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

Ignacio Santamaría-Caballero, Member, IEEE, Carlos J. Pantaleón-Prieto, Jesús Ibáñez-Díaz, Enrique Gómez-Cosío

Abstract—In this paper, we propose two procedures for accurate amplitude and phase estimation of multifrequency signals in 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 that appears when the rotation frequency changes slowly. This second method uses an additional sensor, giving one pulse per revolution, to transform the input (asynchronous) signal into a synchronous signal having a fixed number of samples per revolution. The performance and effectiveness of both procedures are illustrated by means of simulation examples.

Index Terms—Amplitude estimation, discrete Fourier transforms, frequency estimation, phase estimation, rotating machine measurements, spectral analysis.

I. 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 at 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 lines, in addition to the leakage caused by windowing, which introduces an even larger degradation.

Most of the techniques proposed in the literature to reduce the errors caused by leakage are based on interpolating the FFT samples surrounding the true frequency [3]–[6]. On the other hand, in [7] this kind of distortion is eliminated by synchronizing the sampling rate to the signal fundamental frequency. Finally, in [8] a flat-top window is proposed.

In this paper, we propose two procedures for obtaining accurate estimates of the amplitudes and phases. The first approach is based on the simple idea that selecting a shorter window, thus broadening its main lobe, can improve the amplitude and phase estimates. A procedure for obtaining an estimate of the window length that leads to a given error is presented.

The second technique can be applied when there are fluctuations of the running speed; in this case only the method described in [7] provides reasonable estimates. The alternative proposed in the paper uses a tachometer which gives one pulse per revolution. This signal is used to transform the input signal taken at equispaced time intervals into a signal sampled at equispaced angular intervals. 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 resampling procedure is performed using oversampling plus linear interpolation techniques. After this transformation, the discrete Fourier transform (DFT) evaluates the spectrum at integer multiples of the running frequency; therefore, windowing and leakage effects are avoided and the amplitudes and phases are accurate.

II. PROPOSED METHOD I

A. 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 \( f_s = 1/T \) and that the signals are observed during \( T \) seconds, we have the following discrete-time signals at the transducer's output:

\[
x[n] = \sum_{i=1}^{P} A_i \cos(\omega_i n + \theta_i) + \eta[n] \quad n = 0, \ldots, N - 1
\]

where \( A_i \) are the amplitudes of the harmonics, \( \theta_i \) are their phases, \( \eta[n] \) is the measurement noise and \( T = NT_s \).

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

\[
N \gg \frac{(2\pi)}{(\omega_{k+1} - \omega_k)}
\]

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

\[
X[k] = \sum_{n=0}^{N-1} x[n] e^{-j(2\pi k/n)n} \quad k = 0, \ldots, N - 1.
\]

Since only a finite data register is available when we estimate the amplitude and phase of the \( k \)th 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 known 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.

To illustrate this effect, let us consider a discrete-time sinusoidal signal \( x[n] \) of frequency \( \omega_0 \), amplitude \( A_0 \) and phase \( \theta_0 \). If we take \( L \) samples of \( x[n] \), its spectrum is given by

\[
X(\omega) = \frac{A_0}{2} e^{j\theta_0} W(\omega - \omega_0) + \frac{A_0}{2} e^{j\theta_0} W(\omega + \omega_0)
\]

where

\[
W(\omega) = e^{-j\omega(L-1)/2} \frac{\sin(\omega L/2)}{\sin(\omega/2)}
\]

is the Fourier transform of a rectangular window of length \( L \).

Now, let us assume that \( \Delta \omega_0 \) is the error in the frequency estimate; then, the amplitude and phase estimates are given by

\[
\hat{A}_0 = \frac{A_0}{L} \frac{\sin(\Delta \omega_0 L/2)}{\sin(\Delta \omega_0/2)}
\]

and

\[
\hat{\theta}_0 = \theta_0 - \frac{(L-1)}{2} \Delta \omega_0.
\]

From (6) and (7) it is clear that if the length of the window is large, even a small error in frequency can lead to large errors in both amplitude and phase estimates. In fact, since the phase error increases linearly with \( \Delta \omega_0 \), phase estimates can be useless without correction.

### B. Optimal Window Length Selection

In this section we propose a simple procedure for improving the estimates of the amplitudes and phases. It does not require the estimation of the frequency deviation; neither additional hardware. The 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.

Our objective is to obtain the maximum window length \( L \) for a given amplitude error \( \Delta A_{\text{max}} = |A_0 - \hat{A}_0|/A_0 \). Considering that the frequency deviation is a small value \( \Delta \omega_0 L/2 < 1 \), we can obtain a bound for the amplitude error by using the following Taylor series expansion

\[
\sin(\Delta \omega_0/2) = \frac{\Delta \omega_0}{2} - \frac{1}{3!} \left( \frac{\Delta \omega_0}{2} \right)^3.
\]

Substituting (8) and (9) in (6), we obtain

\[
\Delta A_{\text{max}} \leq \frac{1}{6} \left( \frac{\Delta \omega_0}{2} \right)^2 L^2.
\]

Now, assuming that the frequency estimate was obtained in a previous step using the whole register length (\( N \) samples), the frequency deviation is less than half frequency bin, i.e., \( \Delta \omega_0 \leq \pi/N \). Substituting this value in (10), we obtain the maximum window length for a given error in amplitude

\[
L \leq \frac{2\sqrt{6}}{\pi} N \sqrt{\Delta A_{\text{max}}}
\]

For instance, if we have a register of \( N = 2048 \) samples, the above expression indicates that we should select a window of approximately \( L = 325 \) samples in order to keep the amplitude error below 1%. The use of suitable windows can reduce the number of samples needed to reduce the interference between harmonics to a negligible level.

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

1. Select the maximum admissible amplitude error: \( \Delta A_{\text{max}} \).
2. From the register \( x[n] \) of length \( N \) (\( n = 0, \ldots, N - 1 \)), estimate the frequencies \( \omega_1, \ldots, \omega_p \) as the largest peaks of the periodogram.
3. Select a new shorter window of length \( L \) (\( n = 0, \ldots, L - 1 \)) given by (11).
4. For \( k = 1, \ldots, p \) estimate the amplitude and phase of the \( k \)th harmonic as

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

and

\[
\hat{A}_k = \frac{2}{L} |X(\hat{\omega}_k)|
\]
\[ \hat{\theta}_k = \arg(X(\omega_k)). \] (14)

A straightforward improvement of this method consists of averaging the amplitude and phase estimates obtained using different (maybe overlapped) windows of length \( L \). In order to average the phase estimates, it is necessary to take into account the additional shift produced between delayed windows.

### III. Proposed Method II

#### A. Asynchronous to Synchronous Conversion

In this section we propose an alternative procedure for accurate amplitude and phase estimation which is able to eliminate completely the distortion caused by leakage. Unlike the previous method, it also provides a solution for situations where the running frequency changes slowly. This second method requires an additional sensor (tachometer) giving a once-per-turn pulse. The information given by this new sensor 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 technique 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 (ATSC).

In fact, the ATSC method is an off-line resampling procedure to transform a signal sampled at equispaced time intervals into a signal sampled at equispaced angular intervals. This transformation can be carried out using the ideal interpolation schemes derived in [9] and [10], but as we will show in the next subsection a simpler procedure consists of using oversampling plus linear interpolation.

Finally, the proposed method can be summarized in the following steps: first, the signal is acquired during an integer number of revolutions; second, the asynchronous to synchronous conversion is performed. Finally, the frequencies, amplitudes and phases are estimated from the largest peaks of the DFT. Let us note that even if the running frequency changes slowly, this second method requires an additional sensor (tachometer) giving a once-per-turn pulse. The information given by this new sensor is used to synchronize (approximately) the transducer’s output with the running frequency.

#### B. ATSC Analysis

In this section we evaluate the distortion caused by the ATSC procedure when we use linear interpolation to obtain the new samples. To simplify the analysis, let us consider a single sinusoidal signal of frequency \( f_m \) (the frequency of the maximum harmonic of interest), amplitude \( A \) and a sampling frequency \( f_s = 1/T_s \) (\( t = 0, T_s, 2T_s, \cdots \)). The running frequency of the machine is \( f_0 = 1/T_{0} \) and \( f_m = m f_0 \).

Assuming that the tachometer provides the true time instants in which each revolution starts, the first step of the ATSC consists of finding the new sampling instants by locating \( N_r \) uniformly spaced samples between consecutive tachometer pulses \( t'_r = 0, T_0/N_r, \cdots, (N_r - 1)T_0/N_r, \cdots \). If \( f_0 \) does not change with time, the ATSC reduces to a conventional sampling rate change from \( f_s \) to \( f'_s = N_r f_0 \). In this case, the acquisition of an integer number of revolutions avoids leakage due to windowing. On the other hand, if \( f_0 \) changes with time, the final sampling rate will change accordingly: in this case the ATSC eliminates, in addition to leakage, the spectral broadening caused by the fluctuations of \( f_0 \).

To obtain the new samples at \( t' \) any interpolation technique could be used [9], [10]. In particular, if we use linear interpolation, the maximum interpolation error will occur at times \( t' = t + T_s/2 \), i.e., when the new time instant lies halfway between two consecutive samples of the original signal. It can be shown [11] that this error is given by

\[ \Delta \theta_{\text{max}} = A(1 - \cos(\pi f_m/f_s)), \] (15)

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

\[ \sigma^2_{\text{e}} = \frac{A^2(1 - \cos(\pi f_m/f_s))^2}{3}. \] (16)

Taking into account that the variance of the signal is \( \sigma^2_{\kappa} = A^2/2 \), the distortion due to nonideal interpolation can be measured in terms of signal-to-noise ratio (SNR)

\[ \text{SNR} = 10 \log \left( \frac{\sigma^2_{\kappa}}{\sigma^2_{\text{e}}} \right) = 1.76 - 20 \log(1 - \cos(\pi f_m/f_s)). \] (17)

For example, if \( f_s = 2 f_m \) (the Nyquist sampling rate), then SNR = 1.76 dB and it becomes clear that we would obtain a reduced performance from any frequency, amplitude and phase estimation procedure.

As a final example, if we want to keep the distortion lower than 80 dB (representing the signal with a resolution of 12 bits), the required oversampling ratio is \( M \approx 100 \). It would be 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, to use linear interpolation with an oversampling ratio within the interval \( 5 < M < 10 \), leads to highly accurate estimates.

### IV. Simulation Results

#### A. Example I

In this section we present some simulation results in order to evaluate the performance of both methods. 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 \( \theta_1 = \pi/2, \theta_2 = \pi/4, \theta_3 = -0.5 \pi, \) and \( \theta_4 = 0.5 \pi \).

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.

Tables I and II show the results for amplitude and phase estimation obtained averaging 500 independent simulations, respectively. In these tables we perform a comparison of the mean values and the percentage error (below) obtained by the following algorithms.

- **Raw FFT**: amplitude and phase estimation using the whole register length $N = 2048$.
- **Interpolated FFT (Hanning window) [6]**: using two spectral lines surrounding the true frequency, the frequency deviation is estimated and then, the amplitude and phase estimates are corrected. A Hanning window was used to reduce the spectral leakage.
- **Proposed method I**: the frequencies were obtained using the whole register length; then, the sequence $x[n]$ was divided into segments of length $L = 325$ samples with a 25% of overlapping. The amplitudes and phases of the harmonics were obtained by averaging the estimates obtained from all the segments.
- **Proposed method II (ATSC)**: the signal was over-sampled by a factor of 5 giving a sampling rate of 1225 Hz; then the ATSC was performed fixing $N_r = 14$ samples per revolution. Taking into account that the rotation frequency is 18 Hz, the final sampling frequency is 252 Hz.

The results obtained for the amplitude $A_3$ and the phase $\theta_4$ are detailed in Figs. 1 and 2, respectively, which show the mean square error (MSE) versus the SNR. For comparison we have included the Cramer–Rao bound indicating the minimum variance that can be attained by any unbiased estimator [12]. We can conclude that the proposed method I gives similar results than the interpolated FFT when a suitable window is used [6], but with a slightly lower computational cost. Moreover, the proposed method I does not require the estimation of the frequency deviation. In this example, due to the large number of acquired samples ($N = 2048$), the use of windows with sidelobes lower than the Hanning window does not provide a noticeable improvement in the interpolated FFT method. Finally, the best results are obtained by the ATSC method, which attains the Cramer–Rao bound for SNR’s over 2 dB.

### Table I

<table>
<thead>
<tr>
<th></th>
<th>$A_1 = 1$</th>
<th>$A_2 = 0.5$</th>
<th>$A_3 = 0.25$</th>
<th>$A_4 = 0.125$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw FFT</td>
<td>0.681</td>
<td>0.397</td>
<td>0.191</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>31.4%</td>
<td>6.5%</td>
<td>23.6%</td>
<td>3.2%</td>
</tr>
<tr>
<td>Interp.</td>
<td>1.000</td>
<td>0.500</td>
<td>0.250</td>
<td>0.125</td>
</tr>
<tr>
<td>FFT (Hanning) [6]</td>
<td>0.2%</td>
<td>0.5%</td>
<td>1.0%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.992</td>
<td>0.502</td>
<td>0.250</td>
<td>0.125</td>
</tr>
<tr>
<td>method I (N=325)</td>
<td>0.8%</td>
<td>0.5%</td>
<td>0.9%</td>
<td>1.9%</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.999</td>
<td>0.498</td>
<td>0.249</td>
<td>0.124</td>
</tr>
<tr>
<td>method II: ATSC</td>
<td>0.2%</td>
<td>0.4%</td>
<td>0.7%</td>
<td>1.7%</td>
</tr>
</tbody>
</table>

### Table II

<table>
<thead>
<tr>
<th></th>
<th>$\theta_1 = 1.571$</th>
<th>$\theta_2 = 0.785$</th>
<th>$\theta_3 = -0.943$</th>
<th>$\theta_4 = 1.885$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw FFT</td>
<td>3.033</td>
<td>0.655</td>
<td>0.299</td>
<td>1.449</td>
</tr>
<tr>
<td></td>
<td>93.1%</td>
<td>28.1%</td>
<td>131.8%</td>
<td>23.1%</td>
</tr>
<tr>
<td>Interp.</td>
<td>1.571</td>
<td>0.784</td>
<td>-0.941</td>
<td>1.887</td>
</tr>
<tr>
<td>FFT (Hanning) [6]</td>
<td>0.3%</td>
<td>1.8%</td>
<td>2.6%</td>
<td>3.2%</td>
</tr>
<tr>
<td>Proposed</td>
<td>1.459</td>
<td>0.795</td>
<td>-1.052</td>
<td>1.895</td>
</tr>
<tr>
<td>method I (N=325)</td>
<td>7.1%</td>
<td>1.7%</td>
<td>11.6%</td>
<td>2.3%</td>
</tr>
<tr>
<td>Proposed</td>
<td>1.570</td>
<td>0.785</td>
<td>-0.942</td>
<td>1.885</td>
</tr>
<tr>
<td>method II: ATSC</td>
<td>0.1%</td>
<td>0.3%</td>
<td>0.7%</td>
<td>0.8%</td>
</tr>
</tbody>
</table>

**Fig. 1.** Mean square error of $A_3$ versus SNR.

**Fig. 2.** Mean square error of $\theta_4$ versus SNR.
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:

\[ f_i(t) = f_i(1 + 0.02 \sin(2\pi(18/25)t)). \quad (18) \]

In order to show the performance of the ATSC we use the same parameters described in the previous example (SNR = 20 dB). The spectrum of the signals before and after the conversion are shown in Figs. 3 and 4, respectively. The advantages of the proposed procedure in the presence of fluctuations of the running frequency are obvious.

To confirm this idea, Tables III and IV show the results for amplitude and phase estimation obtained by averaging 500 independent simulations, respectively. As in the previous example, they indicate the mean values and percentage errors obtained by a raw FFT, an interpolated FFT using a Hanning window [6] and the proposed methods.

### V. 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-range 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 (equivalent to use a windowed/interpolated FFT), with a moderate increase in the computational cost. The second method performs an asynchronous to synchronous conversion (ATSC) to get a fixed number of samples per revolution by means of linear interpolation. In this way, it is possible to synchronize the sampling frequency with the fundamental one. By means of some simulation examples we have shown that this method obtains better estimates than the previous one, even when there is a slow change in the frequency components. However, the ATSC procedure has a higher computational cost and requires additional hardware.

### References


Ignacio Santamaría-Caballero (M’96) was born in Vitoria, Spain, in 1967. He received the Telecommunication Engineer degree and the Doctoral degree from the Universidad Politécnica de Madrid, Madrid, Spain, in 1991 and 1995, respectively.

In 1992, he joined the Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, Santander, Spain, where he is currently an Associate Professor. His research interests include digital signal processing, digital communication systems, and neural networks.

Carlos J. Pantalón-Prieto was born in Badajoz, Spain, in 1966. He received the Telecommunication Engineer degree and the Doctoral degree from the Polytechnic University of Madrid, Madrid, Spain, in 1990 and 1994, respectively.

He is currently Associate Professor at the Universidad de Cantabria, Santander, Spain. His research interests include nonlinear models, neural networks, and digital signal processing.

Ignacio Santamaría-Caballero

Carlos J. Pantalón-Prieto

Jesús Ibáñez-Díaz was born in Santander, Spain, in 1971. He received the Radiocommunication Bachelor Engineer degree and the Telecommunication Engineer degree from the Universidad de Cantabria, Santander, Spain, in 1992 and 1995, respectively.

In 1995, he joined the Departamento de Ingeniería de Comunicaciones, Universidad de Cantabria, where he is currently an Associate Professor. His research interest include digital signal processing, digital communication systems, and nonlinear systems.

Jesús Ibáñez-Díaz

Enrique Gómez-Cosío, photograph and biography not available at the time of publication.