IEEE Instrumentation and Measurement
Technology Conference
Ottawa, Canada, May 19-21, 1997

Improved Procedures for Estimating Amplitudes and Phases of Harmonics with Application to Vibration Analysis

R. González-Blanco, I. Santamaría-Caballero, C. Pantaleón-Prieto,
Jesús Ibáñez-Díaz, Enrique Gómez-Cossío ∗
DICOM, ETSII Telecom. University of Cantabria, Avda. Los Castros, 39005 Santander, Spain
Phone: 34-42-201552, Fax: 34-42-201488, e-mail: labgts@dicom.unican.es
∗ SETEL S.A., Avda Marqués de Valdecilla s/n, Soto de la Marina, Cantabria, Spain

Abstract- **In this paper we propose two procedures for accurate amplitude and phase estimation of vibration signals of rotating machinery. The first method reduces the amplitude attenuation and phase shift caused by the "nonflat" top of the main lobe of the window. The second procedure is able to reduce not only the leakage effects due to windowing, but also the distortion in the amplitude and phase estimates when there is a slow change in the frequencies. Using an additional sensor giving one pulse per revolution, this method transforms the input (asynchronous) signal into a synchronous signal having a fixed number of samples per revolution. The performance of both procedures is illustrated by means of simulation examples.**

1. INTRODUCTION

Many large, slow speed, rotating machines are monitored by vibration analysis. A typical problem in these systems is the estimation of the frequencies and amplitudes of harmonics of the running speed. Usually, a change in the root mean square (rms) value of the vibration at an integer multiple of the running speed can indicate the development of a fault.

In some monitoring systems it is necessary to estimate not only the frequencies and amplitudes, but also the phases of each harmonic component. This happens, for instance, in systems using a couple of orthogonally-mounted transducers located in a bearing of the shaft. The objective of these systems is to obtain orbits describing the displacement of the shaft centerline.

A typical approach to solve the whole problem (frequency, amplitude and phase estimation) consists of two steps. First, the frequencies are obtained by applying a Fourier-based method, or a high resolution method if the harmonics cannot be resolved by Fourier techniques. Second, the amplitudes and phases are estimated by solving a linear least squares problem [1]. In vibration analysis, however, long data records are usually available and resolution is not a problem. Therefore, the frequencies can be estimated by selecting the largest peaks of the periodogram, which is implemented using an FFT, while the amplitudes and phases are estimated directly from the spectral lines.

However, since the FFT is evaluated in a grid of discrete frequencies, it introduces a bias in the frequency estimate when the sampling period is not an integer multiple of the fundamental period of the input signal. This poses a practical problem since even small errors in the frequency estimates can cause large errors in both amplitude and phase estimates [2]. Moreover, in rotating machines, fluctuations of the running speed cause a broadening of the spectral

[0]This work has been partially financed by BRITE-EURAM project 7289

lines, in addition to the broadening caused by windowing, which can degrade the estimates even more.

To overcome these problems in this paper we propose two procedures for obtaining accurate estimates of the amplitudes and phases. The first approach is based on the idea that selecting a shorter window, thus broadening its main lobe, can improve the amplitude and phase estimates. This approach is useful when we know in advance that there are small errors in the frequency estimates. In the paper we propose a procedure for obtaining the optimal window length for a given error.

The second technique can be applied when there are fluctuations of the running speed. It uses a tachometer which gives one pulse per revolution. This signal is used to transform the data acquired during an integer number of revolutions (asynchronously with the running frequency) into synchronous data. This means that in each revolution the vibration (or displacement) is measured at the same physical positions of the shaft, i.e., in each revolution we obtain exactly the same number of samples. The asynchronous to synchronous transformation is performed using oversampling plus linear interpolation techniques. After this transformation, the spectral lines obtained from the discrete Fourier transform (DFT) coincide with the harmonics; therefore, windowing and leakage effects are avoided and the amplitudes and phases are precise.

2. PROPOSED METHOD I

2.1. Short-Range Leakage

The response of the shaft of a rotating machine can be modeled by a sum of harmonics (or subharmonics of the running speed). Considering a sampling frequency $f_s = 1/T_s$ and that the signals are observed during T seconds, we have the following discrete-time signals at the transducer's output

$$x[n] = \sum_{i=1}^{p} A_i \cos(\omega_i n + \theta_i) + r[n] \qquad n = 0, \cdots, N-1; \tag{1}$$

where A_i are the amplitudes of the harmonics, θ_i are their phases, $r[n]$ is the measurement noise and $T = NT_s$.

Considering that the observation interval is long enough to resolve the harmonics, i.e.,

$$N >> (2\pi)/(\omega_{k+1} - \omega_k) \tag{2}$$

we can obtain the frequency estimates by selecting the p largest peaks of the Discrete Fourier Transform (DFT), which is given by

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{-j(2\pi k/N)n} \qquad k = 0, \cdots, N-1. \tag{3}$$

Since only a finite data register is available, when we estimate the amplitude and phase of the kth harmonic from the k spectral peak, two kinds of leakage errors may exist. The first one is concerned with the interference among the harmonics: the energy in the main lobe of a spectral component "leaks" into the sidelobes, distorting other spectral components. This kind of leakage is know as long-range leakage. In this paper we assume (2), therefore long-range leakage will not be considered.

The second kind of leakage is denoted as short-range leakage. It is caused by the nonflat-top main lobe of the window. When there is an error in the frequency estimate the amplitude and phase estimates are attenuated and shifted, respectively. Some techniques have been proposed to reduce the errors caused by short-range leakage. In [3], this kind of distortion is eliminated by synchronizing the sampling rate to the signal fundamental frequency. Other approaches obtain an accurate frequency estimate using the information given by two consecutive spectral lines surrounding the true frequency; then, the amplitude and phase can also be accurately estimated [2,4]. Finally, in [5] a flat-top window is proposed.

2.2. Optimal Window Length Selection

In this section we propose a simple procedure for obtaining accurate estimates of amplitudes and phases. It does not require the estimation of the frequency deviation neither additional hardware. This method is based on the idea that using a shorter window is equivalent to broadening its main lobe. Therefore, for the same frequency deviation, the amplitude and phase errors are reduced. This idea is illustrated in Fig. 1.

For simplicity, let us consider a discrete-time sinusoidal signal $x[n]$ of frequency ω_0, amplitude A_0 and phase θ_0. If we take L samples of $x[n]$, its spectrum is given by

$$X(\omega) = \frac{A_0}{2} e^{j\theta_0} W(\omega - \omega_0) + \frac{A_0}{2} e^{-j\theta_0} W(\omega + \omega_0) \tag{4}$$

where

$$W(\omega) = e^{-j\omega(L-1)/2} \frac{\sin(\omega L/2)}{\sin(\omega/2)} \tag{5}$$

is the Fourier transform of a rectangular window of length L.

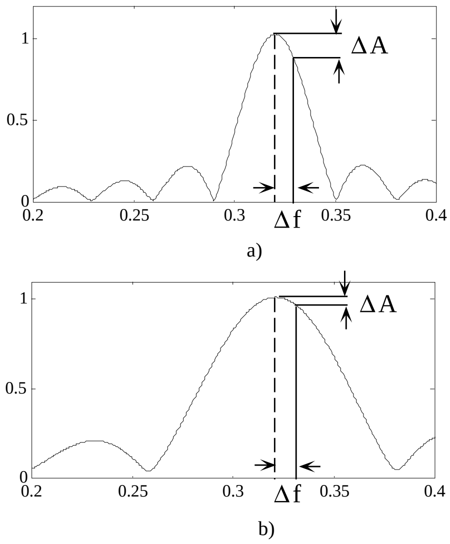

Figure 1. Comparison of the amplitude errors for two different window lengths: a) 32 samples, b) 16 samples.

Let $\hat{\omega}_0 = \omega_0 + \Delta\omega_0$ be the frequency estimate obtained by selecting the largest peak of the DFT of $x[n]$. This estimate is obtained in a previous step using a larger window of length N ($N > L$) to increase resolution. For a frequency deviation $\Delta\omega_0$, the amplitude and phase estimates are given by

$$\hat{A}_0 = \frac{A_0}{L}\frac{\sin(\Delta\omega_0 L/2)}{\sin(\Delta\omega_0/2)} \tag{6}$$

and

$$\hat{\theta}_0 = \theta_0 - \Delta\omega_0\frac{(L-1)}{2} \tag{7}$$

Our objective is to obtain the maximum window lengths L for a given amplitude error $\Delta A_{max} = |A_0 - \hat{A}_0|/A_0$. Considering that the frequency deviation is a small value, we can obtain a bound for the amplitude error by using the following Taylor series expansion

$$\begin{aligned} \sin(\Delta\omega_0/2) &= \frac{\Delta\omega_0}{2}, & (8)\\ \sin(\Delta\omega_0 L/2) &= \frac{\Delta\omega_0 L}{2} - \frac{1}{3!}\left(\frac{\Delta\omega_0}{2}L\right)^3 & (9) \end{aligned}$$

Substituting (8) and (9) in (6), we obtain that the maximum amplitude error can be bounded by

$$\Delta A_{max} \leq \frac{1}{6}\left(\frac{\Delta\omega_0}{2}\right)^2 L^2 \tag{10}$$

Now, assuming that the frequency estimate is obtained using the whole register length (N samples), the frequency deviation is less than half frequency bin, i.e.,

$$\Delta\omega_0 \leq \frac{\pi}{N} \tag{11}$$

Substituting (11) in (10), we obtain the maximum window length for a given error in amplitude

$$L \leq \frac{2\sqrt{6}}{\pi} N \sqrt{\Delta A_{max}} \tag{12}$$

Similar expressions can be obtained for other common windows in spectral analysis (Hamming, Hanning, Blackman,etc.). Nevertheless, the bound (12), obtained for a rectangular window, can be considered as a worst case. Any other window reduces the leakage (short and long-range), but also reduces the resolution.

2.3. The Overall Method

Finally, the proposed method can be summarized in the following steps:

1. Select the maximum admissible amplitude error: ΔA_{max}.
2. From the register $x[n]$ of length N ($n = 0, \cdots, N-1$), estimate the frequencies $\omega_1, \cdots \omega_p$ as the largest peaks of the periodogram.
3. Select a new shorter window of length L given by (12)
4. For $k = 1, \cdots, p$ estimate the amplitude and phase of the kth harmonic as

$$X(\hat{\omega}_k) = \sum_{n=0}^{L-1} x[n]\mathrm{e}^{-j\hat{\omega}_k n} \tag{13}$$

and

$$\begin{aligned} \hat{A}_k &= \frac{2}{L}\left|X(\hat{\omega}_k)\right| & (14)\\ \hat{\theta}_k &= \arg\left(X(\hat{\omega}_k)\right). & (15) \end{aligned}$$

Let us expose some comments about the proposed method. First, the selected window length L must be long enough to resolve clearly the harmonics and to make the long-range leakage negligible, i.e., $L >> 2\pi/(\omega_{k+1} - \omega_k)$. Second, from (6) and (7) it can be seen that short-range leakage distorts the phase estimates more than the amplitude estimates. Finally, in high noise situations the estimates obtained using different (maybe overlapped) windows of length L should be averaged to improve the performance.

3. PROPOSED METHOD II

3.1. Description

In this section we propose an alternative procedure for accurate amplitude and phase estimation. It is able to reduce not only the distortion caused by short-range leakage, but also the effects of long-range leakage. Moreover, it provides a solution in situations for which the running frequency changes slowly. This method requires an additional sensor (tachometer) giving a once-per-turn pulse. This signal is used to synchronize (approximately) the transducer's output with the running frequency.

To give an intuitive idea about how the method works, let us consider that the running frequency of the shaft is changing, in this case we will observe a different number of samples in each revolution. Since we know the starting and ending points of each revolution (i.e., the pulse locations given by the tachometer), we can interpolate the signal in order to obtain a fixed number of samples in each revolution. Using this procedure we have transformed a signal acquired asynchronously with the running frequency into a synchronous signal acquired with a sampling rate which is an integer multiple of the running frequency. We will denote this procedure asynchronous to synchronous conversion.

Therefore, the proposed method can be summarized in the following steps: first, the signal is acquired during an integer number of revolutions; second, asynchronous to synchronous conversion is performed. Finally, the frequencies are estimated as the largest peaks of the periodogram, while the amplitudes and phases are estimated from the spectral lines. Let us note that even if the running frequency does not change the acquisition of an integer number of revolutions avoids leakage effects.

3.2. Asynchronous to Synchronous Conversion

In this section we evaluate the distortion caused by the asynchronous to synchronous conversion when we use linear interpolation to obtain the new samples. Let us consider a sinusoidal analog signal of frequency f_m (the frequency of the maximum harmonic of interest), and a sampling frequency $f_s = 1/T_s$. Assuming that the tachometer provides the true time instants in which each revolution starts, the first step in the conversion consists of finding the new sampling instants by locating f_s uniformly spaced samples between consecutive tachometer pulses. To obtain the new signal values any interpolation technique could be used.

In particular, if we use linear interpolation, the maximum interpolation error will occur at times $T_s/2$. It can be shown [6] that this error is given by

$$\Delta e_{max} = A(1 - \cos(\pi f_m/f_s)) \tag{16}$$

The resultant signal after the conversion can be viewed as the true signal plus some additive noise due to non-ideal interpolation. A reasonable model for this noise is a uniform distribution within the interval $[-\Delta e_{max}, \Delta e_{max}]$; therefore its variance is given by

$$\sigma_e^2 = \frac{A^2(1 - \cos(\pi f_m/f_s))^2}{3} \tag{17}$$

The distortion due to non-ideal interpolation can be measured in terms of Signal to Noise Ratio (SNR)

$$SNR = 1.76 - 20\log\left(1 - \cos(\pi f_m/f_s)\right) \tag{18}$$

For example, if $f_s = 2f_m$ (the Nyquist sampling rate), then $SNR = 1.76$ dB and it becomes clear that the performance of any frequency, amplitude and phase estimation procedure can be reduced. This puts in evidence the necessity of using oversamplig: the signals must be acquired at a sampling rate higher than the Nyquist rate.

As a final example, if we want to keep the distortion lower than 80 dB (representing the signal with a finite resolution of 12 bits), the required oversamplig ratio is $M \approx 100$. It is possible to reduce this ratio using a higher order interpolator.

Finally, we want to remark that the above worst case analysis is very pessimistic. In practice, an oversamplig ratio of $M = 10$ leads to accurate estimates using linear interpolation.

4. SIMULATION RESULTS

4.1. Example I

In this section we present some simulation results in order to evaluate the performance of the first method. We generated a sinusoidal signal composed of four harmonics with frequencies f_1=18 Hz, f_2=36 Hz, f_3=54 Hz and f_4=72 Hz; amplitudes A_1=1, A_2=0.5, A_3=0.25 and A_4=0.125; and phases $\theta_1 = \pi/2$, $\theta_2 = \pi/4$, $\theta_3 = -0.3\pi$ and $\theta_4 = 0.6\pi$. We acquire 2048 samples of this signal using a sampling frequency of f_s=245 Hz, and finally we added white Gaussian noise to obtain a final SNR of 20 dB.

This represents a case where, due to the high number of samples acquired, a small error in the frequency estimate can lead to larger errors in amplitudes and phases. Applying the proposed method we obtain that using a shorter window of length $N = 325$, it is possible to keep the amplitude estimation error lower than 1%. This value is only approximated since due to noise and long-range leakage the estimates are distorted.

Table 1 shows the results obtained averaging 500 independent simulations. The first row shows the mean values of the amplitude estimates and below the error percentage when we use the whole register length $N = 2048$. The second row shows the results obtained with the proposed method which uses a shorter window of length $N = 325$. In particular, the estimates of the amplitude of the fundamental frequency and the third harmonic are clearly improved.

	A_1	A_2	A_3	A_4
N=2048	0.681	0.497	0.191	0.121
	31.9%	0.6%	23.6%	3.2%
N=325	0.992	0.495	0.256	0.120
	0.8%	1.3%	3.6%	4.9%

Table 1. Table 1: Mean values and error percentage (below) for the amplitude estimates using the whole register length N=2048 and using the proposed method with N=325.

For the phase estimates the improvement is even more noticeable. These results are shown in Table 2.

	θ_1	θ_2	θ_3	θ_4
N=2048	3.03	0.564	0.298	1.449
	93.1%	28.1%	131.6%	23.1%
N=325	1.57	0.785	-0.942	1.88
	15.1%	5.2%	19.5%	5.3%

Table 2. Table 1: Mean values and error percentage (below) for the phase estimates using the whole register length N=2048 and using the proposed method with N=325.

4.2. Example 2

In this second example we illustrate the effects of a change in the running frequency and we show that the asynchronous to synchronous conversion can reduce this distortion. In order to model the fluctuation of the running speed, we consider that each harmonic component of the signal generated in the previous example changes according to the following FM model:

$$f_i(t) = f_i(1 + 0.01 sin(2\pi/25t)) \qquad (19)$$

In order to improve the asynchronous to synchronous conversion we acquire 50 revolutions of the signal at a sampling frequency f_s =1225 Hz. This means an oversampling ratio of approximately $M = 8.5$ since the maximum frequency of interest is the fourth harmonic $f_4 = 72$ Hz. We perform the conversion fixing 72 samples per revolution, in this way the running frequency and the final sampling frequency are synchronized. The spectrum of the signals before and after the conversion are shown in Figs. 2 and 3, respectively. The advantages of the proposed procedure in the presence of fluctuations of the running frequency are obvious.

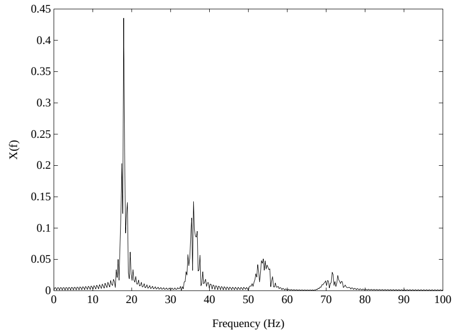

Figure 2. Spectrum obtained without asynchronous to synchronous conversion.

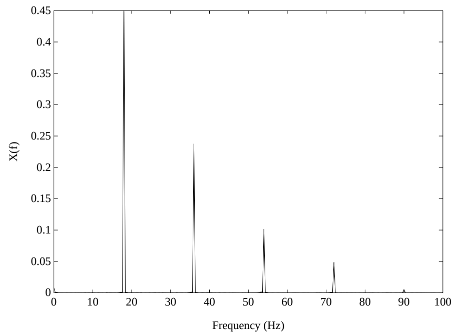

Figure 3. Spectrum obtained after asynchronous to synchronous conversion.

5. CONCLUSIONS

In this paper we have proposed two improved procedures for amplitude and phase estimation of sinusoidal signals in vibration analysis. The first one allows to reduce the short-rage leakage effects by working with a window with a broader main lobe. In vibration analysis of rotating machinery, where the acquired registers can be very long, this method achieves a noticeable improvement. The second method performs an asynchronous to synchronous conversion to get a fixed number of samples per revolution. In this way, it is possible to synchronize the sampling frequency with the fundamental one. Therefore, leakage effects are avoided. By means of a simulation example we have shown that this method obtains accurate estimates even when there is a slow change in the frequency components.

REFERENCES

[1] S. M. Kay, *Modern Spectral Estimation*, Prentice Hall, Englewood Cliffs, New Jersey, 1988.

[2] X. Ming, D. Kang, "Corrections for frequency, amplitude and phase in a fast Fourier transform of a harmonic signal", *Mechanical Systems and Signal Processing*, vol. 10, no. 2, pp. 211-221, 1996.

[3] A. Ferrero, R. Ottoboni, "High-accuracy Fourier analysis based on synchronous sampling techniques", *IEEE Trans. Instrum. Meas.*, vol. 41, pp. 780-785, Dec. 1992.

[4] H. Dishan, "Phase error in fast Fourier transform analysis", *Mechanical Systems and Signal Processing*, vol. 9, no. 2, pp. 113-118, 1995.

[5] X. Dai, I. R. Gretsch, "Quasi-synchronous sampling algorithm and its applications", *IEEE Trans. Instrum. Meas.*, vol. 43, pp. 204-209, April 1994.

[6] T. A. Ramstad, "Digital methods for conversion between arbitrary sampling frequencies", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 32, pp. 577-591, 1984.