

DFT-based frequency estimation under harmonic interference

Aníbal Ferreira and Ricardo Sousa

Abstract—In this paper we address the accurate estimation of the frequency of sinusoids of natural signals such as singing, voice or music. These signals are intrinsically harmonic and are normally contaminated by noise. Taking the Cramér-Rao Lower Bound for unbiased frequency estimators as a reference, we compare the performance of several DFT-based frequency estimators that are non-iterative and that use the rectangular window or the Hanning window. Tests conditions simulate harmonic interference and two new ArcTan-based frequency estimators are also included in the tests. Conclusions are presented on the relative performance of the different frequency estimators as a function of the SNR.

I. INTRODUCTION

Many signal processing problems require the estimation of the frequency, magnitude and phase parameters of sinusoids, particularly those involving speech or audio coding, PCM to MIDI transcription, and real-time accurate singing analysis [1]. The magnitude and phase estimation typically depends on frequency estimation and therefore, in this paper we focus exclusively on the frequency parameter.

The spectrum of speech or singing typically consists of several sinusoids that very approximately follow a harmonic organization [2]. The estimation of the exact frequency of a single sinusoid when other interfering signals are also present, including noise or other sinusoids, is frequently implemented by taking the Discrete Fourier Transform (DFT) of a discrete signal and by extracting information from the DFT spectrum. As the DFT spectrum is intrinsically discrete in frequency and is limited by the natural frequency resolution of the DFT ($2\pi/N$, where N is the size of the DFT), the accurate estimation of the frequency of a sinusoid involves interpolation using several samples of the DFT spectrum (or DFT bins).

Many DFT interpolation algorithms have been proposed in the literature for more than 40 years [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14]. Test conditions may however differ significantly which makes the evaluation of the relative performance quite difficult. For example, the signal may be weighted by different window functions before the DFT which has a strong impact on the frequency selectivity of the DFT and on the ‘leakage effect’ [15]. The test sinusoid may be a complex sinusoid (*i.e.*, a cisoid) or

This work was supported by the Portuguese Foundation for Science and Technology, an agency of the Portuguese Ministry for Science, Technology and Higher Education, under grant SFRH/BD/24811/2005

A. Ferreira is with the Department of Electrical and Computer Engineering, University of Porto - Faculty of Engineering, Porto, Portugal, e-mail: ajf@fe.up.pt

R. Sousa is with the Department of Electrical and Computer Engineering, University of Porto - Faculty of Engineering, Porto, Portugal, e-mail: sousa.ricardo@fe.up.pt

a real sinusoid. The DFT interpolation procedure may be iterative or non-iterative. A more realistic test scenario in the case of singing signals for example, must involve several interfering sinusoids because of the harmonic nature of these signals.

Since we are interested in the real-time and accurate analysis of each individual sinusoid in the harmonic structure of singing, which may easily contain 100 or more sinusoids, we focus on frequency estimation algorithms that:

- avoid iterative procedures and are computationally simple,
- avoid the computation of a DFT larger than the strict length of the data vector,
- maximize the estimation accuracy and robustness to noise when not only noise but also other interfering sinusoids are present.

In order to assess performance, we take as a reference the Cramér-Rao lower bound (CRLB) for the variance of the estimation error obtained with an unbiased maximum likelihood (ML) estimator [7], [8], [9], [11].

This paper is structured as follows. In section II we state the estimation problem, identify the CRLB and specify the test conditions. In section III we define and clarify the relation between the different windows used in our tests. In sections IV and V we present the reference non-iterative frequency estimators evaluated in our study and using the rectangular or the Hanning window, respectively. In section VI we present two new ArcTan estimators that are described in a separate paper [16] and that are also included in our comparative study. Section VII discusses the relative performance of the different frequency estimators and section VIII concludes this paper.

II. THE ESTIMATION PROBLEM

For simplicity but without loss of generality, we consider that $x(n)$ represents a sinusoidal signal of arbitrary frequency ω_ℓ and that is corrupted by additive white Gaussian noise, $r(n)$. The frequency of the sinusoid is given by $\omega_\ell = \frac{2\pi}{N}(\ell + \Delta\ell)$, where ℓ and $\Delta\ell$ represent respectively the integer part ($0 < \ell < N/2$) and the fractional part ($0.0 \leq \Delta\ell < 1.0$ or, depending on the interpolation rule, $-0.5 \leq \Delta\ell < 0.5$) on the DFT bin scale. In the case of a real sinusoid

$$x(n) = A \sin [n\omega_\ell + \phi] + r(n), \quad (1)$$

where A represents the magnitude of the sinusoid, and ϕ represents the phase of the sinusoid.

The estimation can be formulated in a simple way: given the input signal according to (1), find the values of ℓ and $\Delta\ell$ after the signal has been multiplied by an analysis window,

$h(n)$, and transformed to the frequency domain using an N -point DFT

$$V(k) = \sum_{n=0}^{N-1} h(n)x(n) e^{-j\frac{2\pi}{N}kn}. \quad (2)$$

The main objective of the estimation problem is therefore to find an algorithm or formula that uses the values of $V(k)$ and interpolates the value of $\Delta\ell$ since ℓ can be easily identified as it corresponds to a local maximum of $|V(k)|$.

The CRLB for an unbiased ML frequency estimator, in the case of a complex sinusoid, is given by [9]

$$\text{var}\{\hat{\omega}_\ell\} \geq \frac{6\sigma^2}{A^2N(N^2-1)}, \quad (3)$$

where A represents the magnitude of the sinusoid, N represents the length of the DFT and σ^2 represents the variance of the noise which we assume to be zero-mean, white, complex, and Gaussian. In order to simplify the relative evaluation of the performance of different frequency interpolators, we will normalize $\text{var}\{\hat{\omega}_\ell\}$ by the natural frequency resolution of the DFT ($2\pi/N$), take its square root and express the result as a percentage of the normalized bin width:

$$\text{RMSE} = \frac{\sqrt{\text{var}\{\hat{\omega}_\ell\}}}{2\pi/N} \geq \sqrt{\frac{3\sigma^2}{2A^2\pi^2N(1-1/N^2)}}. \quad (4)$$

In our simulations we have considered $A=1$, $N=512$, $\phi=\pi/6$, and $\ell=20$. We made $\Delta\ell$ to vary from 0.0 till 1.0 in steps of 0.01 and, for each case, we have generated 100 instances of the complex noise vector $r(n)$ according to a desired target $\text{SNR} = 10 \log_{10}(A^2/\sigma^2)$. By adjusting the variance of the noise, the RMSE due to each frequency estimator has been simulated for SNR values between -10 dB and 50 dB in steps of 2.5 dB.

A more realist test scenario in singing must however include an interfering sinusoid on each side of the (real, target) sinusoid whose frequency is estimated. We have therefore expanded (1) to a more suitable form:

$$\begin{aligned} x(n) &= A \sin[n\omega_\ell + \phi] + \\ &A \sin\left[\frac{2\pi}{N}9.7n + \frac{\pi}{4}\right] + \\ &A \sin\left[\frac{2\pi}{N}31.3n + \frac{\pi}{3}\right] + r(n). \end{aligned} \quad (5)$$

As in our simulations $A=1$, $\phi=\pi/6$, $\ell=20$, and $\Delta\ell$ varies in the range $[0.0, 1.0[$ (or, for some estimators, $\ell=20$ or $\ell=21$ and $\Delta\ell$ varies in the range $[-0.5, 0.5]$). This means that the separation between the target sinusoid and the two adjacent sinusoids varies between 10.3 bins and 11.3 bins.

The normalized CRLB (4) has thus to be scaled by $\sqrt{2/3}$ according to the conditions of the new test scenario which we will assume from now on.

III. WINDOWS AND THEIR IMPLICATIONS

Three windows of decreasing selectivity and increasing main-side lobe attenuation are considered in this paper: the

rectangular window

$$h_R(n) = 1, \quad (6)$$

$$n = 0, 1, \dots, N-1,$$

the sine window

$$h_S(n) = \sqrt{h_H(n)} = \sin\frac{\pi}{N}(n+0.5), \quad (7)$$

$$n = 0, \dots, N-1,$$

and the Hanning window

$$h_H(n) = \frac{1}{2} \left[1 - \cos\frac{2\pi}{N}(n+0.5) \right], \quad (8)$$

$$n = 0, \dots, N-1.$$

The main lobe width of the rectangular, sine and Hanning window is $4\pi/N$, $6\pi/N$, and $8\pi/N$, respectively. The narrower the main lobe the better the selectivity because closely-spaced sinusoids can be better resolved. In this perspective, the rectangular window has the best selectivity and the Hanning window has the poorest selectivity. On the other hand, the larger the main lobe, the better the attenuation between the main lobe and the side lobes, a feature also referred to as ‘leakage’ [15]. In this perspective, the rectangular window has the largest leakage and the Hanning window has the smallest leakage. The lower the leakage, the lower the mutual influence between two resolved sinusoids in the DFT spectrum. These are important aspects that are likely to influence the performance of the estimation process when noise and other interfering sinusoids are present in the signal.

In order to avoid leakage as much as possible, it is therefore appropriate to interpolate the value of $\Delta\ell$ using the two, three or four largest DFT spectral lines around a spectral peak, when the rectangular, sine or Hanning window is used, respectively.

IV. RECTANGULAR WINDOW-BASED ESTIMATORS

Four non-iterative, DFT-based frequency estimators that presume the rectangular window have been selected in our evaluation based on the reported simplicity and performance [4], [8], [9], [13], and based on our preliminary simulation results.

The Jain *et al.* frequency estimator [4] recognizes that if $\omega_\ell = \frac{2\pi}{N}(\ell + \Delta\ell)$, then the two spectral lines $|V(\ell)|$ and $|V(\ell+1)|$ will be the largest associated with the local peak at ω_ℓ . By neglecting the leakage due to the spectral peak on the negative frequency axis (and assuming that $\ell > 20$), frequency estimation is achieved using

$$\hat{\Delta\ell} = \frac{|V(\ell+1)|}{|V(\ell)| + |V(\ell+1)|}, \quad (9)$$

$$0 \leq \hat{\Delta\ell} < 1.$$

Jain *et al.* have suggested that the loss in performance due to ‘harmonic interference’ is rather small [4].

Quinn has proposed a frequency estimator that uses the DFT spectral lines on each side of a local maximum (at $k = \ell$) in order to improve robustness to noise. Taking $\alpha_L =$

$\frac{-\alpha}{\alpha-1}$ where $\alpha = \Re \left\{ \frac{V(\ell-1)}{V(\ell)} \right\}$, and $\alpha_R = \frac{\alpha}{\alpha-1}$ where $\alpha = \Re \left\{ \frac{V(\ell+1)}{V(\ell)} \right\}$, frequency estimation is obtained as

$$\hat{\Delta\ell} = \frac{\alpha_L + \alpha_R}{2} - \Upsilon(\alpha_L^2) + \Upsilon(\alpha_R^2), \quad (10)$$

$$-\frac{1}{2} \leq \hat{\Delta\ell} < \frac{1}{2},$$

where

$$\Upsilon(x) = \frac{1}{4} \log(3x^2 + 6x + 1) - \frac{\sqrt{6}}{24} \log \left(\frac{1+x - \sqrt{\frac{2}{3}}}{1+x + \sqrt{\frac{2}{3}}} \right).$$

Quinn has shown that the asymptotic variance of this estimator in the case of a single real sinusoid is less than $\pi^2/6 \sim 1.65$ times that of the CRLB.

Macleod has developed a three-sample frequency interpolator that also involves a peak sample in the DFT spectrum and its two neighbors [9]. By considering that in order to improve performance it is necessary to use DFT phase as well as magnitude information in the frequency estimator, $\alpha = \Re\{V(\ell)V^*(\ell)\}$, $\alpha_L = \Re\{V(\ell-1)V^*(\ell)\}$ and $\alpha_R = \Re\{V(\ell+1)V^*(\ell)\}$ are first computed, which lead to

$$\Upsilon = \frac{\alpha_L - \alpha_R}{2\alpha + \alpha_L + \alpha_R},$$

and finally to

$$\hat{\Delta\ell} = \frac{\sqrt{1+8\Upsilon^2} - 1}{4\Upsilon}, \quad (11)$$

$$-\frac{1}{2} \leq \hat{\Delta\ell} < \frac{1}{2}.$$

It is reported that the associated average variance (in the case of a complex sinusoid) is about 1.32 times that of the CRLB.

Jacobsen *et al.* [13] have suggested recently very simple and yet efficient DFT frequency estimators using the rectangular window. As in the previous two cases, three DFT samples centered on a spectral peak are used to provide a frequency estimate according to

$$\hat{\Delta\ell} = \Re \left\{ \frac{V(\ell+1) - V(\ell-1)}{2V(\ell) - V(\ell+1) - V(\ell-1)} \right\}, \quad (12)$$

$$-\frac{1}{2} \leq \hat{\Delta\ell} < \frac{1}{2}.$$

The authors indicate that this simple estimator is surprisingly accurate even for very low SNR which is in part explained by the fact that it has an intrinsic ability to cancel statistical bias.

V. HANNING WINDOW-BASED ESTIMATORS

We have selected five non-iterative, DFT-based frequency estimators that presume the Hanning window, based on the reported simplicity and performance [5], [9], [17], [12], [13], and considering a few preliminary simulation results.

Grandke recognizes that leakage due to ‘harmonic interference’ is a problem with rectangular window-based frequency interpolators and has suggested a frequency estimator to be

used with the Hanning window (providing almost no ‘long-range’ leakage [5]):

$$\hat{\Delta\ell} = \frac{2|V(\ell+1)| - |V(\ell)|}{|V(\ell+1)| + |V(\ell)|}, \quad (13)$$

$$0 \leq \hat{\Delta\ell} < 1.$$

Grandke shows that this DFT frequency estimator (ou ‘interpolator’) provides much more accurate results than Jain’s [4] when the same test conditions are used (and involving three real sinusoids separated by 20 DFT bins). Regarding harmonic interference, Grandke also anticipates that more ‘sophisticated windows’ than the Hanning window may circumvent existing restrictions due to fact that ‘tones have to be sufficiently spaced’.

In the addition to a frequency estimator using the rectangular window, as described in the previous section, Macleod has also proposed a frequency estimator using the Hanning window and that has ‘intrinsic leakage rejection’ [9]. By computing first $\alpha = \Re\{V(\ell)V^*(\ell)\}$, $\alpha_L = \Re\{V(\ell-1)V^*(\ell)\}$ and $\alpha_R = \Re\{V(\ell+1)V^*(\ell)\}$, frequency is estimated using

$$\hat{\Delta\ell} = 2 \frac{\alpha_L - \alpha_R}{2\alpha - \alpha_L - \alpha_R}, \quad (14)$$

$$-\frac{1}{2} \leq \hat{\Delta\ell} < \frac{1}{2}.$$

The average variance of this estimator has been evaluated as about 2.13 times that of the CRLB [9] when the estimation of a single complex sinusoid is considered.

A very simple, popular and frequently used DFT frequency estimator is the parabolic interpolator [17]. The parabolic interpolator involves a parabola fitting the main lobe of the magnitude of the frequency response of the analysis window on the logarithmic scale. As the top of the main lobe has a convex shape, a convenient model on a X-Y plane is given by $y = y_0 - m(x - x_0)^2$, where the unknown parameters are the horizontal shift (x_0), the vertical shift (y_0), and the convexity factor (m). Thus, three equations are required. Let us admit that $y_0 = A_{dB}$ represents the unknown magnitude of the sinusoid, and that $x_0 = \ell + \Delta\ell$ of which only $\Delta\ell$ is unknown. Using $A_{dB}(k) = 20 \log_{10} |V(k)|$, three equations are readily obtained as $A_{dB}(\ell-1) = A_{dB} - m(-\Delta\ell-1)^2$, $A_{dB}(\ell) = A_{dB} - m(-\Delta\ell)^2$, and $A_{dB}(\ell+1) = A_{dB} - m(1-\Delta\ell)^2$. Solving for $\Delta\ell$ one obtains

$$\hat{\Delta\ell} = \frac{A_{dB}(\ell+1) - A_{dB}(\ell-1)}{4A_{dB}(\ell) - 2A_{dB}(\ell+1) - 2A_{dB}(\ell-1)}, \quad (15)$$

$$-\frac{1}{2} \leq \hat{\Delta\ell} < \frac{1}{2}.$$

The parabolic interpolator requires that for all possible values of $\Delta\ell$, three DFT lines fit inside the main lobe of the frequency response of the analysis window. Therefore, this implies that the parabolic interpolator is not valid if used with the rectangular or the sine window. When the parabolic interpolator is used with the Hanning window (8), the maximum absolute estimation error has been found to be in the order of 1.6% of the bin width, which is relatively poor [16], probably due to the fact that the parabolic interpolator is window agnostic. Reliable estimation requires that the

frequency separation between two sinusoids is at least 4 DFT frequency lines (or $8\pi/N$) [16]. The performance of quadratic interpolator may however be improved if a preliminary stage of frequency interpolation is used such as zero-padding [17].

Similarly to the frequency estimator proposed by Quinn as discussed in section IV, Quinn has also proposed recently a new estimator to be used with the Hanning window [12]. Defining $\alpha_L = \frac{2\alpha+1}{1-\alpha}$ where $\alpha = \Re \left\{ \frac{V(\ell-1)}{V(\ell)} \right\}$, and $\alpha_R = \frac{2\alpha+1}{\alpha-1}$ where $\alpha = \Re \left\{ \frac{V(\ell+1)}{V(\ell)} \right\}$, frequency is estimated as

$$\hat{\Delta\ell} = \frac{\alpha_L + \alpha_R}{2} + \Upsilon(\alpha_L^2) - \Upsilon(\alpha_R^2), \quad (16)$$

$$-\frac{1}{2} \leq \hat{\Delta\ell} < \frac{1}{2},$$

where

$$\Upsilon(x) = -\frac{5}{14} \log(35x^2 + 120x + 32) - \frac{\sqrt{155}}{140} \log \left(\frac{12 + 7x - 4\sqrt{\frac{31}{5}}}{12 + 7x + 4\sqrt{\frac{31}{5}}} \right).$$

The reported performance results for a real sinusoid are not too informative as they were obtained for a specific SNR [12].

In addition to the frequency interpolator by Jacobsen *et al.* [13] and described in section IV, the authors also describe in the same paper two frequency interpolators to be used with the Hanning window, including:

$$\hat{\Delta\ell} = 1.36 \frac{|V(\ell+1)| - |V(\ell-1)|}{|V(\ell)| + |V(\ell+1)| + |V(\ell-1)|}, \quad (17)$$

$$-\frac{1}{2} \leq \hat{\Delta\ell} < \frac{1}{2}.$$

Performance results are however presented for a single tone and in a rather small range of SNR (-2 dB till 10 dB).

VI. NEW ARCTAN INTERPOLATORS

In a companion paper [16] we derive two ArcTan-based DFT frequency estimators that are matched to the rectangular window (6) and to the sine window (7). The first estimator makes use of $|V(\ell)|$ and $|V(\ell+1)|$, the two largest DFT spectral lines because at most two DFT bins fit within the main lobe of the frequency response of the rectangular window:

$$\hat{\Delta\ell} \approx \frac{N}{\pi} \arctan \frac{\sin \frac{\pi}{N}}{\cos \frac{\pi}{N} + \frac{|V(\ell)|}{|V(\ell+1)|}}. \quad (18)$$

For complex sinusoids and in the absence of noise, (18) is exact. For real sinusoids however, the accuracy of this frequency interpolator depends on ℓ and $\Delta\ell$ and is essentially independent of the magnitude and phase of the sinusoid. The dependency on ℓ is justified by the fact that for ‘small’ ℓ , in the order of 20 or less, leakage is significant and introduces systematic estimation errors.

The second estimator makes use of the three largest DFT bins because at most three DFT bins fit within the main lobe

of the frequency response of the sine window. Considering that

$$v(n) = x(n)h_S(n) \xleftrightarrow{\text{FT}} V(e^{j\omega}) = \frac{1}{2\pi} X(e^{j\omega}) * H_S(e^{j\omega}), \quad (19)$$

where $H_S(e^{j\omega}) = \sum_{n=0}^{N-1} h_S(n)e^{-j\omega n}$, $|V(\ell)|$ is a local maximum and either $|V(\ell+1)| > |V(\ell-1)|$, which denotes that the exact frequency may be estimated as $\ell + \Delta\ell$; or $|V(\ell+1)| < |V(\ell-1)|$ in which case the exact frequency may be estimated as $\ell - \Delta\ell$ [16]. The estimation of $\Delta\ell$ must therefore deliver a number between 0.0 and 0.5.

In the first case, when $0.0 \leq \Delta\ell \leq \gamma$, where γ is an optimization parameter, improved estimation accuracy is obtained by using $Q = \frac{|V(\ell-1)|}{|V(\ell+1)|} \approx \left[\frac{\cos \frac{\pi}{3}(\Delta\ell+1)}{\cos \frac{\pi}{3}(1-\Delta\ell)} \right]^{\frac{2}{G}}$, then

$$\hat{\Delta\ell} \approx \frac{3}{\pi} \arctan \frac{1}{\sqrt{3}} \frac{1-Q^G}{1+Q^G}. \quad (20)$$

When $\gamma \leq \Delta\ell \leq 0.5$, improved estimation accuracy is obtained by using $S = \frac{|V(\ell+1)|}{|V(\ell)|} \approx \left[\frac{\cos \frac{\pi}{3}(1-\Delta\ell)}{\cos \frac{\pi}{3}\Delta\ell} \right]^{1/F}$, from which we obtain

$$\hat{\Delta\ell} \approx \frac{3}{\pi} \arctan \frac{2S^F - 1}{\sqrt{3}}. \quad (21)$$

The optimization of the estimation error in the *minmax* sense is performed as in [18] which specifies γ , G and F .

In the second case, the same expressions are obtained after redefining $Q = \frac{|V(\ell+1)|}{|V(\ell-1)|}$ and $S = \frac{|V(\ell-1)|}{|V(\ell)|}$ [16], [18].

As shown in [18], the maximum absolute estimation error relative to the bin width is consistently below 0.1%.

VII. SIMULATION RESULTS

In this section we evaluate the relative performance of the rectangular window-based frequency estimators described in section IV and including the ArcTan estimator (18), and the relative performance of the Hanning window-based frequency estimators described in section V and including the ArcTan estimator combining (20) and (21).

A. Performance of rectangular window-based interpolators

Simulation results have been obtained for the DFT-based frequency interpolators identified as Jain99, Quinn97, Macleod98, ArcTanR and Jacobsen99 and corresponding to (9), (10), (11), (18), and (12), respectively.

Fig. 1 represents the square root of the error variance due to each frequency estimator after normalization by the bin width. It can be concluded that three regions may be identified where certain estimators show an advantage approaching the CRLB:

- 1) a region corresponding to SNRs between -10 dB and about 0 dB where the Jain99 and the ArcTanR estimator show a small advantage relative to the remaining estimators,
- 2) a region corresponding to SNRs between about 0 dB and about 10 dB where the Macleod98 estimator shows a marginal relative advantage,

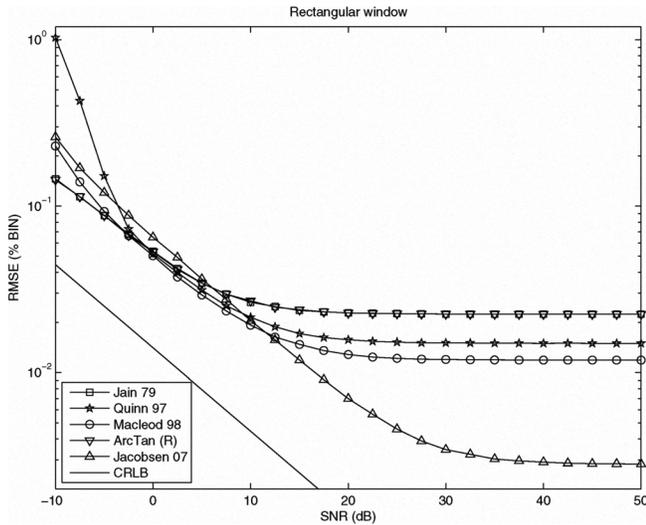


Fig. 1. RMSE (in % of the normalized bin width) of five rectangular window-based frequency estimators as a function of the SNR. The CRLB is also represented as a reference.

3) and a region corresponding to SNRs above about 10 dB where the Jacobsen07 essentially has no competitor.

As a conclusion, although at moderate (*i.e.*, around 10 dB) and high SNRs, all frequency estimators exhibit an RMSE around or better than 0.02% of the bin width, which may be considered fairly acceptable for practical purposes, the Quinn97 estimator should perhaps be avoided due to the poor performance at very low SNR, while the Jacobsen07 estimator is probably the best choice due to its remarkable performance at high SNRs. This result shows that the Jacobsen07 estimator not only is able ‘to cancel statistical bias’ [13], but is also robust to harmonic interference.

A final remark regards the quite similar performance between the Jain79 and ArcTanR estimators and their overall poor performance relative to that of the remaining rectangular window-based estimators. A plausible explanation is that these estimators use magnitude information only from two DFT bins while the remaining three estimators use magnitude and phase information from three DFT bins.

B. Performance of Hanning/Sine window-based interpolators

Simulation results are presented that have been obtained for the DFT-based frequency interpolators identified as Grandke83, Macleod98, Quadratic, Quinn06, ArcTanS and Jacobsen79 and corresponding to (13), (14), (15), (16), (20-21), and (17), respectively.

Results regarding the square root of the normalized error variance due to each frequency estimator are represented in Fig. 2 and reveal that the performance of the Quadratic estimator is the first to saturate for SNRs around 15 dB and above, followed by the Jacobsen07 estimator whose performance saturates for SNR around 30 dB and above, and followed by the ArcTanS estimator whose performance saturates around 40 dB SNR. The remaining three estimators (Grandke83, Quinn06 and Macleod98) follow a similar trend at an almost constant distance from the CRLB.

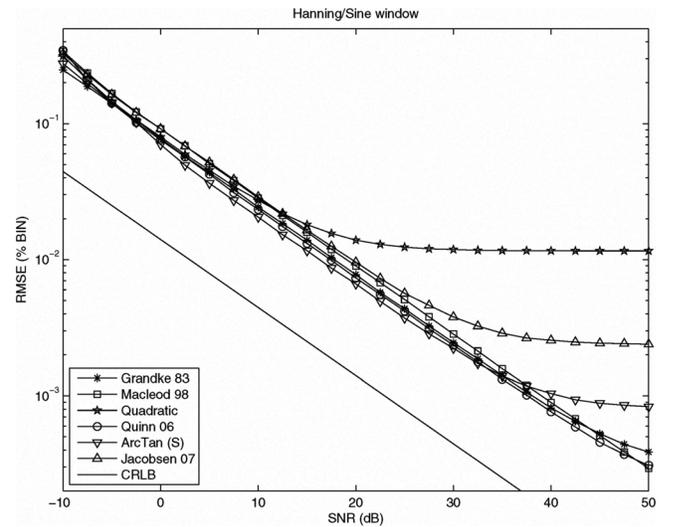


Fig. 2. RMSE (in % of the normalized bin width) of six Hanning/Sine window-based frequency estimators as a function of the SNR and when interfering sinusoids are about 10 bins apart from the target sinusoid. The CRLB is also represented as a reference.

It is interesting to note however that the ArcTanS estimator shows a marginal advantage over the remaining estimators since its normalized RMSE approaches better the CRLB for SNRs between about 0 dB and about 30 dB.

In order to understand better how the close proximity between different sinusoids (possibly harmonic sinusoids) as it happens frequently in singing, affects performance, the two sinusoids in (5) surrounding the target sinusoid have been separated from the latter by 5 bins instead of 10 bins. The simulation results for the new test conditions are represented in Fig. 3. This figure is very informative regarding the

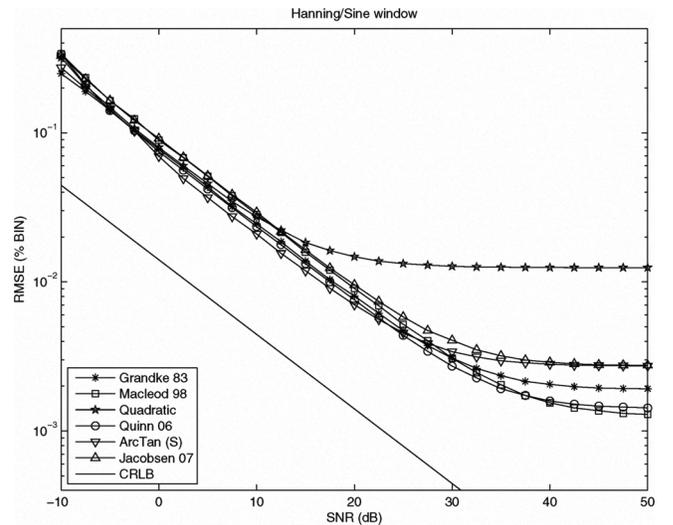


Fig. 3. RMSE (in % of the normalized bin width) of six Hanning/Sine window-based frequency estimators as a function of the SNR and when interfering sinusoids are about 5 bins apart from the target sinusoid. The CRLB is also represented as a reference.

impact on the RMSE performance due to the proximity of the interfering sinusoids to the target sinusoid. In fact, performance saturation occurs for all frequency estimators

between 10 dB SNR and 30 dB SNR. It is interesting to confirm however that the ArcTanS estimator keeps a marginal advantage over the remaining frequency estimators in the SNR range between about 0 dB and 20 dB since its RMSE approaches better the CRLB. This fact may possibly be related to the selectivity of the windows.

Fig. 4 represents the short-time Fourier spectrum of the resulting signal after windowing by the Hanning or sine window. It can be seen that in this case there is a better separation between the different sinusoids. In other words, the sine window allows better selectivity than the Hanning window.

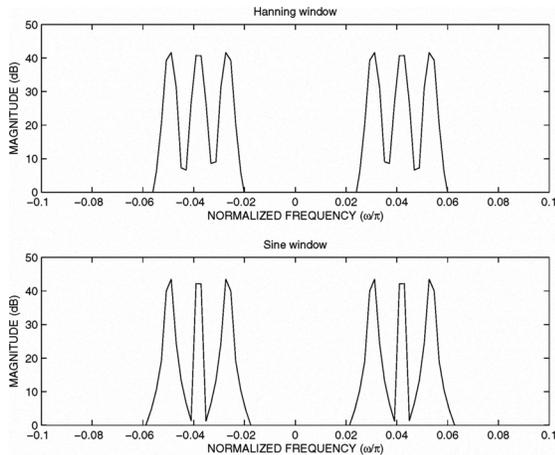


Fig. 4. Short-time Fourier spectrum of a target sinusoid ($\ell + \Delta\ell = 20.5$ bins) and two surround interfering sinusoids whose frequencies are 5.3 bins apart from the target. $N = 512$.

As an overall conclusion, Fig. 3 suggests that the Grandke83, ArcTanS, Macleod98 and the Quinn06 frequency estimators are quite acceptable choices for a wide range of SNRs and when sinusoids are closely spaced.

VIII. CONCLUSION

In this paper we have compared the performance of several non-iterative DFT frequency estimators that use either the rectangular window or the Hanning window, and including also two new ArcTan-based frequency estimators that use the rectangular window and the sine window. The test conditions have considered the influence due to noise and due to interfering sinusoids as this is the most probable scenario with natural audio signals such as singing. Results have revealed that at very low SNRs rectangular window-based interpolators may perform better than Hanning window-based interpolators. The inverse is true for moderate and high SNRs and the advantage may even exceed an order of magnitude. It has also been concluded that sinusoidal interference has the effect of reducing the performance of all estimators in three aspects: performance deviates more from the CRLB, performance is limited at high SNRs by an asymptotic trend, the performance of the different estimators becomes more ‘harmonized’ when sinusoidal interference is stronger, particularly when the Hanning window is used.

A final word is dedicated to the parabolic interpolator. This is a quite popular frequency estimation method when

used with the Hanning window and yet results show that its performance is the worse among all the tested estimators that use this window, which is easily explained by the fact that the parabolic interpolator is window agnostic. However, the parabolic interpolator may deliver improved results if a first stage of frequency interpolation is implemented, such as zero-padding [17]. Zero-padding is a technique frequently used to interpolate the information delivered by an N -point DFT by adding zeros to a data vector and computing a DFT larger than N [14], [10], [17]. Because zero-padding represents however a significant penalty in terms of computational complexity, we have not consider it in this paper.

REFERENCES

- [1] D. Hoppe, M. Sadakata, and P. Desain, “Development of real-time visual feedback assistance in singing training: a review,” *Journal of Computer Assisted Learning*, no. 22, pp. 308–316, 2006.
- [2] Johan Sundberg, *The Science of the Singing Voice*, Northern Illinois University Press, 1987.
- [3] D. C. Rife and G. A. Vincent, “Use of the discrete Fourier transform in the measurement of frequencies and levels of tones,” *Bell Systems Technical Journal*, vol. 49, no. 2, pp. 197–228, February 1970.
- [4] V. K. Jain, W. L. Collins, and D. C. Davis, “High-accuracy analog measurements via interpolated FFT,” *IEEE Transactions on Instrumentation and Measurement*, vol. 28, no. 2, pp. 113–122, June 1979.
- [5] Thomas Grandke, “Interpolation algorithms for discrete Fourier transforms of weighted signals,” *IEEE Transactions on Instrumentation and Measurement*, vol. 32, no. 2, pp. 350–355, June 1983.
- [6] H. Renders, J. Schoukens, and G. Vilain, “High-accuracy spectrum analysis of sampled discrete frequency signals by analytical leakage compensation,” *IEEE Transactions on Instrumentation and Measurement*, vol. 33, no. 4, pp. 287–292, December 1984.
- [7] J. Schoukens, R. Pintelon, and H. Van Hamme, “The interpolated fast Fourier transform: a comparative study,” *IEEE Transactions on Instrumentation and Measurement*, vol. 41, no. 2, pp. 226–232, April 1992.
- [8] Barry G. Quinn, “Estimation of frequency, amplitude, and phase from the DFT of a time series,” *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 814–817, March 1997.
- [9] Malcolm D. Macleod, “Fast nearly ml estimation of the parameters of real or complex single tones or resolved multiple tones,” *IEEE Transactions on Signal Processing*, vol. 46, no. 1, pp. 141–148, January 1998.
- [10] M. Betsler, P. Collen, G. Richard, and B. David, “Review and discussion on classical stft-based frequency estimators,” *120th Convention of the Audio Engineering Society*, 2006, Paper n. 6765.
- [11] Jeffrey D. Klein, “Fast algorithms for single frequency estimation,” *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1762–1770, May 2006.
- [12] B. G. Quinn, “Frequency estimation using tapered data,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2006, pp. III–73–III–76.
- [13] Eric Jacobsen and Peter Kootsookos, “Fast, accurate frequency estimators,” *IEEE Signal Processing Magazine*, pp. 123–125, May 2007.
- [14] Barry G. Quinn, “Recent advances in rapid frequency estimation,” *Digital Signal Processing (Elsevier)*, no. 19, pp. 942–948, 2009.
- [15] Alan V. Oppenheim, Ronald W. Schaffer, and John R. Buck, *Discrete-time Signal Processing*, Prentice-Hall Inc., 1998, 2nd ed.
- [16] Ricardo Sousa and Anibal J. S. Ferreira, “Non-iterative frequency estimation in the DFT magnitude domain,” in *ISCCP 2010*, 2010, submitted for publication.
- [17] M. Abe and J. O. Smith III, “Design criteria for simple sinusoidal parameter estimation based on quadratic interpolation of the FFT magnitude peaks,” *117 Convention of the Audio Engineering Society*, 2004, Paper n. 6256.
- [18] Anibal Ferreira and Deepen Sinha, “Accurate and robust frequency estimation in the ODFT domain,” in *2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 2005, pp. 203–206.