MSSP—20001321 Susan yamuna BG
Mechanical Systems and Signal Processing (2000) **14**(0), 000–000
doi:10.1006/mssp.2000.1321, available online at http://www.idealibrary.com on IDEAL®

A COMPARATIVE STUDY OF HIGH-ACCURACY FREQUENCY ESTIMATION METHODS

IGNACIO SANTAMARÍA, CARLOS PANTALEÓN AND JESÚS IBAÑEZ

DICOM, ETSII y Telecommunication University of Cantabria, Avda, Los Castros, 39005 Santander, Spain.
E-mails: nacho@gtas.dicom.unican.es, carlos@gtas.dicom.unican.es, jesus@gtas.dicom.unican.es

This paper presents a comparative study of three high-accuracy frequency estimation methods for application in vibration analysis of rotating machinery. The first two techniques are non-parametric methods based on the fast Fourier transform (FFT) : the interpolated fast Fourier transform (IFFT) and the iterative weighted phase averager (IWPA). The third method is a parametric high-resolution technique known as ESPRIT. The FFT-based methods combine techniques to reduce the effects of windowing with an iterative procedure which, at each iteration, detects the strongest peak and subtracts its effect (to reduce the interference resulting from spectral leakage). The paper compares their variance, resolution and computational requirements by means of simulation examples and also using end winding vibration data taken from a hydroelectric turbogenerator. It is found that, in situations with a moderate or high level of spectral interference, the IWPA method outperforms the IFFT and is even competitive with ESPRIT. Moreover, the IWPA has the ability to separate sinusoids more closely spaced than the periodogram's resolution limit. The IFFT method, on the other hand, has the lowest computational cost.

1. INTRODUCTION

A typical problem in vibration analysis of rotating machinery consists of estimating the frequencies of a collection of sampled sinusoidal signals contaminated by additive noise. Usually, the detection of a new spectral peak or a change in the root-mean-square (rms) value of the vibration at an integer multiple of the rotation frequency can indicate the development of a fault.

In vibration analysis, long data records are typically acquired and therefore a simple fast Fourier transform (FFT) algorithm provides enough resolution for frequency estimation. Moreover, the requirement of real-time analysis is another reason that explains why classical frequency estimation methods such as the periodogram (probably windowed and averaged) still remain the most used tools in vibration analysis and monitoring of rotating machinery.

However, as it is shown in this paper, if we can afford a slight increase in the computational cost, it is possible to obtain high-accuracy frequency estimates working with a relatively short data set. Specifically, in this paper we compare the performance of three of these high-accuracy methods: two of them are based on the FFT and the third is a high-resolution method.

Most of the frequency estimation methods [1, 2] can be grouped into two classes: parametric or high-resolution methods and non-parametric or periodogram-based methods. The high-resolution methods are able to resolve spectral peaks separated in

0888–3270/00/000000 + 00 $35.00/0

frequency less than $1/T$ (T being the observation time of the signal), which is the resolution limit for the methods based on the periodogram. To this class pertain methods such as Pisarenko's method, the MUSIC and the Root-MUSIC, the Min-Norm method and more recently, the ESPRIT method [3]. All the high-resolution methods provide very accurate frequency estimates, with only small differences in their statistical properties. For our comparative study, we have selected the ESPRIT method for two reasons: first, it provides slightly more accurate estimates than the other methods [1], and second, ESPRIT is not widely known to the vibration analysis community.

On the other hand, the main advantage of classical periodogram-based methods, which employ FFT algorithms, is that they have a low computational cost and, therefore, can be efficiently implemented. However, in general, periodogram- or FFT-based methods cannot resolve closely spaced frequencies.

To obtain high-accuracy estimates with FFT-based approaches we have to deal with two different kinds of problems: first, it is the problem of spectral leakage due to interference among the sinusoidal components, which occurs even when tapered windows are used. This kind of leakage is sometimes denoted as long-range leakage. Second, since the FFT is evaluated in a grid of discrete frequencies, it introduces a bias in the frequency estimates. The classical method of zero padding to mitigate this effect (also referred to as short-range leakage) may be computationally expensive if high accuracy is required.

Several techniques have been proposed to overcome the short-range leakage problem. The simplest group of methods comprises those based on interpolating the FFT samples surrounding the true frequency [4–7]. Interpolated FFT (IFFT) methods have the advantage that they are easy to implement and very fast. However, when the sinusoids are not well separated in frequency, their results are not very accurate. As a second example for our comparative study, we have selected the original Rife's IFFT method [4, 5].

Finally, another family of high-accuracy frequency estimation methods based on the FFT, comprises those relying on some type of weighted linear regression on the phase data [8, 9]. They are denoted as weighted phase averager (WPA) methods. The WPA methods are not widely known, eventhough they achieve, in general, better results than the IFFT methods. One of the goals of this paper is to describe these high-accuracy methods and also to propose a modification denoted as the iterative weighted phase averager (IWPA), which is able to provide accurate frequency estimates when the sinusoids are not well separated in frequency.

The rest of the paper is organised as follows. In Section 2, we describe in some detail the proposed IWPA algorithm. The IFFT method and the ESPRIT algorithm are summarised in Sections 3 and 4, respectively. In Section 5, the performances of these three methods in different scenarios are compared using simulation examples. In Section 6, we apply these methods to a real vibration analysis problem using end winding vibration data taken from a hydroelectric turbogenerator. Finally, the main conclusions are summarised in Section 7.

2. THE IWPA METHOD

2.1. SHORT-RANGE LEAKAGE CORRECTION

In order to explain the basis of the frequency estimation procedure, let us consider, first, the case of a single sinusoidal signal

$$x[n] = A_0 \cos(2\pi f_0 n + \theta_0), \quad n = 0, \ldots, N-1 \tag{1}$$

where (f_0, A_0, θ_0) are unknown parameters. Throughout this paper we will use discrete frequencies f and discrete-time (sampled) sequences. The corresponding analog frequency F_0

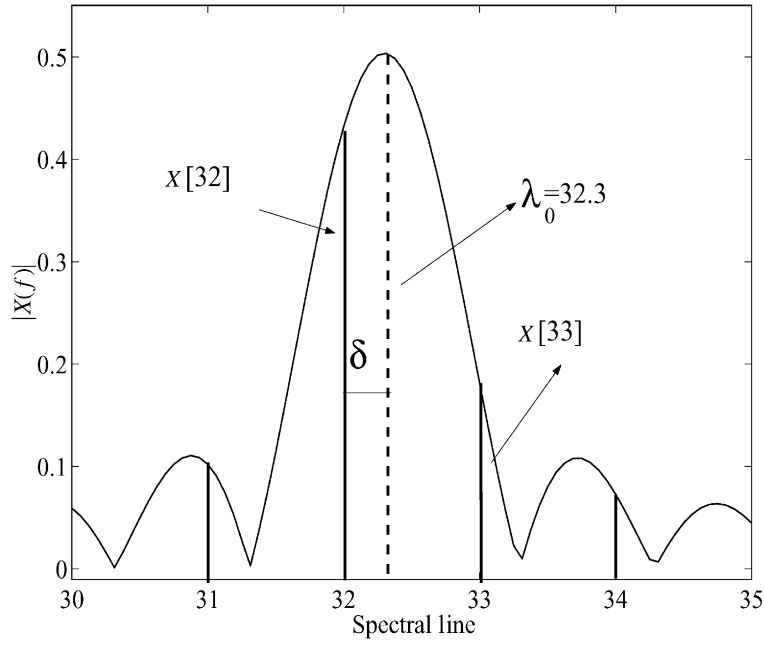

Figure 1. Fourier transform of a windowed sinusoid of frequency $f_0 = 32.3/N$. The spectral lines obtained from a DFT as well as the frequency deviation error due to discretization δ are indicated.

can be obtained from $F_0 = F_s f_0$, where F_s is the sampling frequency. The observation time of the signal is $T = N/F_s$ seconds.

The discrete frequency f_0 can be written as

$$f_0 = (L + \delta)/N = \lambda_0/N \tag{2}$$

where L and $-1/2 \leqslant \delta < 1/2$ are the integer part and the fractional part of λ_0, respectively, and λ_0 denotes the number of signal cycles contained in the observation window. Let us give an example to clarify this notation (see Fig. 1): suppose that we acquire a sinusoidal signal of frequency 32.3 Hz sampled at 128 Hz during 1 s ($N = 128$ samples); then, $f_0 = 32.3/128$. In this case $\lambda_0 = 32.3$ (the frequency cycles), $L = 32$, and $\delta = 0.3$.

With this notation, the discrete Fourier transform (DFT) of $x[n]$ at spectral line k is given by

$$X[k] = \frac{A_0}{2} e^{j\theta_0} W((k - \lambda_0)/N) + \frac{A_0}{2} e^{-j\theta_0} W((k + \lambda_0)/N) \tag{3}$$

where $W(f)$ is the Fourier transform of the selected time-domain window. For instance, considering a rectangular window of length N, we have

$$W(f) = \frac{\sin(N\pi f)}{\sin(\pi f)} e^{-j\pi f(N-1)}. \tag{4}$$

We assume from now on that the frequency of the sinusoid is far enough from the origin ($f = 0$) so that the leakage coming from the negative part of the spectrum can be neglected.

In this situation equation (3) reduces to

$$X[k] = \frac{A_0}{2} e^{j\theta_0} W((k - \lambda_0)/N). \tag{5}$$

As it is well-known, equation (3) can be efficiently calculated using an FFT algorithm and the frequency can be estimated from its largest peak

$$\hat{f_0} = \frac{\text{argmax}(|X[k]|)}{N} = L/N. \tag{6}$$

If λ_0 is not an integer (i.e. if the signal does not contain an integer number of periods within the observation interval), the results provided by equation (6) can have low accuracy because of the spacing between two adjacent samples of the DFT: this effect is denoted as short-range spectral leakage. For instance, in the example given above we would estimate the frequency as the largest spectral line at $L = 32$.

Obviously, the short-range leakage problem can be alleviated by applying zero padding to the signal before computing the DFT. However, a more efficient approach was recently proposed by Umesh and Nelson [10]. The method, which belongs to the WPA family, is based on averaging the weighted phase estimates obtained using different non-overlapping segments of $x[n]$. In particular, it proceeds as follows: first, the whole original sampled signal, $x[n]$, is used to obtain a coarse estimate of the frequency, $\hat{f_0}$, by applying equation (6). Second, the signal is divided into M non-overlapping segments of length P:

$$x_s[n] = x[n + sP], \quad 0 \leqslant n \leqslant P - 1 \tag{7}$$

and then the spectrum of each segment is evaluated at the previously estimated $\hat{f_0}$,

$$X_s(\hat{f_0}) = \sum_{n=0}^{P-1} x_s[n] e^{-j2\pi \hat{f_0}(n+sP)}. \tag{8}$$

Now, substituting equation (1) into equation (8) and taking into account that $\hat{f_0} = (f_0 - \delta)/N$, it is straightforward to show that the spectrum of the signal can be written as

$$X_s(\hat{f_0}) = \frac{A_0}{2} \left(\frac{\sin(\pi\delta P/N)}{\sin(\pi\delta/N)} \right) e^{j\theta_0} e^{-j(2\pi\delta P/N)(s+1)} \tag{9}$$

that is the amplitude of the peak is attenuated by a factor which depends on the particular window applied as well as on the frequency error δ, while its phase is shifted by a quantity which is linearly related to the frequency error

$$\arg(X_s(\hat{f_0})) = \theta_0 - \frac{2\pi\delta P}{N}(s + 1). \tag{10}$$

From equation (10), it is clear that δ can be estimated from the phase differences between consecutive segments. From a different point of view, a similar idea has been recently explored in [11].

From our experience, for most of the examples, splitting the original register into two segments ($P = N/2$) is enough to achieve accurate estimates. In this case, the frequency error can be estimated as

$$\hat{\delta} = \frac{N}{2\pi L}(\arg(X_1(\hat{f_0})) - \arg(X_2(\hat{f_0}))). \tag{11}$$

As shown in [10], when the input data consist of a single sinusoid the estimator described in this section almost attains the Cramer–Rao (CR) bound at high SNRs, which is the

lowest variance that can be attained by any unbiased estimator [12]. However, it does not perform adequately for multiple sinusoids when they are not well separated in frequency. In the following section, we describe a method to reduce the long-range spectral leakage in these situations and to extend the applicability of the WPA estimator.

2.2. LONG-RANGE LEAKAGE REDUCTION

Now, let us consider a signal consisting of p real sinusoids,

$$x[n] = \sum_{i=1}^{p} A_i \cos(2\pi f_i n + \theta_i) \tag{12}$$

in this situation, besides the short-range spectral leakage due to the granularity of the FFT, long-range leakage appears, which is concerned with the interference among the harmonics. Although the use of a proper window can reduce this interference, it does not solve the problem.

Recently, a simple iterative algorithm has been proposed in [13] to reduce the long-range spectral leakage. At each iteration the frequency, amplitude and phase of the strongest sinusoid are estimated. Then, this peak is subtracted from the original data and a new peak is estimated. A key point of the method is that each time a new frequency is estimated, all previously computed frequencies, amplitudes and phases are reestimated.

The main difficulty with this method is that it requires very accurate initial frequency estimates. Otherwise, even a small error in the frequency estimate could cause large errors in both the amplitude and phase estimates [see equation (8)]; therefore the interference caused by that sinusoid could not be effectively eliminated. In [13] this disadvantage is alleviated by applying zero padding. From a computational point of view, however, a simpler solution could be to apply the WPA method given by (11).

2.3. THE OVERALL IWPA METHOD

Finally, the proposed method is a straightforward combination of the procedures discussed in Sections 2.1 and 2.2, which is able to reduce both the short- and long-range spectral leakage. Specifically, on the first iteration of the procedure, which we denote as the IWPA, the strongest spectral line in $X(f)$ is detected and its frequency $\hat{f}_1$, amplitude $\hat{A}_1$ and phase $\hat{\theta}_1$, are estimated as follows. The frequency error deviation is corrected using two non-overlapped windows of length $P = N/2$ and applying equation (11). Once $\hat{f}_1$ has been obtained, the corresponding amplitude and phase are estimated as

$$(\hat{A}_1, \hat{\theta}_1) = \min \sum_{n=0}^{N-1} (x[n] - A_1 \cos(2\pi \hat{f}_1 n + \theta_1))^2. \tag{13}$$

Denoting $X_1 = A_1 \cos(\theta_1)$ and $Y_1 = A_1 \sin(\theta_1)$, the minimum of equation (13) is obtained by solving

$$\begin{pmatrix} -\sum_n \cos(2\pi \hat{f}_1 n)^2 & \frac{1}{2}\sum_n \sin(4\pi \hat{f}_1 n) \\ \frac{1}{2}\sum_n \sin(4\pi \hat{f}_1 n) & -\sum_n \sin(2\pi \hat{f}_1 n)^2 \end{pmatrix} \begin{pmatrix} X_1 \\ Y_1 \end{pmatrix} = \begin{pmatrix} \sum_n x[n]\cos(2\pi \hat{f}_1 n) \\ \sum_n x[n]\sin(2\pi \hat{f}_1 n) \end{pmatrix} \tag{14}$$

and then we estimate $\hat{A}_1 = \sqrt{X_1^2 + Y_1^2}$ and $\hat{\theta}_1 = \arctan(Y_1/X_1)$.

Having obtained the parameters of the first sinusoid, a residual signal is calculated by subtracting the previously obtained components

$$r[n] = x[n] - \hat{A}_1 \cos(2\pi \hat{f}_1 n + \hat{\theta}_1) \tag{15}$$

and now the algorithm obtains the parameters of the strongest sinusoid remaining in $r[n]$, $(\hat{f}_2, \hat{A}_2, \hat{\theta}_2)$. We can use these parameters to refine the estimates of the first sinusoid by

applying the same procedure to a new residual

$$r[n] = x[n] - \hat{A}_2 \cos(2\pi \hat{f}_2 n + \hat{\theta}_2). \tag{16}$$

To summarize, at each iteration we obtain a new component as the strongest peak of the remaining residual and then we reestimate the parameters of all the previously obtained components.

As we will show later, unlike other FFT-based procedures, the technique can even distinguish sinusoids with a frequency separation smaller than the periodogram's resolution limit $(1/N)$.

3. THE INTERPOLATED FFT

Let us consider again the case of single sinusoid of frequency f_0. As we have seen previously, the frequency can be written as $f_0 = (L + \delta)/N = \lambda_0/N$ [see equation (2)]. The position of the largest peak of the FFT, L, provides a coarse estimate of the frequency. Therefore, the problem of obtaining an improved frequency estimate reduces to estimate the frequency deviation $\hat{\delta}$.

In [4]. Rife proposed to estimate the frequency deviation by interpolating the magnitudes of the FFT outputs. Looking again at Fig. 1 and taking into account equation (4), we can see that the amplitude of the largest spectral line is given by

$$|X[L]| = \frac{A_0}{2} \frac{\sin(\pi\delta)}{\sin(\pi\delta/N)} \tag{17}$$

while the amplitude of the second largest spectral line (in the situation depicted in Fig. 1 this line is located at $k = L + 1$) is given by

$$|X|[L + 1]| = \frac{A_0}{2} \frac{\sin(\pi(1 - \delta))}{\sin(\pi(1 - \delta)/N)}. \tag{18}$$

In general, using a window with Fourier transform $W(f)$, the ratio of the two magnitudes can be expressed as

$$\frac{|X[L + \alpha]|}{X[L]} = \frac{|W(1 - \delta)|}{|W(\delta)|} \tag{19}$$

where α is defined by

$$|X[L + \alpha]| = \max(|X[L + 1]|, |X[L - 1]|) \tag{20}$$

i.e. α is equal to 1 if the true frequency lies to the right of the largest FFT peak and is -1 otherwise.

In the literature, approximate solutions for the inversion of equation (19) have been proposed for different windows [4, 5, 14–16]. In particular, for the rectangular window we have the following approximate solution [5]:

$$\hat{\delta} = \alpha \frac{|X[L + \alpha]|}{|X[L + \alpha]| + |X[L]|} \tag{21}$$

and for the Hanning window [7, 14]

$$\hat{\delta} = \alpha \frac{2|X[L + \alpha]| - |X[L]|}{|X[L + \alpha]| + |X[L]|}. \tag{22}$$

Similar expressions can be found in the literature for the class of Rife–Vincent windows [4, 15, 16]. Recently, some interpolation techniques have been proposed which consider three or even five FFT samples centred on the largest peak [17]. From our experience, however, in practical situations where the noise is the main cause of disturbance, the use of interpolators using three or five samples does not result in any noticeable improvement. Therefore, we limit out study to the simplest and more efficient IFFT techniques given by equations (21) and (22).

Obviously, in situations with closely spaced sinusoids in which long-range leakage is the main source of error, the IFFT method described above should be combined with the iterative procedure of Section 2.2. In fact, the complete IFFT algorithm considered in this paper is the same as that described in Section 2.3, but substituting the expression to correct the frequency (11) by equation (21) or (22) depending on the used window.

4. THE ESPRIT ALGORITHM

The ESPRIT algorithm was proposed to estimate the frequencies of a set of complex exponentials in noise and it was further developed in the context of array signal processing [3, 18]. In this section we briefly summarise its main characteristics.

Without loss of generality, let us consider a signal composed of p sinusoids corrupted by noise,

$$x[n] = \sum_{i=1}^{p} A_i \cos(2\pi f_i n + \theta_i) + r[n], \quad n = 0, \ldots, N-1 \tag{23}$$

where $r[n]$ is a white Gaussian noise of power σ^2. Assuming that the phases are random with uniform distribution, the autocorrelation of $x[n]$ is given by

$$R_x[m] = E[x[n]x[n+m]] + \sum_{i=1}^{p} \frac{A_i^2}{2} \cos(2\pi f_i m) + \sigma^2 \delta[m] \tag{24}$$

where $\delta[m] = 1$ if $m = 0$, and 0 otherwise. Using this model, the $(M \times M)$ (we assume that $2p < M < N$) data covariance matrix is given by

$$\mathbf{R} = \begin{pmatrix} R_x[0] & R_x[1] & \cdots & R_x[M-1] \\ R_x[1] & R_x[0] & \cdots & R_x[M-2] \\ \vdots & \ddots & \ddots & \vdots \\ R_x[M-1] & R_x[M-2] & \cdots & R_x[0] \end{pmatrix} \tag{25}$$

which can be rewritten as

$$\mathbf{R} = \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma^2 \mathbf{I} \tag{26}$$

where H denotes conjugate and transpose, $\mathbf{I}$ is the $(M \times M)$ identify matrix and $\mathbf{P}$ is the following $(2p \times 2p)$ diagonal matrix:

$$\mathbf{P} = \mathrm{diag}(A_1^2/2, \ldots, A_p^2/2, A_p^2/2, \ldots, A_1^2/2). \tag{27}$$

Finally, defining

$$a(f) = [\mathbf{1} \mathrm{e}^{-i2\pi f} \cdots \mathrm{e}^{-i(M-1)2\pi f}]^T (M \times 1) \tag{28}$$

the matrix $\mathbf{A}$ is given by

$$\mathbf{A} = [a(f_1) \cdots a(f_p)a(-f_p) \cdots a(f_1)](M \times 2p). \tag{29}$$

Analogous to other high-resolution algorithms, ESPRIT relies on properties of the data covariance matrix (25). Specifically, performing the singular-value decomposition of $\mathbf{R}$, we can write

$$\mathbf{R} = \mathbf{U}\Lambda\mathbf{U}^H \tag{30}$$

where Λ is a diagonal matrix with real eigenvalues ordered such that $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_M > 0$. Note that the matrix $\mathbf{APA}^H$ in equation (25) is rank-deficient: $\text{rank}(\mathbf{APA}^H) = 2p < M$. This property allows to partition the eigenvalues/eigenvectors pairs into noise eigenvectors, corresponding to eigenvalues $\lambda_{2p+1} = \cdots = \lambda_M = \sigma^2$; and signal eigenvectors corresponding to eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_{2p} > \sigma^2$. Hence, we can decompose $\mathbf{R}$ as

$$\mathbf{R} = \mathbf{U}_s\Lambda_s\mathbf{U}_s^H + \mathbf{U}_n\Lambda_n\mathbf{U}_n^H. \tag{31}$$

It can be shown that the matrix $\mathbf{U}_s$, which contains the signal eigenvectors, can be written as [3]

$$\mathbf{U}_s = \mathbf{AT} \tag{32}$$

where $\mathbf{T}$ is a full-rank matrix. This means that $\mathbf{A}$ and $\mathbf{U}_s$ span the same subspace.

Unlike other subspace-based approaches, ESPRIT exploits the special structure of matrices $\mathbf{A}$ and $\mathbf{U}_s$. In particular, $\mathbf{A}$ can be partitioned into sub-matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ as follows:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \text{last row} \end{bmatrix} = \begin{bmatrix} \text{first row} \\ \mathbf{A}_2 \end{bmatrix}. \tag{33}$$

By the structure of $\mathbf{A}$ (denoted as *shift-structure*), $\mathbf{A}_1$ and $\mathbf{A}_2$ are related by the formula

$$\mathbf{A}_2 = \mathbf{A}_1\Phi \tag{34}$$

where Φ is a diagonal matrix with elements $\mathrm{e}^{-j2\pi f_i}$, $i = 1, \ldots, p$; on the diagonal. In this way, the frequency estimation problem reduces to that of estimating Φ. Note that the *shift-structure* of $\mathbf{A}$ is also exploited in the IWPA method through equation (10).

Similarly to $\mathbf{A}$, the matrix $\mathbf{U}_s$ can be partitioned into sub-matrices $\mathbf{U}_1$ and $\mathbf{U}_2$. Now, combining equations (34) and (32) we arrive at

$$\mathbf{U}_2 = \mathbf{U}_1\Psi \tag{35}$$

where Ψ is related to Φ by

$$\Psi = \mathbf{T}^{-1}\Phi\mathbf{T} \tag{36}$$

since equation (36) is a similarity transformation, both Ψ and Φ have the same eigenvalues, from which we can obtain the estimated frequencies. Since in practice $\mathbf{U}_1$ and $\mathbf{U}_2$ are noisy estimates, the matrix Ψ is estimated in equation (35) by applying a total-least-squares (TLS) algorithm [19].

Similar to the other frequency estimation methods considered in this paper, here we assume that the number of sinusoids is known; thus, the only parameter to be selected is the order M of the matrix $\mathbf{R}$. If we want to estimate p real sinusoids the lowest value for M is $2p$, higher values for M will increase significantly the performance of the method. However, M cannot be increased too much since the computational burden grows as M^3.

Finally, the ESPRIT algorithm, as applied in this paper, can be summarised as follows:

1. Compute the eigendecomposition of the data covariance matrix of order $M\mathbf{R}$.
2. Form $\mathbf{U}_s$ by selecting the $2p$ eigenvectors corresponding to the largest eigenvalues.
3. Partition $\mathbf{U}_s$ into $\mathbf{U}_1$ and $\mathbf{U}_2$ by deleting the last row and the first row as in equation (33).
4. Estimate $\hat{\Psi}$ by solving $\mathbf{U}_2 = \mathbf{U}_1\Psi$ in a TLS sense.
5. Estimate the frequencies $\hat{f_i}$ as $-\arg(v_i)/2\pi$, where v_i, $i = 2, 4, \ldots, 2p$, are the eigenvalues of $\hat{\Psi}$.

5. SIMULATION RESULTS

In this section, we present some simulation results to compare the performance of the three frequency estimation methods considered in the paper. We generated sinusoidal signals of length $N = 256$ samples composed of two tones and added white Gaussian noise. Comparisons are made based on 1000 independent trials for each SNR, which is defined as

$$SNR = 10\log_{10}\frac{\sum_{i=1}^{p} A_i^2/2}{\sigma^2}. \tag{37}$$

The most important parameter that limits the performance of the three methods is the lowest distance between sinusoids $\min|f_i - f_j| = \Delta\lambda/N$. In order to evaluate the effect of long-range leakage, we have considered three different values of $\Delta\lambda$:

- low leakage: $\Delta\lambda = 10.7$,
- high leakage: $\Delta\lambda = 2.6$,
- very high leakage: $\Delta\lambda = 0.7$.

Specifically, for the first example we generate the following signal, $f_1 = 60.2/N, f_2 = 70.9/N$ ($\Delta\lambda = 10.7$):

$$x[n] = \cos\left(2\pi\frac{60.2}{N}n + \pi/10\right) + \cos\left(2\pi\frac{70.9}{N}n\right) + r[n].$$

We compare the performance of the following methods:

- Raw N-points FFT.
- IFFT: we apply the Rife's interpolation FFT procedure with a rectangular window [4, 5] and a Hanning window [14].
- IWPA: rectangular window; segments of length $P = N/2$.
- ESPRIT: data covariance matrix of order $M = 32$.

As a figure of merit we take the mean-squared error (MSE): for each trial we obtain two estimates $\hat{f_1}$ and $\hat{f_2}$; then, if we perform a total of N_s trials for each experiment, the MSE is defined as

$$MSE(f_i) = \frac{1}{N_s}\sum_{j=1}^{N_s}(\hat{f_{ij}} - f_i)^2 \tag{38}$$

this figure of merit takes into account both the variance and the bias of the estimator.

Figure 2 shows the MSE *vs* the SNR for the frequency estimate of $\hat{f_1}$. We see that, in a low-leakage situation (typical of vibration analysis)

- The IWPA method obtains better results than the IFFT at low and moderate SNRs. However, it gets biased at high SNRs.

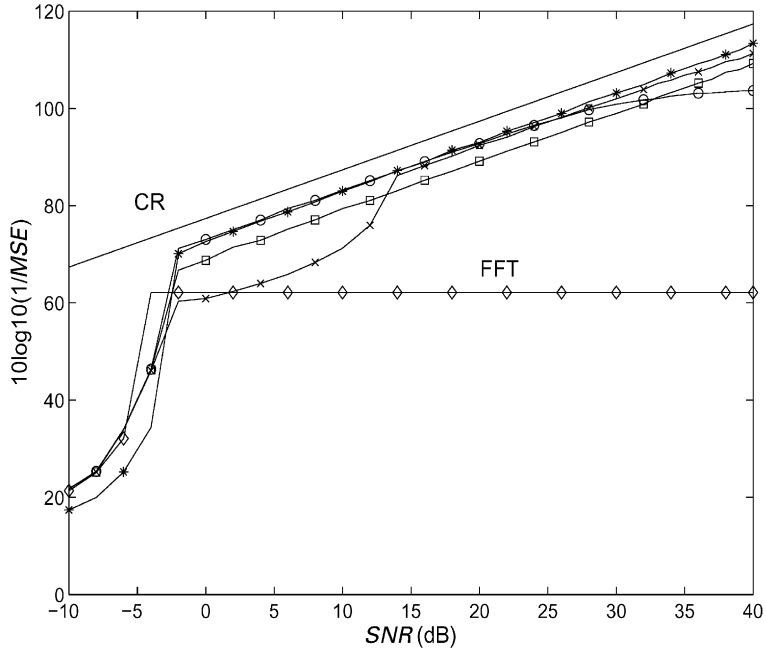

Figure 2. Mean-square error of $\hat{f}_1$ *vs* SNR in a situation with a low level of spectral interference (Example 1: $f_1 = 60.2/256$, $f_2 = 70.9/256$): ○—○—○ IWPA; ∗— ∗— ∗ ESPRIT; ×—×—× IFFT (rect.); □—□—□ IFFT (Hanning).

- The use of a suitable window (Hanning, for instance) extends the threshold SNR of IFFT significantly.
- The best results are provided by ESPRIT.

For the second experiment, we change f_2 from $70.9/N$ to $62.8/N$; in this way, the frequency separation is reduced to $\Delta f = 2.6/N$. In this situation, with a high level of interference between harmonics, the IWPA clearly outperforms the IFFT methods; moreover, the IWPA is competitive with ESPRIT. At high SNRs, both IWPA and IFFT tend to become biased (see Fig. 3).

Finally, for the last example we move f_2 to $60.9/N$. In this scenario, the frequency separation is $\Delta\lambda = 0.7$, which is below the periodogram's resolution limit. Therefore, conventional methods based on the FFT, would fail to resolve the two sinusoids. However, in Fig. 4 we see that the IWPA method still obtains accurate estimates. We can understand this fact, by describing how the algorithm proceeds in a typical example: the IWPA method locates the first peak at $\hat{f} = 60/N$. Then, although we are in a situation with a very high level of spectral interference, the IWPA method is able to correct the frequency giving an estimate closer to $60.2/N$. The iterative procedure of Section 2.2 subtracts the effect of that component and a new spectral line is now detected at $\hat{f} = 61/N$ and corrected towards $60.9/N$. From Fig. 4 we see that, at low SNRs, the IWPA method clearly outperforms ESPRIT, but again it becomes biased at high SNRs. Unlike the IWPA, the IFFT method is not able to correct the frequency estimates in a situation with such a high level of leakage and thus the frequencies cannot be resolved.

In vibration analysis, where real-time measurements are often necessary, the computational cost is a critical factor. Table 1 gives an idea of the computational requirements

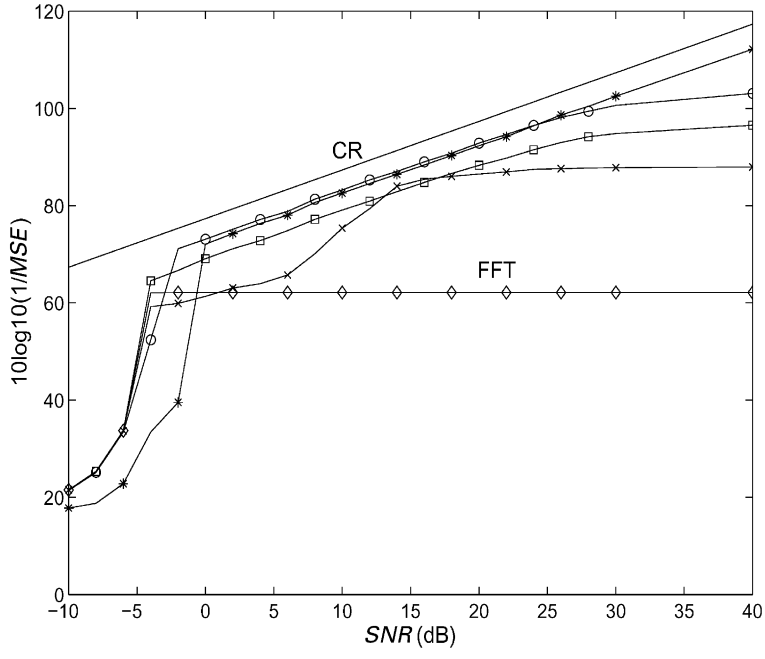

Figure 3. Mean-square error of $\hat{f}_1$ *vs* SNR in a situation with a high level of spectral interference (Example 2: $f_1 = 60.2/256$, $f_2 = 62.8/256$): ○—○—○ IWPA; ∗— ∗— ∗ ESPRIT; ×—×—× IFFT (rect.); □—□—□ IFFT (Hanning).

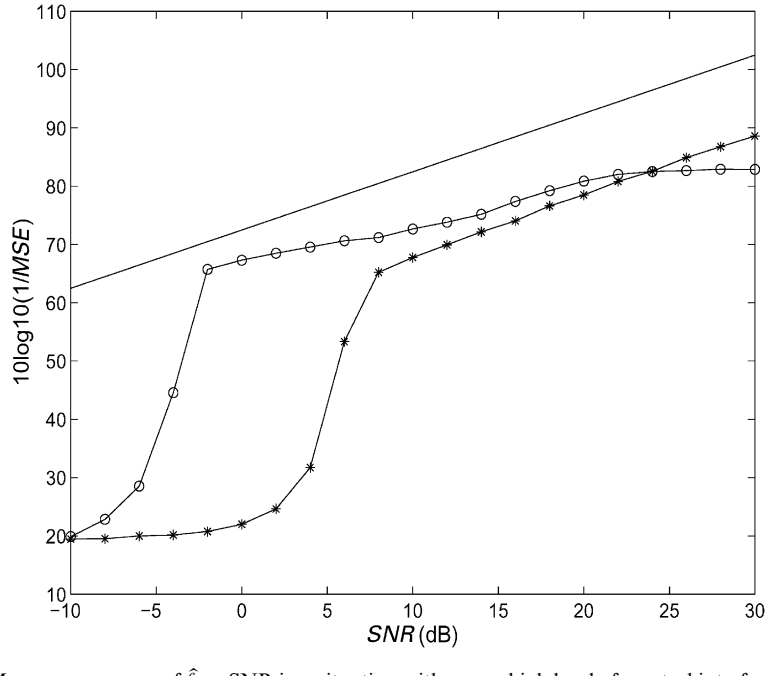

Figure 4. Mean-square error of $\hat{f}_1$ *vs* SNR in a situation with a very high level of spectral interference (Example 3: $f_1 = 60.2/256, f_2 = 60.9/256$). The figure only shows the performance of the IWPA method and ESPRIT, since the IFFT method fails to resolve the two sinusoids: ——— Cramer–Rao; ○—○—○ IWPA; ∗— ∗— ∗ ESPRIT.

TABLE 1

Computational cost of each method in floating point operations (Flops)

N	FFT	IFFT	IWPA ($P = N/2$)	ESPRIT ($M = N/4$)
128	3.4e3	3.8e3	1.2e4	5.1e6
256	7.3e3	8.1e3	1.7e4	4.0e7

associated with each method for different register lengths. The three methods were programmed using Matlab [20] and the computational cost was measured as the number of floating point operations (Flops). The computational cost of IWPA is three times that of IFFT. Finally, the high computational requirements of high-resolution methods such as ESPRIT explain why these methods are not widely used in vibration analysis.

6. A REAL EXAMPLE: VIBRATION ANALYSIS OF HYDROELECTRIC TG-SETS

The reliability of rotating machines such as a hydroelectric turbogenerator is critical to the overall reliability and operation of an electrical power plant. It is becoming more and more important to receive an early warning of any problem before failure and long outage occurs. The analysis of vibration signals is the most popular monitoring tool for its capability to detect most of the mechanical-related and hydraulic malfunctions.

In this section, we apply the frequency estimation methods to real end winding vibration data acquired from a hydroelectric generator set located in Villarino (Spain). The set is a reversible Francis turbine with 135 MW of nominal power. The overall monitoring system measures the shaft's displacement at three guide-bearing levels using a couple of orthogonal sensors at each level (six channels), the vibration of the bearing support (six channels) and the end winding vibration (three channels). Here, we only consider the end winding vibration channels.

The winding vibration signals are acquired using optoelectronic accelerometers, a sensor is set in each electric phase in the cups that have the highest tension and are most distant among them, in order to pick up the tangential vibration induced in the cups when the rotor is turning. The optical vibration signal are preprocessed and conditioned using an optoelectronic unit, which includes the optical source (laser diode), the detection unit (photodetectors and preamplifiers), analog filters with selectable gains and the power supply. Then, the electrical signals are sampled (analog-to-digital conversion) and stored in a personal computer for further processing. A complete description of the system can be found in [21, 22].

The stator of the monitored turbine has three phase bornes and three neutral and its windings are connected in star. On the other hand, the rotor has 10 poles and it turns at 600 rpm; the excitation is made by means of collector rings and brushes. Typically, a winding signal is composed of a number of harmonics of the electric network frequency. In particular, for the turbine monitored in this example the electric network frequency is $f_n = 50$ Hz and the winding vibration signal has a main harmonic at $2f_n$, and subharmonics at f_s, $3f_n$, $4f_s$, etc. (see Fig. 6).

An increase in the overall vibration level of these signals may anticipate malfunctions such as a loosening of the stator wedges, the axial core bolts or crore's radial junctions. Besides the overall vibration level, of the most valuable spectral information is the location of the largest peaks as well as their amplitudes and phases. This information can be used for

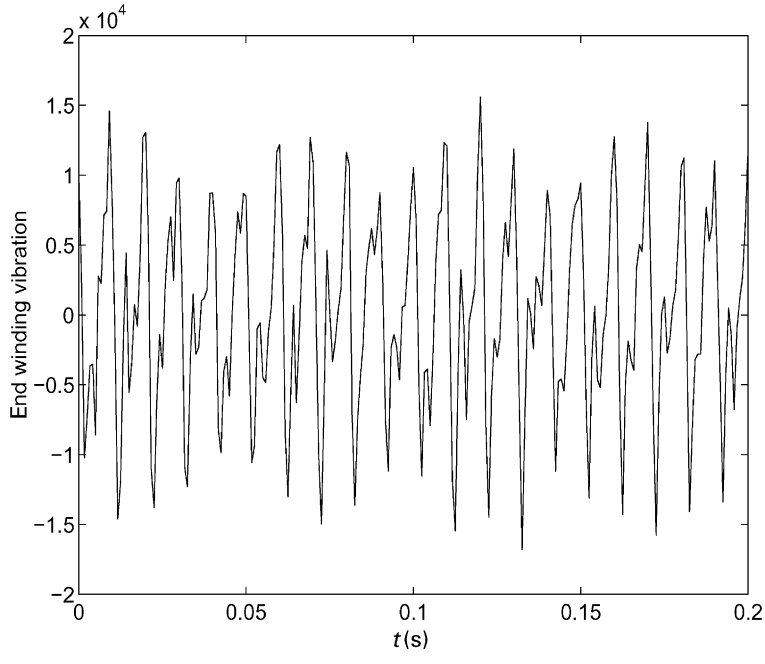

Figure 5. End winding vibration.

data trend analysis, i.e. the evolution of these spectral parameters can indicate the development of a failure. Therefore, it is a key step to have accurate frequency estimates.

The first goal of this section is to show that by applying any of the procedures described in this paper it is possible to obtain accurate frequency estimates using temporal sequences shorter than those required when just an FFT (probably with zero padding) is used. The second goal is to perform a comparison among the presented methods.

As an example, Fig. 5 shows a winding vibration register of $N = 256$ samples and Fig. 6 shows its spectrum obtained using a 512-FFT (we applied zero padding); the sampling frequency is $F_s = 1200$ Hz. Considering that the register is composed of $p = 6$ sinusoids, we compare the performance of the following methods:

- Using the whole temporal sequence ($N = 256$), we apply zero padding to perform a 512-FFT and then we select the six largest peaks.
- Using a shorter temporal sequence ($N = 64$), we apply the IWPA method with $P = 32$.
- Using a shorter temporal sequence ($N = 64$), we apply the IFFT with a Hanning window.
- Using a shorter temporal sequence ($N = 64$), we apply the ESPRIT method with $M = 32$.

From the results of Table 2, we can conclude that if we use any of the proposed high-accuracy methods it is possible to acquire a shorter sequence. Obviously, this may render some computational and memory savings. In vibration analysis, for what concerns frequency estimation, in general the temporal length of the acquired sequences is overestimated. The frequency estimates provided by the three method are even better than those of a 512-FFT. For instance, unlike the 512-FFT, they are able to detect the small subharmonic located near 400 Hz.

Among the three proposed methods the best compromise between performance and computational cost is provided by the IWPA: it is the method selected for this application.

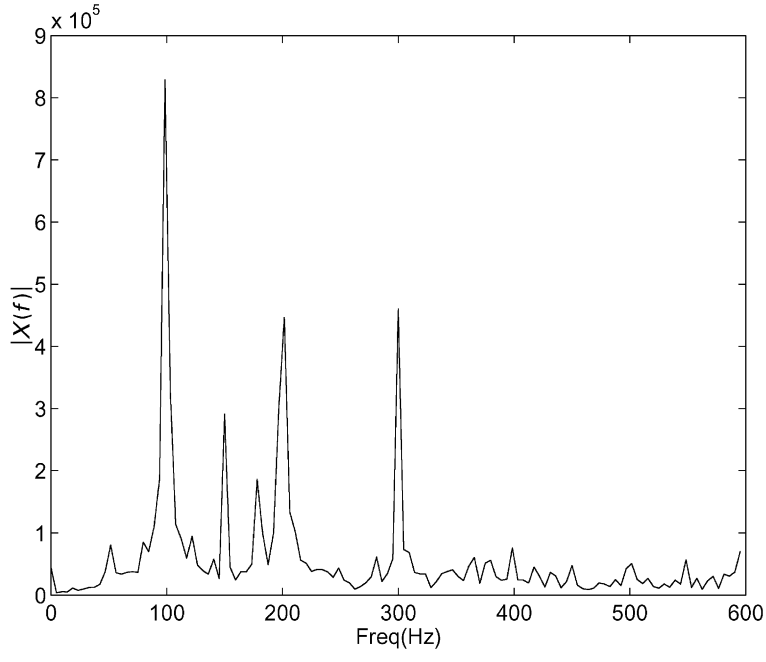

Figure 6. Spectrum of the end winding vibration.

TABLE 2

Frequencies estimated from the end winding vibration data

	FFT $N = 256$ + zero padding	IFFT $N = 64$	IWPA $N = 64$, $P = 32$	ESPRIT $N = 64$, $M = 32$
f_1	98.4	100.1	100.4	99.9
f_2	300.0	199.4	299.8	197.0
f_3	201.6	302.4	150.3	297.8
f_4	150.0	156.3	200.0	150.1
f_5	178.2	67.0	178.7	384.8
f_6	107.8	383.7	399.5	0

7. CONCLUSIONS

In this paper, three different high-accuracy frequency estimation methods: IWPA, IFFT and ESPRIT are compared. Only the IFFT is widely known and applied in vibration analysis problems. Using simulation examples as well as real data from a hydroelectric turbogenerator, we have found that when the harmonics are well separated in frequency and there is a low level of spectral interference, all the methods have similar performance. In this situation, the IFFT technique should be applied since it has the lowest computational cost. On the other hand, when the level of spectral interference is moderate or high, the IWPA outperforms the IFFT (with a moderate increase in computational cost). Moreover, the IWPA has the ability to separate sinusoids more closely spaced than the periodogram's resolution limit. At low SNRs it provides better results than a high-resolution method such as ESPRIT. However, it tends to get biased at high SNRs. Finally, despite its good

statistical properties and high-resolution capability, the high computational cost of ESPRIT advises against its use in real-time vibration analysis applications.

ACKNOWLEDGEMENTS

This work has been partially financed by BRITE-EURAM project 7289 and by CICYT grant TIC96-0500-C10-07.

REFERENCES

1. P. STOICA and R. L. MOSES 1997 *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall.
2. S. M. KAY 1988 *Modern Spectral Estimation, Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall.
3. R. ROY and T. KAILATH 1989 *IEEE Transactions on Acoustics, Speech and Signal Processing* **37**, 984–995. ESPRIT—estimation of signal parameters via rotational invariance techniques.
4. D. C. RIFE and G. A. VINCENT 1970 *Bell System Technical Journal* **49**, 197–228. Use of the discrete Fourier transform in the measurement of levels and tones.
5. V. K. JAIN, W. L. COLLINS and D. C. DAVIS 1979 *IEEE Transactions on Instrumentation and Measurement* **28**, 113–122. High-accuracy analog measurements via Interpolated FFT.
6. C. OFFELLI and D. PETRI 1990 *IEEE Transactions on Instrumentation and Measurement* **39**, 106–111. Interpolation techniques for real-time multifrequency waveform analysis.
7. X. MING and D. KANG 1996 *Mechanical Systems and Signal Processing* **10**, 211–221. Corrections for frequency, amplitude and phase in a fast Fourier transform of a harmonic signal.
8. S. M. KAY 1989 *IEEE Transactions on Acoustics, Speech and Signal Processing* **37**, 1987–1990. A fast and accurate single frequency estimator.
9. S. A. TRETTER 1985 *IEEE Transactions on Information Theory* **31**, 832–835. Estimating the frequency of a noisy sinusoid by linear regression.
10. S. UMESH and D. NELSON 1996 *Proceedings of the 1996 IEEE International Conference on Acoustics Speech and Signal Processing*, Atlanta (GA), 2797–2800. Computationally efficient estimation of sinusoidal frequency at low SNR.
11. J. M. DANTHEZ and C. GIROUSSENS *Mechanical Systems and Signal Processing* **12**, 753–768. The autocorrelation spectrum: a useful spectral estimator for vibration analysis of rotating machinery. Accurate estimation and cancellation of pure tones.
12. D. C. RIFE and R. R. BOORSTYN 1976 *Bell System Technical Journal* **55**, 1389–1410. Multiple tone parameter estimation for discrete-time observations.
13. P. T. GOUGH 1994 *IEEE Transactions on Signal Processing* **42**, 1317–1322. A fast spectral estimation algorithm based on the FFT.
14. T. GRANDKE 1983 *IEEE Transactions on Instrumentation and Measurement* **32**, 350–355. Interpolation algorithms for discrete Fourier transform of weighted signals.
15. G. ANDRIA, M. SAVINO and A. TROTTA 1989 *IEEE Transactions on Instrumentation and Measurement* **38**, 856–863. Windows and interpolation algorithms to improve electrical measurement accuracy.
16. C. OFFELLI and D. PETRI 1990 *IEEE Transactions on Instrumentation and Measurement* **39**, 106–111. Interpolation techniques for real-time multifrequency waveform analysis.
17. M. D. MACLEOD 1998 *IEEE Transactions on Signal Processing* **46**, 141–148. Fast nearly ML estimation of the parameters of real or complex single tones or resolved multiple tones.
18. R. ROY, A. PAULRAJ and T. KAILATH 1996 *IEEE Transactions on Acoustics, Speech and Signal Processing* **34**, 1340–1342 . ESPRIT—a subspace rotation approach to estimation of parameters of cisoids in noise.
19. G. H. GOLUB and C. F. VANLOAN 1989 *Matrix Computations*. Baltimore, MD: Johns Hopkins University Press.
20. The Mathworks, Inc. 1998 *Using Matlab*. Natick, MA: The Mathworks, Inc.
21. F. ARREGUI et al. 1996 *EUROMAINTENANCE 96*, Copenhagen, 741–750. Protection and monitoring system for hydroelectric generating sets.
22. J. IBAÑEZ, C. PANTALEÓN, I. SANTAMARÍA, D. LUENGO and M. LÁZARO 1998 *International Conference on Signal Processing Application and Technology*, Toronto, Canada, 932–936. Specific DSP based monitoring system for hydro-generator sets.

APPENDIX: NOMENCLATURE

$x[n]$	sampled sequence
F_s	sampling frequency
f_0	discrete frequency
F_0	analog frequency
N	number of samples of $x[n]$
P	length of the segments in which the signal is split in the IWPA method
$X(f)$ $W(f)$	Fourier transform of the signal and the window, respectively
$X[k]$	discrete Fourier transform (DFT), kth spectral line
L	location of the largest spectral line
δ_0	frequency error deviation
λ_0	frequency in cycles (number of periods within the observation interval)
$R_x[m]$	autocorrelation of $x[n]$
$\mathbf{R}$	autocorrelation matrix
M	dimension of $\mathbf{R}$ ($M \times M$)
σ^2	noise variance
$\mathbf{I}$	identity matrix of dimension ($M \times M$)
λ_i	eigenvalues of $\mathbf{R}$
p	number of sinusoids (harmonics)
$\mathbf{A}$	matrix whose columns are the steering vectors for each frequency ($2p \times M$)
$\mathbf{U}_s$, $\mathbf{U}_n$	matrices with the signal and noise eigenvectors, respectively
$\mathbf{U}_1$, $\mathbf{U}_2$	submatrices formed by deleting the first and last rows of $\mathbf{U}_s$, respectively
$\mathbf{A}_1$, $\mathbf{A}_2$	submatrices formed by deleting the first and last rows of $\mathbf{A}$, respectively

AUTHOR QUERY FORM

HARCOURT PUBLISHERS

JOURNAL TITLE: MSSP DATE: 12/6/2000
ARTICLE NO. : 20001321

Queries and/or remarks

Manuscript Page/line	Details required	Author's response
19 20	Ref. [1] - Pl. provide year Ref. [21] - Pl. provide all authors names	