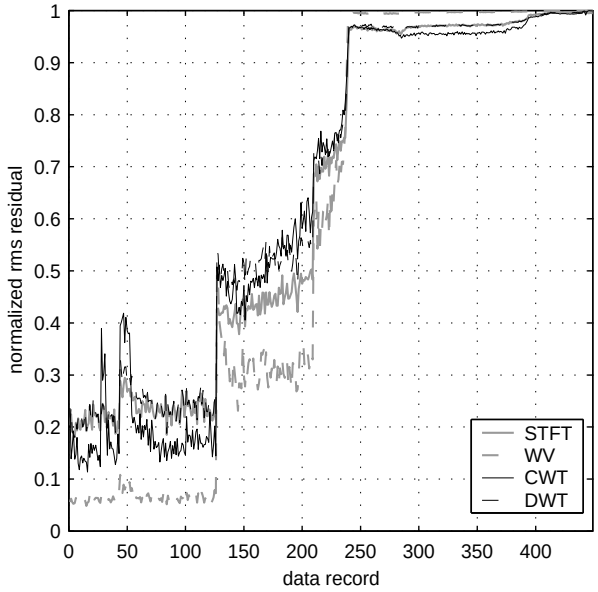

Figure 10: Residual RMS for all Four TF Methods Using the Metric on Accelerometer 1.

level of 125% in data records 127-209 the residual jumps up to a larger level at the torque increase. During the third torque level the residual for the STFT gradually increases, the residual for WV-CW initially jumps up to about 0.5 then decreases to about 0.3 for most of the records, the residual for the CWT initially jumps up then decreases then steadily increases and the residual for the DWT remains about constant. For the fourth torque level of 150% in data records 210-449 the residuals for all transforms initially jump up at data record 210 then rise steadily to record 238, then jump up again and remain at a high level for the remainder of the data records. In this last region there are small changes in the residual occurring around records 300, 410 and 440. All residuals above data record 240 are at high levels, close to one. These high levels indicate that the shape of the vibration signals differ greatly from the shape of those in the training set. For example, the time histories in Figs. 12e and 12f do not resemble those in Figs. 12a and 12b.

The residuals for the other three accelerometers show similar behavior overall with some variations. The variations will be described for the CWT (Fig. 11). The residual for accelerometer number 2, which is also close to the pinion and oriented approximately tangential to the shaft of the pinion, shows jumps near records 400, 410 and 440. Below record number 150, where there is the least damage to the pinion, the residual for accelerometer number 5, with the same orientation as accelerometer number 1 and farthest from the pinion, most closely matches the residual for accelerometer number 1. Above record number 240, the residual for accelerometer number 4, radial to the planetary gear and offset about 45 deg from the input pinion, most closely resembles the residual from accelerometer number 1.

For comparison, the rms levels for the four accelerometers are shown in Fig.12. For accelerometer number 1, the rms is low up through data record number 238 with small decreases at the first and second torque changes. There is a jump up around

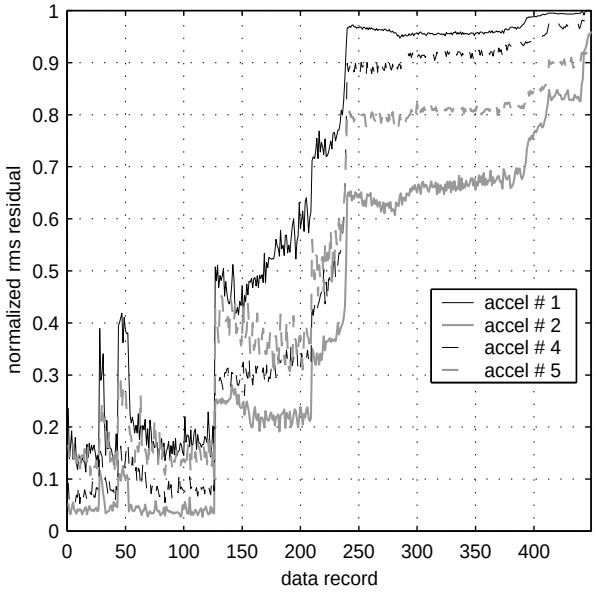

Figure 11: CWT RMS Residual for CWT Using the Metric on all Four Accelerometers.

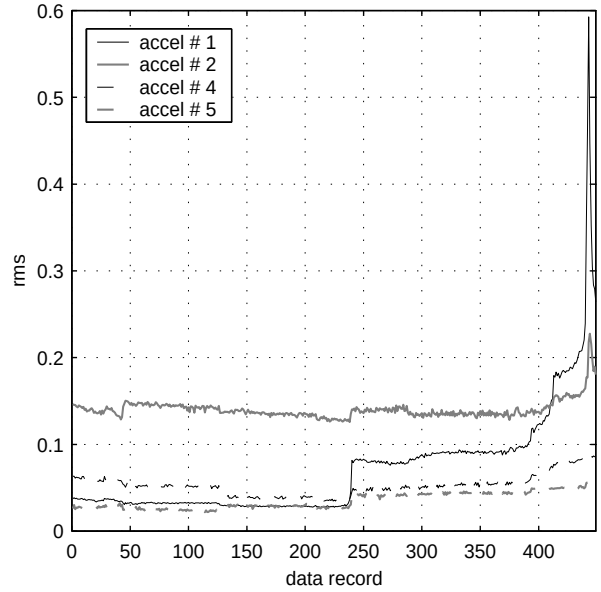

Figure 12: RMS of Time Data on All Four Accelerometers

238, a gradual rise between records 288 and 320, jumps near 398 and 412, a very steep rise from 440 to 443 followed by a drop off over the final records.

The Choi-Williams method gave the best fit for all of the accelerometers from data records $1:127$, the points that were used in the training. For accelerometers 1 and 3, the Discrete Wavelet Transform and the Short Time Fourier Transform both fit poorly for the same data records.

The coefficients generated by the model and not shown here relate to both the torque level of the gear and the health of the gear. In all of the models the coefficients change when the torque changes. The coefficients also change for most of the 16 models at data records 27, 33, 44, 54, 240, 280, 393, 408, 440 and 443. Coefficients from the models made with the DWT show the least amount of change at those records.

Discussion

Computer resources

All of the computations were done on a generic 550 Mhz Pentium III Xeon with 1 Gb of memory running Redhat Linux 6.2.

Referring to Table 2, in increasing order, a single STFT took just .005 seconds to complete. A DWT took .096 seconds and a CWT took 1.37 seconds. The WV-CW which took 2.48 seconds to perform was by far the longest single method.

Also, in increasing order, the SVD times for a single DWT took 4.7 seconds. A STFT took 6.0 seconds and a CWT took 77.9 seconds. Again, the method which took the longest was the WV-CW, taking 176.4 seconds.

Although there were no exact figures kept on the amount of space used for each of the methods, from estimates, the DWT took the least amount of space. The STFT took the second most, and the CWT took the next. The WV-CW, with a very large margin, took the most memory.

The WV-CW method took the most computer resources without showing a clear benefit over other methods.

SVD Modeling and Metric

All four time-frequency analyses clearly show changes in the data over the course of the experiment. The experiment did not provide a way to directly determine the state of the gear during the experiment. The gear contained notches on nine of the teeth before data record one and the gear sustained considerable damage by the end of the experiment. The damage probably occurred in stages. Changes in the vibration signal are indicated by the modeling of the time-frequency transforms at data records 27-32, 44-53, 127-209, 210-240, 240, 285, 400, 410, 440 and 443-449.

The local increases in the residual at data records 27-32 and 44-53 are very evident on the models for the CWT models, but not clear for the other the time-frequency transforms. The model coefficients in these local regions differ for the models of all of the time-frequency analyses and all accelerometers. These findings coincide with data records where the test rig may have had problems with the input shaft. These relatively high levels of the residual are probably due to mechanical problems with the test rig.

The very steep jump in the residual at the second torque increase at data record 127 may indicate some gear damage occurred at this time. The residual change indicates a change in the character of the signal, not the amplitude of the signal. Although the character of the vibration signal of a gear changes with torque, examining measurements from gears known to contain no damage leads to an expectation of a smaller change in the residual for the torque change in this experiment. It is plausible that if a crack in the gear was on the verge of propagating and causing functional damage that an increase in torque would initiate the damage. The time history in Fig. 12c for data

Table 2: Method Comparisons

	Time (sec.)	SVD times (sec.)	Samples per Rev.	Spectrum in shaft harmonics
STFT	.005	6.0	256	128
Choi-Williams	2.48	176.4	256	64
Continuous Wavelet	1.37	77.9	256	128
Discrete Wavelet	.096	4.7	256	128

record 150 shows some of the waveforms with a tooth-width wavelength beginning to merge near 0.1, 0.75 and 0.85 of the rotation when compared with the time histories from the earlier records which more clearly contain 19 individual waveforms for the 19 teeth on the pinion. With minor damage to a gear tooth, the tooth may no longer carry its share of the load thus producing changes in the vibration signal, including the merging of waveforms. The large jump at the second torque change may occur because the torque is increasing from 100% to 125% of full load. The large jump in the residual at record 127 is probably a combination of change to a torque level not included in the training set and change in the shape of the time history due to minor damage to some of the gear teeth. A training set containing data with a good gear running at 125% torque is needed to make more definite statements. The gradual increase in the residual for many of the models in the third torque region may indicate gradual increasing damage.

The steep jump in the residual at the final torque increase (data record 210) and gradual increase for about the next 40 data records can be interpreted in the same manner as the jump and increase after the second torque change. The time history in Fig 12d for data record 230 shows even more merging of waveforms on the scale of the gear-tooth to a longer wavelength.

All changes after the final torque definitively indicate changes in the state of the gear. The steep jump in residual and abrupt changes in the coefficients of the models occurred near data record 240 indicating an abrupt change in the gear. Changes indicating discrete changes in the gear condition also occurred near data records 285, 400 and 410.

For the last 9 data records, the coefficients of the models show more drastic change than the residual. Many of the coefficients increase by a factor of 2 or more over a few data records then decrease to low levels while the residual increases or remains high. Very severe damage probably occurs over the last several data records.

Residuals from models of the different time-frequency analyses have some different characteristics. Residuals from the model of the CWT indicated problems with the test rig more than the other time-frequency transforms. The WV-CW model gave the closest fit to the data in the training set and its residual rose more than for other models. All of the residuals on the different accelerometers showed indications of damage to the pinion gear.

When the level of the residual from modeling indicates damage may exist in machinery, more information can be obtained in the modeling by further examination of the residual. If the damage is of a local nature, the time-frequency representation of the residual will show structures localizing the damage. Such structures were observed in the measurements of current test waveforms changing size from 1/19 of a rotation to about 2/19 of a rotation. Examination of such residuals from different kinds of faults in different machinery might lead to the ability to classify the damage.

The time-frequency analyses distinguish events at distinct times, so signals used in the modeling must be consistently synchronized. When the interrupter signal used for synchronization is on the same shaft as the gear being analyzed, there is no alignment problem. When the phase alignment queue is from a shaft spinning at a different frequency something must be done to account for the proper phase. One option is to align the data with a circular correlation correction. Another option is to use synchronously averaged data from all possible phase alignments of the gear in the training set.

The modeling in the current work does not account for the nonlinear dependence of the vibration signal on torque. The modeling should be developed to incorporate the signal dependence on torque and possibly other operating parameters. A nonlinear extension of Principal Component Analysis [38] might accomplish this.

The SVD modeling of the transformed data can be developed into an automatic fault detector for use on machinery. Extensive training and test sets covering all operating conditions of the gear are needed. The distribution of the normalized rms residual can then be determined by the histogram for the training and test sets. With the distribution the trigger level on the normalized rms residual can be set to yield a given false alarm rate. The sensitivity of the fault detection can be investigated with test rig data of gears with and without faults. The false alarm rate can be investigated with measurements made in flight. This fault detection will have the combined benefits of sensitivity to faults that produce changes in the time-frequency analyses and the ability to automate fault detection with a single number metric.

Conclusion

The Short Time Fourier Transform, the Choi-Williams modified version of the Wigner-Ville Transform, the Continuous Wavelet Transform and the Discrete Wavelet Transform all reveal changes in the vibration measurements of a damaged spiral bevel pinion gear.

The Choi-Williams modified version of the Wigner-Ville Transform and the Continuous Wavelet Transform both use considerably more computational resources than the Short Time Fourier Transform and the Discrete Wavelet Transform.

A single number metric applicable to automatic fault detection was introduced that can be produced from any systematic numerical representation of the vibration signals. Vibration measurements of the gear operating without faults are needed to generate the model using the metric.

The new metric revealed indications of test rig problems with the Continuous Wavelet Transform, thus indicating the additional computational resources needed for the Continuous Wavelet Transform add value.

More testing of the new metric is needed to determine its value for automatic fault detection and to develop methods of setting the threshold for the metric.

Acknowledgments

The authors wish to acknowledge Drs. Ed Huff and Irem Tumer for their helpful discussions about this paper.

References

[1] Staszewski, W.J., Worden, K., Tomlinson, G.R., "Time-Frequency Analysis in Gearbox Fault Detection Using the Wigner-Ville Distribution and Pattern Recognition," Mechanical Systems and Signal Processing, Vol 11, (5), 1997.

[2] Zakrajsek, James J., Handschuh, Robert F., Lewicki, David G., Decker, Harry J., "Detecting Gear Tooth Fracture in a High Contact Ratio Face Gear Mesh," NASA Technical Memorandum 106822, 1995.

[3] Hambaba, Ahmed, "Multiresolution Error Detection On Early Fatigue Cracks in Gears," 2000 IEEE Aerospace Conference, March 2000.

[4] Samuel, Paul D., Pines, Darryll J., Lewicki, David G., "A Comparison of Stationary and Non-Stationary Metrics for Detecting Faults in Helicopter Gearboxes," Journal of the American Helicopter Society, Vol 45 (2), April 2000.

[5] Wang, W.J., McFadden, P.D., "Early Detection of Gear Failure by Vibration Analysis–I. Calculation of the Time-Frequency Distribution," Mechanical Systems and Signal Processing, Vol 7, (3), 1993.

[6] Williams, W.J., Zalubas, E.J., "Helicopter Transmission Fault Detection Via Time-Frequency, Scale and Spectral Methods," Mechanical Systems and Signal Processing, Vol 14, (4), 2000.

[7] Wang, W.J., McFadden, P.D., "Early Detection of Gear Failure by Vibration Analysis–II. Interpretation of the Time-Frequency Distribution Using Image Processing Techniques," Mechanical Systems and Signal Processing, Vol 7, (3), 1993.

[8] Loughlin, P., and Cakrak, F., "Conditional Moments Analysis of Transients with Applications to Helicopter Fault Data," Mechanical Systems and Signal Processing, Vol 14, (4), December 1999.

[9] Oehlmann, H., Brie, D., Tomczak, M., and Richard, A., " A Method for Analysing Gearbox Faults Using Time-Frequency Representations," Mechanical Systems and Signal Processing, Vol 11, (4), October 1996.

[10] Baydar, N., and Ball, A., "Detection of Gear Deterioration Under Varying Load Conditions by Using the Instantaneous Power Spectrum," Mechanical Systems and Signal Processing, Vol 14, (6), November 1999.

[11] Boulahbal, D., Golnaraghi, M. Farid, and Ismail, F., " Amplitude and Phase Wavelet Maps for the Detection of Cracks in Geared Systems," Mechanical and Signal Processing, Vol 13, (3), October 1998.

[12] Lin, Jing, and Qu, Liangsheng, "Feature Extraction Based on Morlet Wavelet and its Application for Mechanical Fault Diagnosis," Journal of Sound and Vibration, Vol 234, (1), January 2000.

[13] Lopez, Jose E., Tenney, Robert R., and Deckert, James C., " Fault Detection and Identification Using Real-Time Wavelet Feature Extraction," Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, 1994.

[14] Ferlez, Robert, and Lang, Derek C., " Gear-Tooth Fault Detection and Tracking Using the Wavelet Transform," Proceedings of Prognosis of Residual Life of Machinery and Stuctures, MFPT, 1998.

[15] Dalpiaz, G., Rivola, A, and Rubini, R., "Effectiveness and Sensitivity of Vibration Processing Techniques for Local Fault Detection in Gears,", Mechanical Systems and Signal Processing, Vol 14, (3), 1999.

[16] Lin, S.T. and McFadden, P.D., "Gear Vibration Analysis by B-Spline Wavelet-Based Linear Wavelet Transform," Mechanical Systems and Signal Processing, Vol 11, (4), April 1997.

[17] Wang, W.J., McFadden, P.D., "Application of Wavelets to Gearbox Vibration Signals for Fault Detection," Journal of Sound and Vibration, Vol 192, (5), 1996.

[18] Lin, J. and Qu, L, "Feature Extraction Based on Morlet Wavelet and its Application for Mechanical Fault Diagnosis," Journal of Sound and Vibration, Vol 234, (1), 2000.

[19] Paya, B.A., and Esat, I.I., "Artificial Neural Network Based Fault Diagnostics of Rotating Machinery Using Wavelet Transforms as a Preprocessor," Mechanical Systems and Signal Processing, Vol 11, (5), March 1997.

[20] Wang, W.J., and McFadden, P.D., "Application of Orthogonal Wavelets to Early Gear Damage Detection," Mechanical Systems and Signal Processing, Vol 9, (5), May 1995.

[21] McFadden, P.D., “Detection of Gear Faults by Decomposition of Matched Differences of Vibration Signals”, Mechanical Systems and Signal Processing, Vol 14, (5), May 2000.

[22] Lewicki, David G., Coy, John J., “Vibration Characteristics of OH-58A Helicopter Main Rotor Transmission,” NASA Technical Paper 2705, 1987.

[23] Kaiser, Gerald, A Friendly Guide to Wavelets, Birkhauser, Boston, 1994.

[24] Choi, H. and Williams, W.J., “Improved Time-Frequency Representation of Multicomponent Signals Using Exponential Kernels,” IEEE Trans. on Acoustics Speech and Signal Processing, Vol 37, (6), June 1989.

[25] Cohen, Leon, “Time-Frequency Distributions-A Review,” Proceedings of the IEEE, Vol 77, (7), July 1989.

[26] Daubechies, Ingrid, Ten Lectures on Wavelets, SIAM, Philadelphia, Pennsylvania 1992.

[27] Stapleton, John C. and Bass, Steven Craig, “Synthesis of Musical Tones Based on the Karhuen-Loève Transform,” IEEE Transaction on Acoustics, Speech and Signal Processing, Vol 36, (3), March 1988.

[28] Kirby, M. and Sirovich, L., “Application of the Karhuen-Loève Procedure for the Characterization of Human Faces,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 12, (1), January 1990.

[29] Beyerbach, Daniel and Nawab, Hamid, “Principal Components Analysis of the Short-Time Fourier Transform,” Proceedings International Conference on Acoustics, Speech and Signal Processing, Vol 3, April 1991.

[30] Newab, Hamid, Beyerbach, Daniel and Dorken, Erkan, “Principal Decomposition of Time-Frequency Distributions,” IEEE Transactions on Signal Processing, Vol 4,(11), November 1993.

[31] Hung, Elmer and Zhao, F., “Diagnostic Information Processing for Sensor-Rich Distributed Systems,” Poc. 2nd International Conference on Information Fusion, Sunnyvale, CA, 1999.

[32] Tumer, Irem Y., Wood, Kristin L., and Busch-Vishniac, Ilene J., “A Mathematical Transform to Analyze Part Surface Quality in Manufacturing,” Journal of Manufacturing Science and Engineering, Vol 122 (1), February, 2000.

[33] Penev, Penio S. and Sirovich, Lawrence, “The Global Dimensionality of Face Space,” Proceedings of Forth IEEE International Conference on Automatic Face and Gesture Recognition, March 2000.

[34] Jolliffe, I. T., Principal Component Analysis, Springer-Verlag, New York, 1986.

[35] Wax, Mati and Kailath, Thomas, “Detection of Signals by Information Theoretic Criteria,” IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol 33, (2), April, 1985.

[36] Williams, Douglas, B., “Counting the Degrees of Freedom When Using AIC and MDL to Detect Signals,” IEEE Transactions on Signal Processing, Vol 42, (11), November, 1994.

[37] Johnson, Don H. and Dudgeon, Dan E., Array Signal Processing Concepts and Techniques, Prentice Hall, New Jersey, 1993.

[38] Kambhatla, Nanda, and Leen, Todd K., “ Fast Non-Linear Dimension Reduction” in Advances in Neural Information Processing Systems 6, Morgan Kaufmann Publishers, San Francisco, California 1994.