SIGNAL PROCESSING FOR EFFECTIVE VIBRATION ANALYSIS

Dennis H. Shreve

IRD Mechanalysis, Inc
Columbus, Ohio
November 1995

ABSTRACT

Effective vibration analysis first begins with acquiring an accurate time-varying signal from an industry standard vibration transducer, such as an accelerometer. The raw analog signal is typically brought into a portable, digital instrument that processes it for a variety of user functions. Depending on user requirements for analysis and the native units of the raw signal, it can either be processed directly or routed to mathematical integrators for conversion to other units of vibration measurement. Depending on the frequency of interest, the signal may be conditioned through a series of high-pass and low-pass filters. Depending on the desired result, the signal may be sampled multiple times and averaged. If time waveform analysis is desired in the digital instrument, it is necessary to decide the number of samples and the sample rate. The time period to be viewed is the sample period times the number of samples. Most portable instruments also incorporate FFT (Fast Fourier Transform) processing as the method for taking the overall time-varying input sample and splitting it into its individual frequency components. In older analog instruments, this analysis function was accomplished by swept filters.

There are a large number of setup parameters to consider in defining the FFT process: (1) lines of resolution, (2) maximum frequency, (3) averaging type, (4) number of averages, and (5) window type. All of these interact to affect the desired output, and there is a distinct compromise to be made between the quality of the information and the time it takes to perform the data collection.

Success in predictive maintenance depends on several key elements in the data acquisition and conversion process: (1) the trend of the overall vibration level, (2) the amplitudes and frequencies of the individual components of the composite vibration signal, and (3) the phase of a vibration signal on one part of a machine relative to another measurement on the machine at the same operating condition.

This paper is intended to take the reader from the vibration sensor output through the various stages in the signal processing path in a typical vibration measurement instrument using modern digital technology. Furthermore, it considers the various data collection setup parameters and tradeoffs in acquiring fast, meaningful vibration data to perform accurate analysis in the field of predictive maintenance.

As they are related to successful vibration analysis, analog signal sampling and conditioning; anti-aliasing measures; noise filtering techniques; frequency banding - low-pass, high-pass, and band-pass; data averaging methods; and FFT frequency conversion are among the topics of detailed discussion.

1. DISCUSSION

Vibration analysis starts with a time-varying, real-world signal from a transducer or sensor. From the input of this signal to a vibration measurement instrument, a variety of options are possible to analyze the signal. It is the intent of this paper to focus on the internal signal processing path, and how it relates to the ultimate root-cause analysis of the original vibration problem. First, let us take a look at the block diagram for a typical signal path in an instrument, as shown in Figure 1.

2. TIME WAVEFORM

A typical time waveform signal in analog form from an accelerometer could take an appearance like that shown in Figure 2.

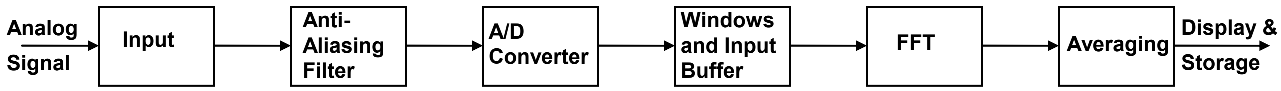


Figure 1. Typical Signal Path

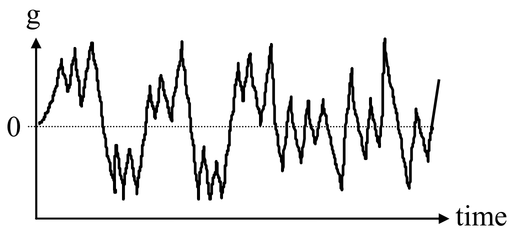


Figure 2. Typical Time Waveform

In a digital instrument, much the same thing is seen. However, it is necessary in a digital instrument to specify several parameters in order to accurately reconstruct the plot. It is important to tell the instrument what sample rate to use, and how many samples to take. In doing this, the following are specified:

a) The time period that can be viewed. This is equal to the sample period times the number of samples. The highest frequency that can be chosen for sampling is an attribute of the instrument and is expressed in Hertz or CPM (where 1 Hz = 60 CPM). Sample rates of up to 150 KHz are not uncommon in modern instruments.

b) The highest frequency that can be seen. This is always less than half the sample frequency.

The number of samples chosen is typically a number like 1024 (this is 2^{10}, a good reference for later computation of FFTs). The resulting time waveform requires a discerning eye to evaluate, but is very popular as an analysis tool in industrial processes. It is important to note that brief transients are often visible in this data, where they could be covered up by further signal processing.

In processing a digital signal for analysis, there are a number of limitations to take into account:

- Low pass filters - to eliminate any high frequencies.

- High pass filters - to eliminate DC and low frequency noise.

- Transducer characteristics - a factor that usually limits effective lowest and highest frequencies, and also has an inherent resonance frequency that magnifies signals at that point.

Additionally, the integration of signals -- producing a velocity or displacement signal from an accelerometer or a displacement signal from a velocity pickup -- will tend to lose low frequency information and introduce noise. Integration of the input signal is generally best accomplished in analog circuits due to the limited dynamic range of the analog-to-digital (A/D) conversion process. Digital circuits typically introduce more errors and if there is any jitter at low frequency, it becomes magnified upon integration.

These are the raw ingredients for digital signal and analysis. Within the limitations discussed and further processing, it becomes quite possible to perform extremely accurate diagnoses of equipment condition.

3. FFT

The most common form of further signal processing is known as the FFT, or Fast Fourier Transform. This is a method of taking a real-world, time-varying signal and splitting it into components, each with an amplitude, a phase, and a frequency. By associating the frequencies with machine characteristics, and looking at the amplitudes, it is possible to pinpoint troubles very accurately. With analog instruments, the same information is provided with a swept filter. This is referred to as constant Q (or constant % bandwidth) filtering, where a low/high pass filter combination of say 2.5 % bandwidth is swept in real time through a signal to produce a plot of amplitude vs. frequency. This gives good frequency resolution at lower frequencies (e.g. 2.5 % of 600 CPM is 15 CPM resolution), and at high frequencies resolution is lower (2.5 % of 120,000 CPM is 3000 CPM). For this reason, the frequency axis is usually a log scale, as shown in Figure 3.

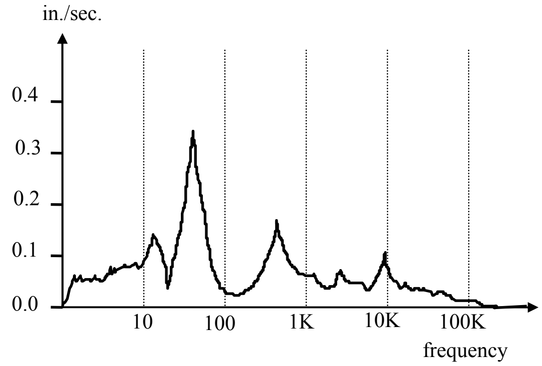


Figure 3. Velocity vs. Log Frequency

This “tuning” technique is much slower than an FFT, especially at low frequencies. It can miss information also because it only looks at each frequency at one instant in time. Swept filters are nevertheless a powerful analysis tool, especially for steady state vibrations.

In modern instruments today, the FFT is more commonly used to provide frequency domain information.

As the theory of Jean Baptiste Fourier states: All waveforms, no matter how complex, can be expressed as the sum of sine waves of varying amplitudes, phase, and frequencies. In the case of machinery vibration, this is most certainly true. A machine's time waveform is predominantly the sum of many sine waves of differing amplitudes and frequencies. The challenge is to break down the complex time-waveform into the components from which it is made. Figure 4 shows an example of this.

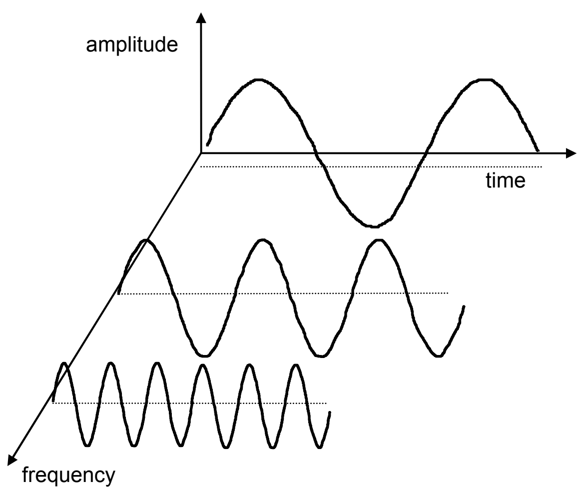


Figure 4. Complex Time Waveform Components

Three waveforms are shown, plotted in a 3-D grid of time, frequency and amplitude. If we add the waves together, we see our composite time waveform (Figure 5); and if we look end on to eliminate the time axis, we get a picture of the frequencies and amplitudes (Figure 6). This is our FFT.

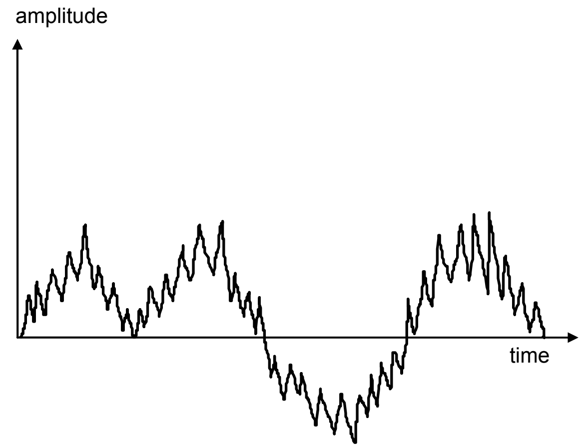


Figure 5. Composite Time Waveform

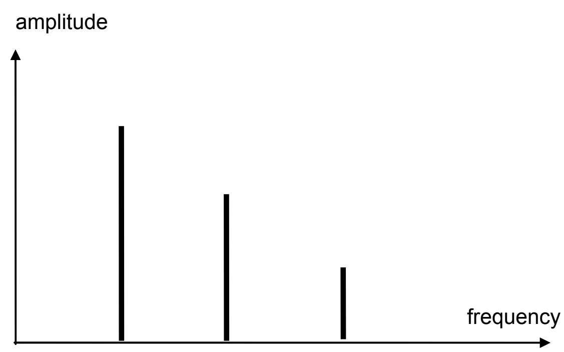


Figure 6. Frequency Components and Amplitudes

When an FFT measurement is specified in an instrument, there are several selections that can be made, as shown in Figure 7.

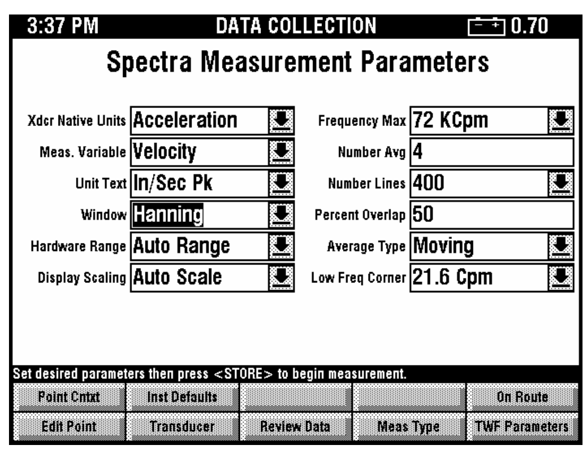


Figure 7. FFT Setup Parameters

Key parameters are as follows:

- Fmax
- Number of Averages
- Number of Lines
- Average Type
- Percent Overlap
- Low Frequency Corner
- Window Type

and each will be discussed in further detail.

4. LINES OF RESOLUTION

FFT resolution describes the number of lines of information that appear on the FFT plot, as shown in Figure 8. Typical values are 100, 200, 400, 800, 1600, 3200, 6400, and 12,800. Each line will cover a range of frequencies, and the resolution of each line can be calculated simply by dividing the overall frequency (Fmax) by the number of lines. For example, an Fmax of 120,000 CPM and 400 lines gives a resolution of 300 CPM per line.

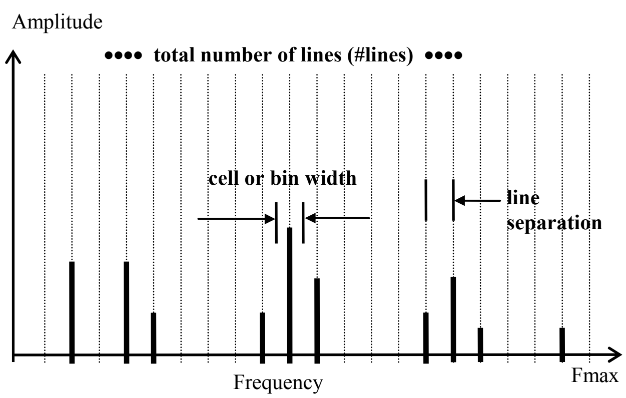


Figure 8. FFT Resolution

5. FMAX

This is the highest frequency that will be captured and displayed by the instrument. In choosing the Fmax, we also set other parameters. One of these is called the anti-aliasing filter.

As the operations used to produce FFTs are digital, and we use a digitized time waveform to produce the FFT, we are really looking at a series of points on the time waveform graph, as shown in Figure 9.

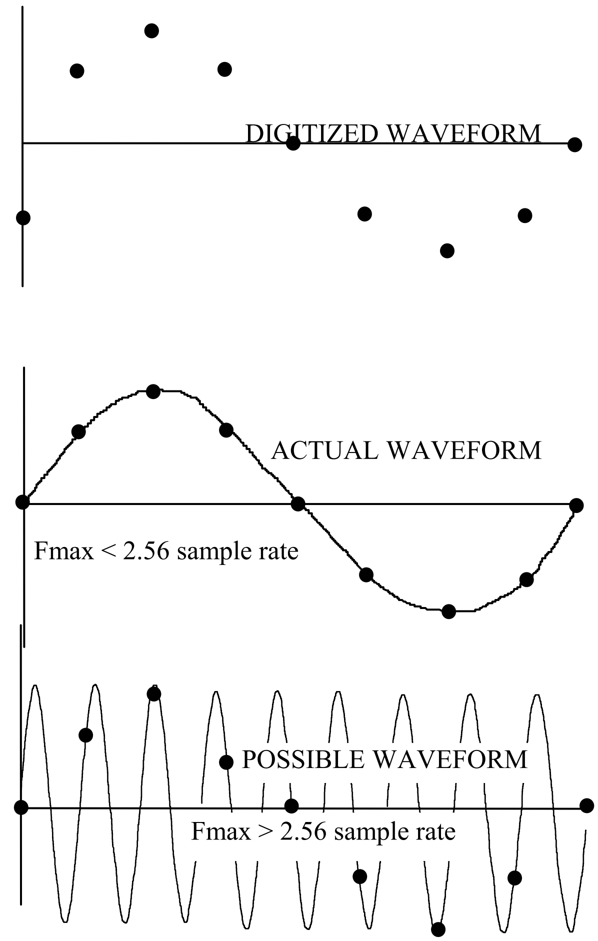


Figure 9. Digital Sampling and Aliasing

6. ALIASING

In order to ensure that sine waves can be generated from the points, we need to sample at a rate which is much higher than the highest frequency that we want to resolve. From a theorem of Claude Shannon and Harry Nyquist, the lowest sample rate we can use is at least double Fmax. This means that it is necessary to sample a pure sine wave at least twice its fundamental frequency in order to adequately define it. Due to the roll-off of the anti-aliasing filter, it is necessary to exceed a doubling of the highest frequency content. A number like 2.5 times would be adequate, but in order to comply with the computer world, 2.56 is usually the number employed. If a lower sampling rate is used, the original time-varying signal cannot be reconstructed and "aliasing" may occur. With this phenomenon, a high frequency component will tend to look like a lower frequency, as shown in Figure 9.

Figure 10 provides an example of filter roll-off and "fold-over" frequency phenomena in aliasing.

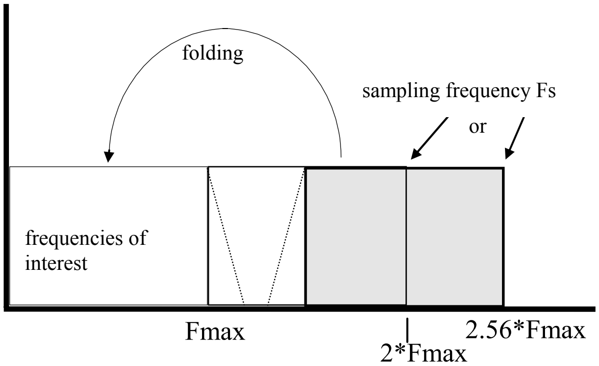


Figure 10. Aliasing Fold-over Phenomena

Two signals are said to alias if the difference of their frequencies falls within the frequency range of interest. This difference frequency is always generated in the process of sampling.

To ensure that we do not have any high frequency components in our signal (higher than the chosen Fmax value), we use an anti-aliasing filter to suppress the raw signal above Fmax. This combination of techniques saves processing time and ensures that the information in the frequency range we have chosen is accurate.

7. DATA CAPTURE TIME

As the parameters Fmax and lines of resolution are selected, the total sample time for capturing valid FFT data is determined.

For a 400-line FFT, due to the calculations involved, we need to take 1024 points on the waveform. This number (N = 2.56*(#lines)) is derived from the following calculations:

Bandwidth (BW) = F_{max}/(#lines)

$T_{(obs)}$ = 1/BW = (#lines)/F_{max}

$T_{(obs)}$ = N*$T_{(sample)}$ = N*(1/(2.56* F_{max}))

N = 2.56*(#lines)

where

(#lines) = total number of lines of FFT resolution

F_{max} = highest analyzed frequency (Hz.)

N = number of samples collected

$T_{(sample)}$ = sample period (sec.)

$T_{(obs)}$ = observation time (sec.).
See Figure 11 for an example on sampling.

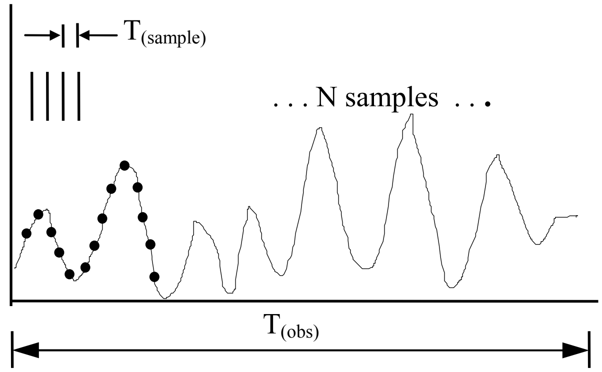


Figure 11. Sampling and Observation Time

If we assume we want an Fmax of 120,000 CPM and 400 lines of resolution, we can now determine how long our sampled time waveform must be.

- To avoid aliasing, a low pass filter of 120,000 CPM is selected
- To avoid aliasing, we sample at 307,200 CPM (=2.56 x 120,000).
- There are 1024 samples to yield 400 lines of resolution

The section of time waveform observed will be 1024 samples at a sample time of 2 msec., for a total of 0.2 sec. Thus, we need an instrument with at least 5 KHz sampling rate (1024 samples in 0.2 secs = 5120 samples/sec).

As another example, a 400 line FFT with an Fmax of 6000 CPM would require an observed time waveform calculated as follows:

$T_{(obs)}$ = N*$T_{(sample)}$ = N*(1/(2.56*F_{max}))

= 1024*(1/(2.56*100 Hz.))

= 1024*(1/256)

= 4 seconds.

While lower values of Fmax offer much improved resolution for the frequencies displayed, it does not come for free. Collection time for data is significantly longer. (The same holds true when low frequency corners are selected for the measurement.)

To illustrate the relationship between the length of the time waveform we need to observe and the resolution achieved, consider how you would need to examine a signal made up of two waveforms with very close frequencies. If the waveforms started off in phase, it would take a long time before they separated enough to show their different frequencies. For example, this can be heard as "beats" when two machines run at nearly the same speed. The bottom line is: *In order to achieve high resolution in the frequency domain, long sample times are required.*

8. NUMBER OF AVERAGES

When an FFT is produced, the instrument uses a digitized time waveform and performs the mathematical operation to produce the FFT. However, observing only one section of time waveform may exclude some peak caused by a random vibration influence. To minimize this, it is common to look at several sections of the time waveform, calculate several FFTs, and display an average result. Four averages are commonly taken.

Averaging is available in most FFT analyzers to assist in interpreting data. Averaging provides more repeatable results in data collection for early warnings of machine deterioration. Averaging also helps in the interpretation of complex, noisy signals.

Types of averaging include: linear, exponential, peak hold, and synchronous time averaging. Each type has certain qualities that allow it to better suited for a given application, and a brief description follows.

linear

In linear averaging, each instantaneous spectrum is added to the next and the sum is divided by the total number of spectra. This method is useful in obtaining repeatable data for fault trending, as used in most predictive maintenance programs. It is also useful for averaging out random background vibrations.

peak hold

Peak hold is not a true averaging method. During sampling time, the peak value registered in each analysis cell is captured and displayed. This method is very useful in viewing transients or for stress analysis calculations.

exponential

This technique takes the most recent spectrum collected and weighs it more heavily than the past data. It is useful in observing conditions that are changing slowly with respect to sampling time -- i.e., a steady-state process.

synchronous

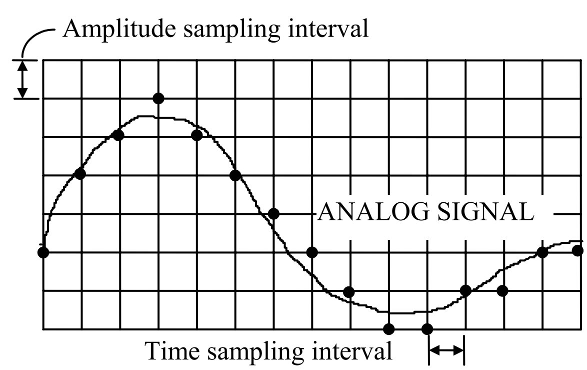
This method utilizes a synchronizing signal from the machine being analyzed. The synchronizing signal is usually derived from a photocell, electromagnetic pickup, or some other form of tachometer input.

The vibration input is sampled at precisely the same moment with respect to shaft rotation during the averaging time period. This method can prove to be a useful tool for filtering out random background vibrations.

9. A/D CONVERSION

In working with real-world analog signals that must be converted to digital format for computer processing, an A/D (analog-to-digital) converter is used. The sampling interval on a time basis is one important parameter, but most often, an A/D is specified by its amplitude resolution.

As computer processing circuits work in powers of 2, or binary numbers, A/D converters are characterized as 12-bit, 14-bit, 16-bit, etc. Thus, an A/D specified with 12-bit resolution offers 4096 intervals (or quantization levels) on an amplitude scale (i.e., $2^{12} = 4096$). The greater the resolution, the better the amplitude resolution, and hence the better dynamic range. An A/D with 16-bit resolution offers accuracy to one part in 65,536, or 96 dB dynamic range. The concept of amplitude resolution is shown in Figure 12.



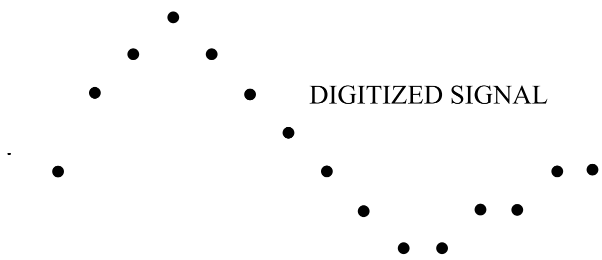


Figure 12. Signal Amplitude Sampling

A 12-bit A/D converter results in a resolution of 0.024% of the full-scale reading, while a 16-bit A/D is 16 times better, or 0.0015% of full scale. This extra esolution provides us with the ability to see both large and small amplitudes at the same time.

10. WINDOW TYPE

One more step that must be performed on the sampled signal is windowing, where the actual time-varying analog signal is "framed" by the multiplication of another known time function. The result of this mathematical operation is to provide a sampled time waveform that appears to be continuous and periodic. Discontinuities are "filled in" by forcing the sampled signal to be equal to zero at the beginning and end of the sampling period (window).

In using a window, however, there is a trade off between the ability to resolve frequencies and the resolution of amplitudes. If we have no window function applied (Rectangular Window), the frequency and amplitude resolution is excellent, provided the signal is periodic and fits the time sample exactly. For example, with a sine wave that starts at zero at the beginning of the sample, it would also need to finish at zero to give good resolution. If it does not, the waveform has the characteristics of a sine wave and a square wave -- that gives rise to "leakage" into the bins on either side of the main frequency on our FFT. Most windows, for this reason, ensure that the signal starts and finishes in our time sample at the same level, thus avoiding the need for a synchronous signal.

Leakage (or smearing) is the result of the FFT algorithm trying to handle discontinuities in the sample. The FFT sees the discontinuities as a modulating frequency. This produces spectral components (sidebands) where none truly exist. The use of windowing also affects our ability to resolve closely spaced frequencies while maintaining amplitude accuracy. One can only be optimized at the expense of the other.

There are many available windowing functions. Rectangular (actually no window), Flat-Top, Hamming, Kaiser-Bessel, and Hanning are among the list available. Perhaps the most commonly used window is Hanning (raised cosine). It is good for analyzing sine waves, as it provides a good compromise on both frequency and amplitude resolution. Its effect is shown in Figure 13.

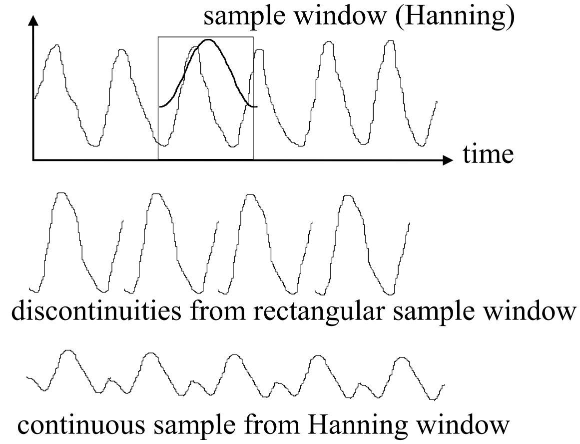


Figure 13. Hanning Window Sampling

Next, the FFT calculation takes the windowed values from the time waveform and calculates an amplitude for each line of resolution.

Each line of resolution is effectively the value of the overall reading for the vibration signal in the range covered by each FFT bin. There is one FFT bin per line, and for a 400 line spectrum of 120 K CPM Fmax, each bin would be 300 CPM wide. This gives the FFT a constant bandwidth (BW) of 300 CPM, and is the same as the resolution.

Each bin can be considered a low/high pass filter. The characteristics of the filter are determined by the window shape. In the case of the Hanning window, each bin has a filter characteristic as shown in Figure 14.

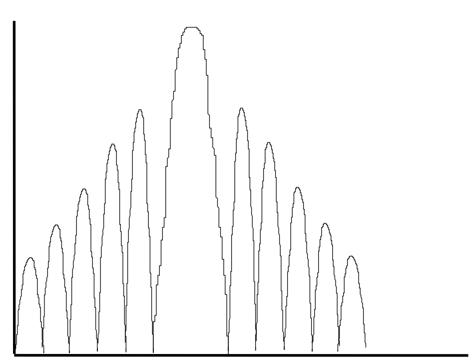


Figure 14. Hanning Filter Characteristics

The filter shape has sloping sides and does not have a flat top. Thus, some errors are introduced. The shapeof the top of the filter can give up to 16% error in amplitude (often called the Picket Fence Error) and due to the slope, a frequency in one bin will be seen in several other bins. This is leakage. See Figure 15.

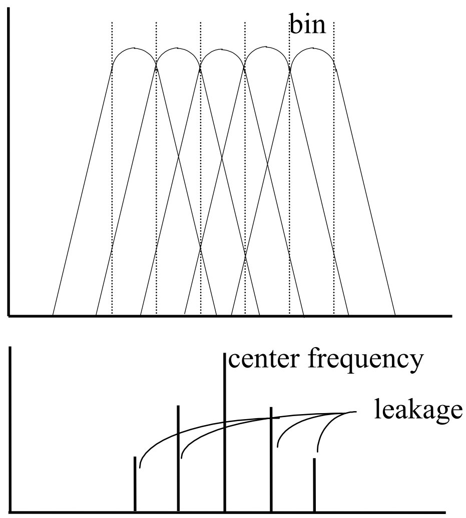


Figure 15. Hanning Leakage Effect

Hanning is a good compromise window as the main frequency is well defined and is usually at a maximum in only 1 bin, and amplitudes are comparatively accurate.

Other windows have uses in particular applications, and several will be discussed below.

flat top

The flat top window has a very wide filter which covers several bins. It shows a signal appearing at several frequencies, but has the advantage of giving very accurate amplitude. Its primary use is for calibration.

rectangular

This is actually no window at all. The advantage of using this comes in run-up or coast-down where if the windows are triggered by a signal in phase with rotation, where very good order tracking can be achieved. This window is also used for transients.

hamming

The Hamming window provides better frequency resolution at the expense of amplitude. Less of the signal leaks into adjacent bins than with the Hanning window. This can be used to separate close frequency components.

kaiser-bessel

This window is even better than the Hamming technique for separating close frequencies because the filter has even less leakage into side bins. The initial main envelope however covers several bins so resolution is less than with Hamming.

blackman-harris

Again, the Blackman-Harris window is a good tool for frequency separation, and it provides good amplitude accuracy.

11. SPEEDING UP THE PROCESS

There are two common ways to speed up FFTs. These are (1) overlap averaging, that works well with the Hanning Window and (2) folded FFT, that works well for any window function.

overlap averaging

When more than 1 average is used to calculate the FFT, it is possible to use overlapping samples, as shown in Figure 16.

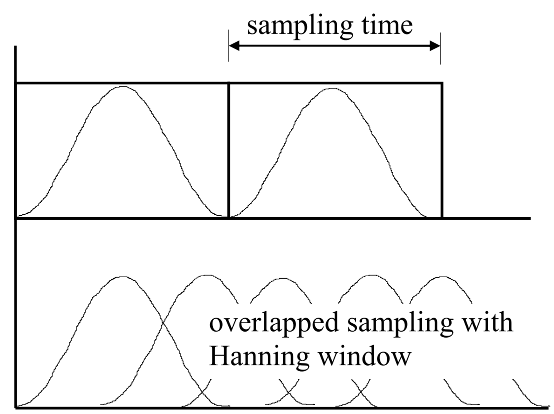


Figure 16. Overlapped Sampling

This works well since the first part and last part of the sample have their amplitudes reduced in normal averaging, while the overlapping sample takes full readings at these positions. The reduction in accuracy is very small, and for FFTs with a low Fmax and a lot of averages, collection times can be reduced considerably. For example, an FFT with 400 lines, an Fmax of 6000 CPM, and 8 averages without overlapping takes 32 seconds to gather the samples. With 50% overlap averaging, sampling requires only 18 seconds.

folded fft

When an FFT is calculated an array of numbers is generated, as shown in Figure 17.

real	n_1	n_2	n_3	n_4			n_{1022}	n_{1023}	n_{1024}
imaginary	0	0	0	0			0	0	0

n_1	n_2	n_3	n_4			n_{511}	n_{512}
n_{513}	n_{514}	n_{515}	n_{516}			n_{1023}	n_{1024}

Figure 17. Folded Array

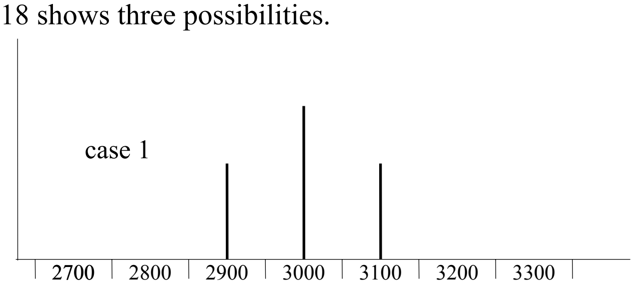
Each number is matched with an imaginary number . In our calculations, we need a number in every position, and there are n^2 multiplications, but we eventually throw away half of the answers. To optimize the time spent, we can replace all the imaginary numbers on input with real data and have an array half as long. This means approximately half the calculation time, as shown by the equation

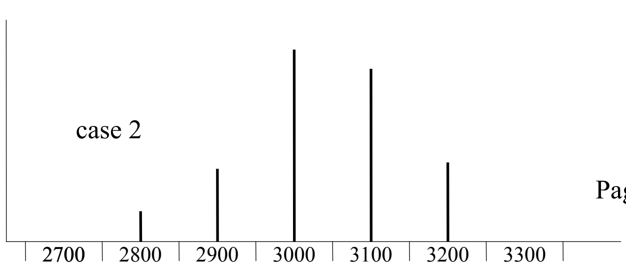
$$\frac{512 \ln (512)}{1024 \ln (1024)}$$

for this example, but we do need to add a little time to sort out the answers at the very end. There is still a very worthwhile saving in time.

12 IMPROVING FREQUENCY RESOLUTION

With the Hanning window, we know that whenever we have an amplitude showing up in a bin, we will have leakage into the bins adjacent to it and possibly in bins farther apart as well. Figure 18 shows three possibilities.





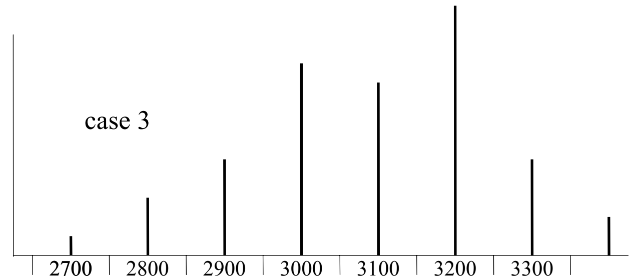


Figure 18. FFT Bin Leakage - 3 Cases

The first has equal height bins adjacent to the main frequency, the second has unequal leakage and the third demonstrates two components with close frequencies.

In the first case, as the adjacent bins are exactly equal, the frequency of the vibration is in the center of its bin -- exactly at 3000 CPM.

In the second case, as the leakage into the 3100 bin is larger, the exact frequency is not 3000 but is somewhere between 3000 and 3050 (where the bin ends). By calculation, it is possible to pin down the frequency of a component very accurately, giving 10 times or better resolution than the raw FFT.

In the third case, there are two large components at close frequencies. The only way to separate these is to use an FFT with higher resolution (e.g. Zoom), or to move the transducer to a position where one or other disappears.

13. OVERALLS, BANDS, and POWER CALCULATIONS

As a general rule, if you want to have a good measurement of an overall level, use the raw analog signal, or calculate it directly from the digitized signal. This is because power calculations (or overalls) from FFTs are subject to several errors.

If a signal has a significant low frequency component, a calculated value from the FFT may not see it because the zero bin and often the first bin are discarded (Figure 19).

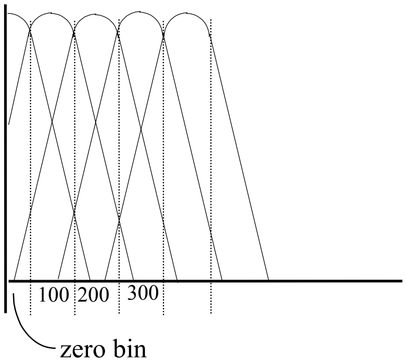


Figure 19. Zero Bin Suppression

The zero bin is discarded because the FFT circuitry is generally not DC coupled. The first bin may be discarded for example due to low end noise caused by integration. If there should be a significant component at these frequencies, its amplitude will be lost.

Some manufacturers of instrumentation use overalls calculated from the FFT because this is fast. Fast it may be; however, accuracy suffers. If insufficient settling time is allowed before signals are sampled, errors can be introduced into the FFT as well.

Secondly, an overall calculated from an FFT will ignore everything above Fmax. Generally analog overalls do not -- hence, there will be a difference.

The net result is that if you want to make sense of overalls that are generated from digital information, you need to know exactly what you are doing. With analog overalls, results are exact, and not subject to setup or interpretation.

Nevertheless, there are excellent reasons to calculate overalls from FFTs. One of these is band alarming where the energy in a frequency range is used as an indicator of trouble (Figure 20).

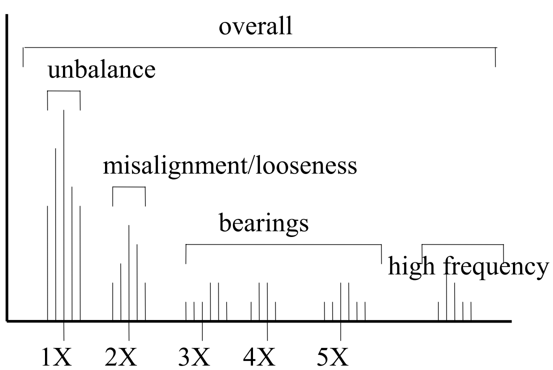


Figure 20. Frequency Bands

In the haystack or bearing areas for example, although any one reading may not be high, a high energy level in these frequencies is an indicator of trouble. Calculation of the overall level is done by a process known as the Root Sum Square. For an FFT which used a Hanning window, the formula is:

$$\text{Overall} = \sqrt{\frac{\sum \text{amplitude}^2}{1.5}}$$

The 1.5 factor is used to correct for the Hanning window characteristics and for leakage of signal into adjacent bins. It is possible to get an alarm without any of the individual lines in a band exceeding the limit. Note also that if the band is only 1 bin wide, there will be no leakage into adjacent bins. Hence, an absolute alarm level should be used rather than a calculated power. (A guideline to avoid this situation is to always set alarm limits to cover at least 4 bins of information.)

14. REAL TIME

The term real time is often applied to instruments on which the screen display changes quickly. While this is a requirement for real time analysis, it is not the whole story. Real time capability can be described as the highest rate at which data can be captured and displayed without leaving any gaps in the analysis. In other words, for FFTs this means the instrument must be capable of taking a full sample, and calculating and displaying that sample while the next sample is being captured. An example could be as follows:

For a 400-line FFT and an Fmax of 12,000 CPM, the sample rate is 512 Hz and each sampled window takes 2 seconds. If the screen updated every 2 seconds the unit could be said to have a real time rate of 200 Hz or 12,000 CPM -- the highest frequency which can be displayed.

Modern instruments today often refer to “live-time” displays. Graphical displays are presented in this format to observe measurements in progress. Generally, we can apply the following formula:

Real Time Rate = (#lines FFT)*(update rate)

example:

RTR = (400)*(8 times/sec.) = 3200 Hz.

15. CONCLUSIONS

The advent of affordable, reliable FFT data collectors and analyzers have introduced a number of new terms to the practicing vibration technician.

An understanding of these basic concepts in signal processing and data manipulation will enable one to select instrumentation and to understand its use.

In order to acquire accurate data, the vibration technician must carefully select the proper parameters: measurement units, Fmax, lines of resolution, averaging, bands/alarms, history size, etc.

The current trend in setting acceptable vibration limits is to use industry-proven severity charts specific to machine type and operating conditions.

16. REFERENCES

1. Bradley, Dan, "Introduction to FFT Terms and Parameters", IRD Mechanalysis, Inc, Technical Report #139, May 1989.

2. Catlin, John Jr., "The Significance of Averaging for Swept Filter and FFT Analysis of Machinery Vibration", Vibration Institute Proceedings, Tenth Meeting, June 1986.

3. Frarey, John L., "Concepts and Use of Real Time Analyzers", Vibrations, Vol. 2, No. 3, December 1986.

4. Harris and Crede, Shock and Vibration Handbook, Mc-Graw-Hill, New York, NY, 1976.

5. Gade and Hurlufsen, "Windows to FFT Analysis", Sound and Vibration, Vol. 22, No. 3, March 1988.

6. "The Applications of Filters to Analog and Digital Signal Processing", Wavetek, 1986.

7. "Signal Processing and Dual-Channel Guide", Nicolet Scientific Corp., 1982.