CENTER FOR COMPUTER RESEARCH IN MUSIC AND ACOUSTICS
DEPARTMENT OF MUSIC, STANFORD UNIVERSITY
REPORT NO. STAN-M-117

CQIFFT: Correcting Bias in a Sinusoidal Parameter Estimator based on Quadratic Interpolation of FFT Magnitude Peaks

October 13, 2004

Mototsugu Abe and Julius O. Smith III
E-Mail: {abemoto, jos}@ccrma.stanford.edu

Abstract

Quadratic peak interpolation in a zero-padded Fast Fourier Transform has been widely used for sinusoidal parameter estimation in audio applications. Since it approximates the true spectral peak shape with a quadratic polynomial, the estimates with non-Gaussian windows are more or less biased. In this paper, we propose an efficient bias correction method as an extension of the quadratic interpolation. We show simple bias-correcting functions with 4 extra multiplications greatly improve the estimation accuracy. The improvement is up to the order of 10 to 100, so that the required zero-padding factors can be reduced to 1.1 for the 0.1% error bound.

1 Introduction

Sinusoidal modeling [1, 2, 3] has been widely used to represent the most salient aspects of tonal sound. A key component of sinusoidal modeling is the estimation of the parameters of multiple sinusoids from recorded data. Among various approximate maximum likelihood (ML) estimators [4, 5, 6, 7, 8], quadratic interpolation of magnitude peaks in a zero-padded Fast Fourier Transform [1] (referred as the Quadratically Interpolated FFT, or QIFFT method) has been widely used due to its simplicity and accuracy, which is sufficient for most audio purposes [9, 10].

However, since the QIFFT method approximates the true spectral peak shape with a quadratic polynomial, the estimates are generally biased. For example, the biases in frequency and amplitude estimates with no zero padded Hann window are 1.6% and 3.8% respectively [14]. If we re-synthesized a sound with these differences, the errors would yield audible artifacts.

To reduce this bias, the QIFFT method is generally used with a certain amount of zero padding, in exchange for some computational costs. If we are to bound the frequency-bias below 0.1% for a Hann window of 1323 samples (30ms at the sampling frequency of 44.1kHz), since a zero-padding factor of 2.4 or greater is needed [14], we have to call for an FFT of 4096 samples.

Other than the zero padding, a commonly used method to correct the bias is to utilize an iterative, nonlinear optimization, such as Newton's method [6, 11]. Or, some non-iterative methods using other functions than a parabola are also proposed. Abeysekera [12] proposes to use Hilbert transform to approximate the peak shape. Zacharov [13] proposes the dichotomous search scheme. However, the computational costs of these methods are generally far more than those of the QIFFT method in which only 6 multiplications and 1 division per peak are needed.

In this paper, we propose an efficient bias-correction method for the frequency and amplitude estimates of the QIFFT. The key point is to introduce information about the spectral peak shape as window-dependent bias correction functions. We show simple cubic and parabolic functions can efficiently correct the frequency and amplitude biases respectively. The improvement is up to the order of 10 to 100, so that the required zero-padding factors can be reduced to 1.1 for the 0.1% error bound. The proposed method only requires additional 4 multiplications per peak to the original QIFFT. Noise robustness of the proposed method is also discussed.

We use the following symbols in this paper: N: FFT size, M: window length, Z_p: zero-padding factor ($= N/M$), and F_s: sampling frequency.

2 QIFFT method

The Quadratically Interpolated FFT (QIFFT) method for estimating sinusoidal parameters from peaks in spectral magnitude data can be summarized as follows:

1. Calculate the amplitude and phase spectrum of audio data, by using an appropriately zero-padded FFT with an appropriate window of an appropriate length (points in Fig. 1).

2. Find the bin number of the maximum peak magnitude ($k_{\max}$).

3. Quadratically interpolate the log-amplitude of the peak using two neighboring samples (dotted line), and define $\hat{\Delta}$ as the distance in bins from $k_{\max}$ to the parabola peak.

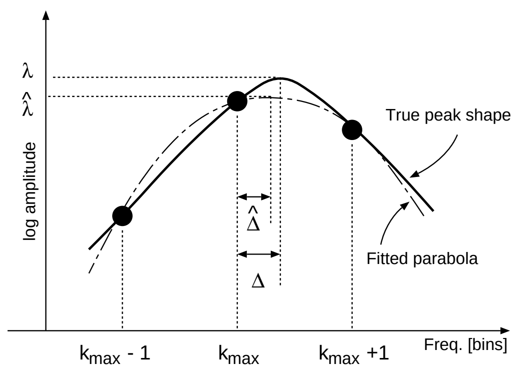

Figure 1: Quadratic interpolation of spectral peak

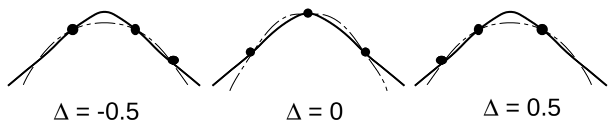

Figure 2: Three extreme cases ($\Delta = -0.5, 0$, and 0.5).

4. Estimate the peak frequency in bins as $k_{\max} + \hat{\Delta}$, and define the estimated peak log-amplitude $\hat{\lambda}$, as the interpolation (based on $\hat{\Delta}$) of the log-amplitude samples at $k = k_{\max} - 1$, $k_{\max}$, and $k_{\max} + 1$.

5. Estimate the phase, if needed, by interpolating[1] the phase spectrum based on the interpolated frequency estimate.

6. Subtract the peak from the FFT data for subsequent processing.

7. Repeat steps 2-6 above for each peak.

Note that quadratic interpolation is not reliably applicable to a rectangular window with a zero-padding factor below 1.5, since the number of sampling points in the main lobe becomes less than 3.

3 Interpolation Bias

A schematic description of the interpolation bias is shown in Fig .1, in which Δ and $\hat{\Delta}$ denote the true and estimated offset of the peak position from the nearest FFT bin respectively (in units of FFT bins), and λ and $\hat{\lambda}$ denote the true and estimated log-amplitude respectively. We can easily deduce from the figure that the amplitude and frequency biases do not depend on $k_{\max}$ or λ. They depend only on the offset from the FFT bin (Δ).

Three extreme cases, which are $\Delta = -0.5, 0$ and 0.5, are shown in Fig .2. When $\Delta = 0$, both the frequency and amplitude estimates are unbiased. When $\Delta = \pm 0.5$, the frequency estimates are again unbiased, but the amplitude estimates are maximally biased. In addition, we can easily confirm that the frequency bias is odd symmetric, whereas the amplitude bias is even symmetric.

Figure 3 shows experimentally obtained biases in the frequency and amplitude estimates with Hann windows for various Δ and zero-padding factors. We can confirm that the frequency bias has three zeros at $\hat{\Delta} = -0.5, 0$ and 0.5, and the amplitude bias has double zeros at $\hat{\Delta} = 0$, as predicted from the extreme cases.

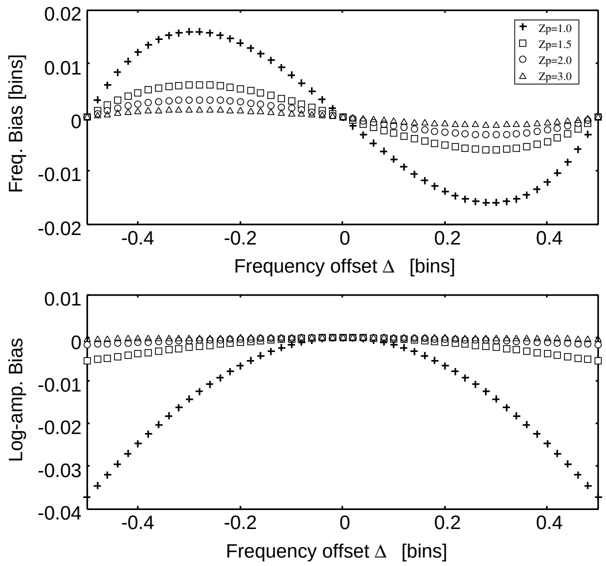

Figure 3: Experimentally obtained biases in the QIFFT (Hann window, FFT size $N = 4096$): frequency bias (top) and amplitude bias (bottom).

Table 1: Coefficients of the correction functions.

	c0	c1	c2	c3
Rect	1.279369	1.756245	−1.173273	−3.241966
Hann	0.247560	0.084372	−0.090608	−0.055781
Hamm	0.256498	0.075977	−0.116927	−0.062882
Black	0.124188	0.013752	−0.038073	−0.006195
KB(1.5)*	0.309479	0.141430	−0.132571	−0.134588
KB(2.0)	0.199657	0.044008	−0.078430	−0.027973
KB(2.5)	0.135819	0.017893	−0.045315	−0.008833
KB(3.0)	0.097632	0.008615	−0.027991	−0.003516

(*) KB(α) : Kaiser-Bessel windows of the indicated α value.

4 Simple Bias-Correction Functions

The simplest bias-correction functions which satisfy the above conditions can be written as

$$\check{\Delta} = \hat{\Delta} + \xi_{Z_p}(\hat{\Delta} - 0.5)(\hat{\Delta} + 0.5)\hat{\Delta}, \tag{1}$$
$$\check{\lambda} = \hat{\lambda} + \eta_{Z_p}\hat{\Delta}^2, \tag{2}$$

where $\check{\Delta}$ and $\check{\lambda}$ are the bias-corrected estimates, and ξ_{Z_p} and η_{Z_p} denote the scales of the cubic and parabolic functions. As shown in Fig. 3, the scales depend on zero-padding factors. To determine the dependency, we independently obtain the scales for some fixed zero-padding factors using least square method, as shown by the points in Fig. 4. From this result, we can deduce that this dependecy can be effectively approximaed by combinations of exponential functions, as

$$\xi_{Z_p} = c_0 Z_p^{-2} + c_1 Z_p^{-4}, \tag{3}$$
$$\eta_{Z_p} = c_2 Z_p^{-4} + c_3 Z_p^{-6}, \tag{4}$$

[1]Though either quadratic or linear interpolation can be used, we use quadratic interpolation in this paper.

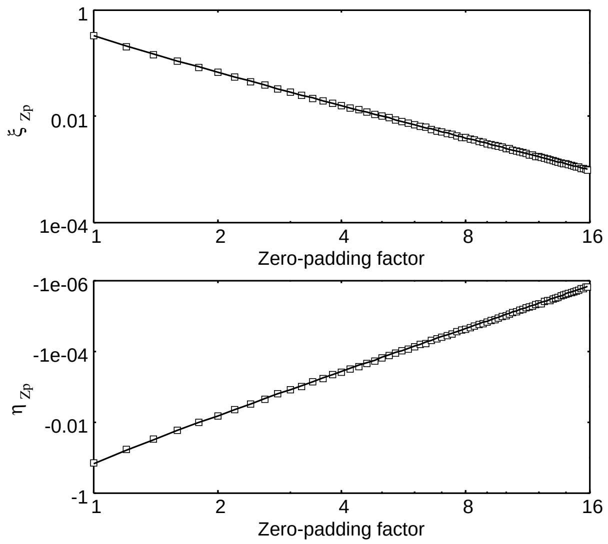

Figure 4: Scales (ξ_{Z_p}, η_{Z_p}) obtained by the least square method independently applied for fixed zero-padding factors (points), and those obtained by the approximation functions (lines). (Hann window, FFT size $N = 64, 128, ..., 8192$.)

where c_0, c_1, c_2 and c_3 are window dependent coefficients, shown in Table 1.

The solid lines in Fig. 4 show the scales calculated by Eqs. (3) and (4). We can confirm that these functions well approximate the scales in wide range of zero-padding factors. The lines in Fig. 5 show the bias-correction functions in Eqs. (1), (2), (3),(4) and Table 1, shown with the experimentally obtained biases (points). Although slight difference can be seen for the zero-padding factor of 1.0, the bias-correction functions quite well fit to the actual biases.

We call the QIFFT method with the correction functions as "the Corrected QIFFT (CQIFFT) method". Note that since $\hat{\Delta}^2$ appears in both Eqs. (1) and (2), only 4 multiplications per peak are needed in addition to the original QIFFT method.

5 Comparison of the CQIFFT with the QIFFT

5.1 Experimental conditions

The signal to be identified is a time-invariant complex sinusoid, defined by

$$x(n) = A_0 e^{j(\omega_0 n + \phi_0)}, \quad n = 0, 1, 2, \dots , \tag{5}$$

where A_0, ω_0, ϕ_0 denote the amplitude, (normalized radian) frequency and phase respectively. We estimate $\{\hat{A}_0, \hat{\omega}_0, \hat{\phi}_0\}$ using the QIFFT method and evaluate the biases, as

$$\begin{aligned} \mathrm{Bias}_\omega &= |\hat{\omega}_0 - \omega_0|/(2\pi/M), & (6)\\ \mathrm{Bias}_A &= |\hat{A}_0 - A_0|/A_0, & (7)\\ \mathrm{Bias}_\phi &= |\hat{\phi}_0 - \phi_0|/\pi. & (8) \end{aligned}$$

Note that since the frequency bias is normalized by $\Omega_M \triangleq 2\pi/M$, where M is the window length, it becomes nearly independent of the actual window length. We prepare 512 sinusoids by randomly changing $\{A_0, \omega_0, \phi_0\}$, using the FFT sizes $N = \{64, 128, ..., 8192\}$, and setting the window length $M \geq 31$ to the maximum odd integer not exceeding N/Z_p. By sweeping the zero-padding factor Z_p from 1.0 to 16.0 with the step of 0.1, and by taking the maximum biases out of the 4096 (8 FFT sizes $\times$ 512 sinusoids) test sets for each zero-padding factor, we get the maximum bias curves.

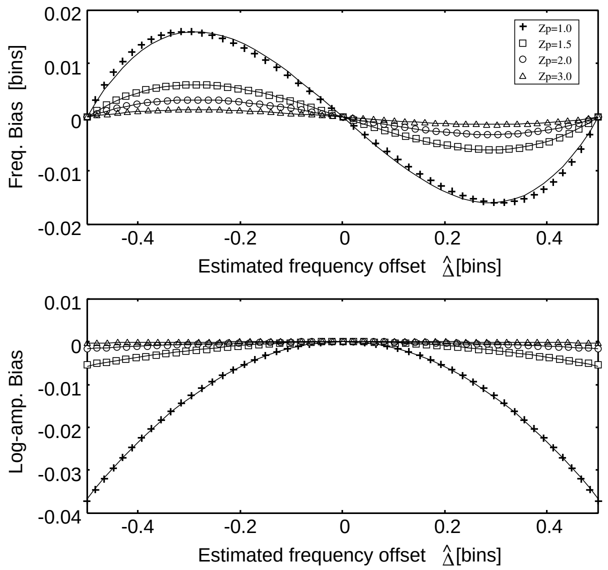

Figure 5: Experimentally obtained biases in the QIFFT (points) and the bias-correction functions (lines).

5.2 Results

Figures 6 to 13 show the frequency and amplitude biases in both the CQIFFT and QIFFT for various windows. Other than a rectangular window with a zero-padding factor below 1.5, which is not appropriate for the QIFFT as mentioned in the previous section, the improvement of the estimation accuracy is up to the order of 10 to 100 for all the windows with any zero-padding factors.

The maximum biases for given zero-padding factors are summarized in Table 2. Conversely, the required minimum zero-padding factors for given error bounds are shown in Table 3. To bound the frequency-bias below 0.1%, zero-padding factors of {1.1, 1.1, 1.0} for {Hann, Hamming, Blackman} windows are sufficient for the CQIFFT, whereas {2.4, 2.4, 1.9} are required for the QIFFT. The reduction of the zero-padding factors is generally more than a half, which greatly reduces the computational costs in the FFT.

6 Noise Robustness

In this section, in order to compare the original and corrected QIFFT methods as an approximate ML estimator, we numerically measure the error variance in the estimated sinusoidal parameters using a fixed zero-padding factor and a variety of signal-to-noise S/N ratios, where the noise is chosen to be additive white Gaussian noise (AWGN). Specifically, we use an FFT size of 4096 and a Hann window with a zero-padding factor of 2.5. The sinusoidal parameters are randomly given, as in the numerical experiments in the previous section.

The results are shown in Fig. 14. We also show the Cramer-Rao Bound (CRB) and audible error limits which we used 0.1% in frequency and 0.1dB ($\approx 1.2\%$) in amplitude. At low S/N ratios (SNR < -10dB), the errors are far larger than the CRBs. This is the so-called "threshold effect", in which a spurious noise peak is detected instead of the sinusoidal peak [4]. At moderate S/N ratios (SNR < 30dB for the QIFFT, SNR < 60dB for the CQIFFT), we find that the QIFFT works essentially as well as the ML estimator. At high S/N ratios, the errors are dominated by the biases. We can confirm that the biases in the CQIFFT are far smaller than those in the QIFFT.

7 Summary

In this paper, we investigated the bias due to quadratic interpolation, and proposed simple bias correction functions for frequency and amplitude estimates. The method improve the estimation accuracy up to the order of 10 to 100 with only 4 additional multiplications per peak.

One drawback of the CQIFFT is the use of the spectral peak shape of a window. Since, in general, frequency chirp in a time-varying sinusoid widen the peak shape, signals to be estimated should be reasonably stable within window length to have the benefit of the bias correction.

References

[1] J. O. Smith III and X. Serra: "PARSHL: A program for the analysis/synthesis of inharmonic sounds based on a sinusoidal representation," in Proc. ICMC'87, available at http://www-ccrma.stanford.edu/˜jos/parshl.

[2] R. J. McAulay and T. F. Quatieri: "Speech Analysis/Synthesis Based on a Sinusoidal Representation," IEEE Acoust. Speech Sig. Proc., Vol.34, No.4, 744/754 (1986).

[3] M. Goodwin: "Residual Modeling in Music Analysis-Synthesis," Proc. IEEE ICASSP'96, 1005/1008 (1996).

[4] D. C. Rife and R. R. Boorstyn: "Single-Tone Parameter Estimation from Discrete-Time Observations," IEEE Trans. Info. Theory, 20, 5, 591/598 (1974).

[5] D. J. Thomson: "Spectrum Estimation and Harmonic Analysis," Proc. of the IEEE, 70, 9, 1055/1096 (1982).

[6] T. J. Abatzoglou: "A Fast Maximum Likelihood Algorithm for Frequency Estimation of a Sinusoid Based on Newton's Method," IEEE Trans. Acoust., Speech, Signal Processing, 33, 1, 77/89 (1985).

[7] B. G. Quinn: "Estimation of Frequency, Amplitude, and Phase from the DFT of a Time Series," IEEE Trans. Signal Processing, 45, 3, 814/817 (1997).

[8] M. D. Macleod: "Fast Nearly ML Estimation of the Parameters of Real or Complex Single Tones or Resolved Multiple Tones," IEEE Trans. Signal Processing, 46, 1, 141/148 (1998).

[9] R. C. Maher and J. W. Beauchamp: "Fundamental Frequency Estimation of Musical Signals Using a Two-way Mismatch Procedure," J. Acoust. Soc. Am., 95, 4, 2254/2263 (1994).

[10] A. Klapuri: "Multipitch Estimation and Source Separation by the Spectral Smoothness Principle," Proc. IEEE ICASSP'01, 3381/3384 (2001).

[11] Ph. Depalle and T. Hélie: "Extraction of Spectral Peak Parameters using a Short-Time Fourier Transform Modeling and No Sidelobe Windows," Proc. IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics (Mohonk'97), (1997).

[12] S. S. Abeysekera: "An Efficient Hilbert Transform Interpolation Algorithm for Peak Position Estimation," Proc. IEEE Signal Processing Workshop on Statistical Signal Processing, 417/420 (2001).

[13] Y. V. Zakharov and T. C. Tozer: "Frequency Estimator with Dichotomous Search of Periodogram Peak," IEEE Electronic Letters, 35, 19, 1608/1609 (1999).

[14] M. Abe and J. O. Smith III: "Design Criteria for the Quadratically Interpolated FFT Method (I): Bias due to Interpolation," Technical Report STAN-M-114, Dept. of Music, Stanford University, October, (2004).

[15] M. Abe and J. O. Smith III: "Design Criteria for the Quadratically Interpolated FFT Method (II): Bias due to Interfering Components," Technical Report STAN-M-115, Dept. of Music, Stanford University, October, (2004).

[16] M. Abe and J. O. Smith III: "Design Criteria for the Quadratically Interpolated FFT Method (III): Bias due to Amplitude and Frequency Modulation," Technical Report STAN-M-116, Dept. of Music, Stanford University, October, (2004).

[17] M. Abe and J. O. Smith III: "AM/FM Rate Estimation and Bias Correction for Time-Varying Sinusoidal Modeling," Technical Report STAN-M-118, Dept. of Music, Stanford University, October, (2004).

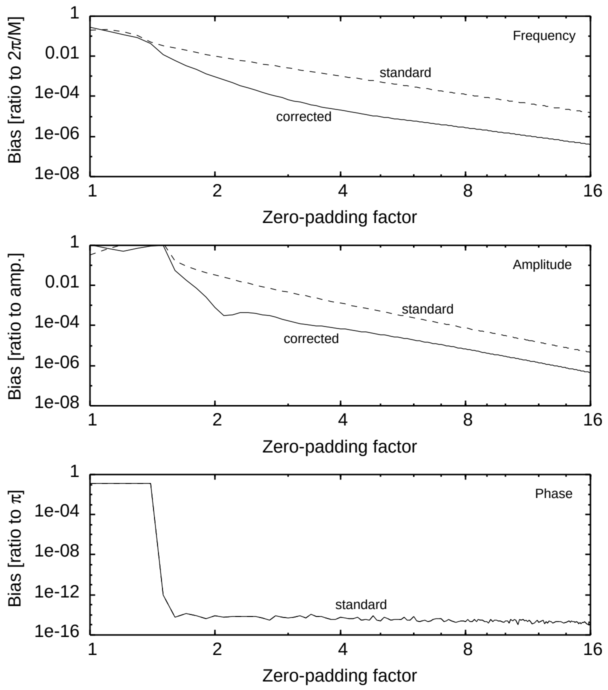

Figure 6: Biases in the CQIFFT and QIFFT with rectangular windows.

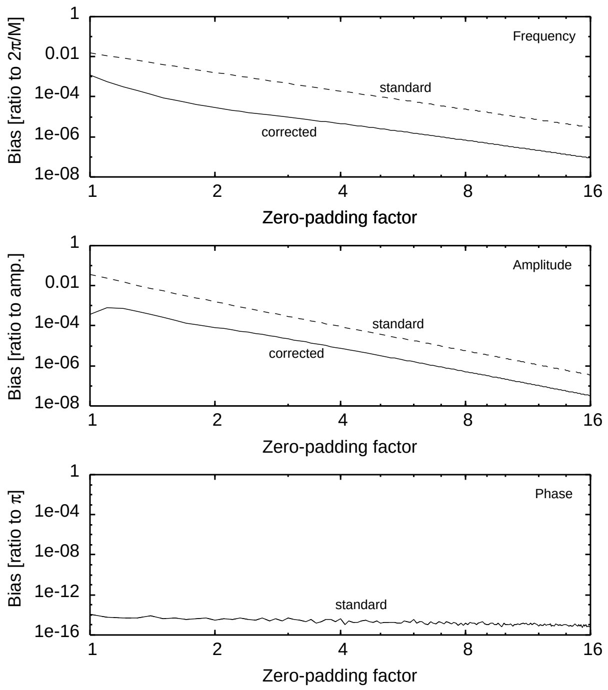

Figure 7: Biases in the CQIFFT and QIFFT with Hann windows.

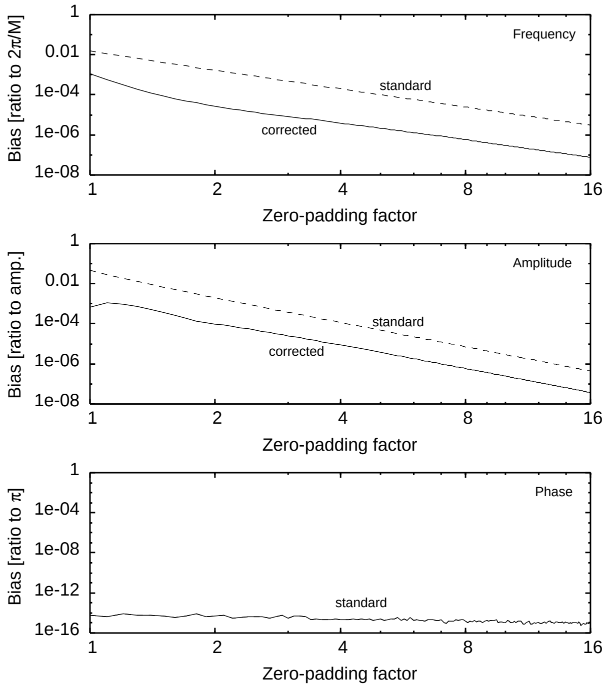

Figure 8: Biases in the CQIFFT and QIFFT with Hamming windows.

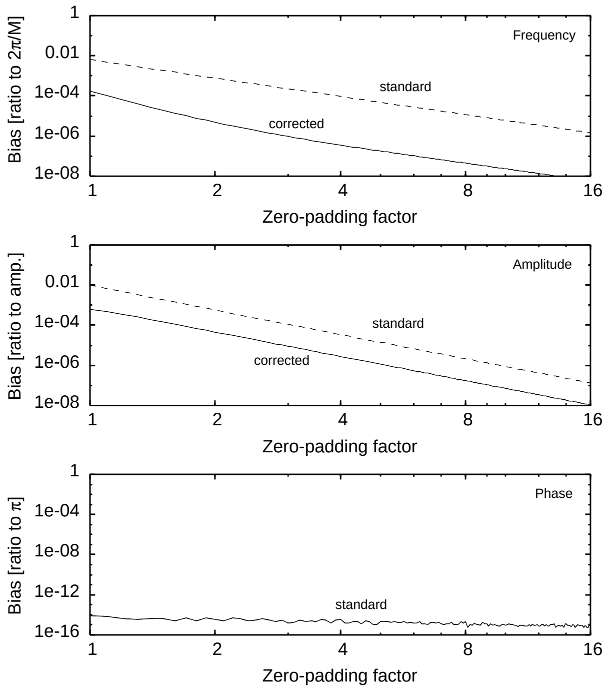

Figure 9: Biases in the CQIFFT and QIFFT with Blackman windows.

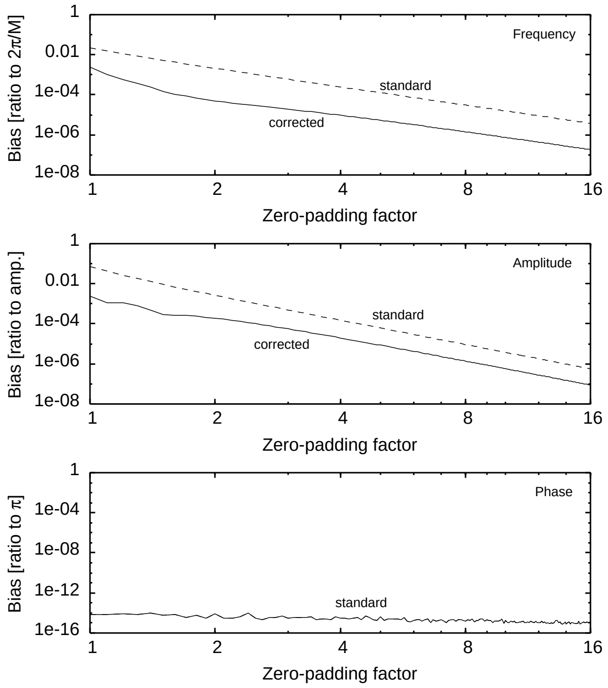

Figure 10: Biases in the CQIFFT and QIFFT with Kaiser-Bessel window of $\alpha = 1.5$.

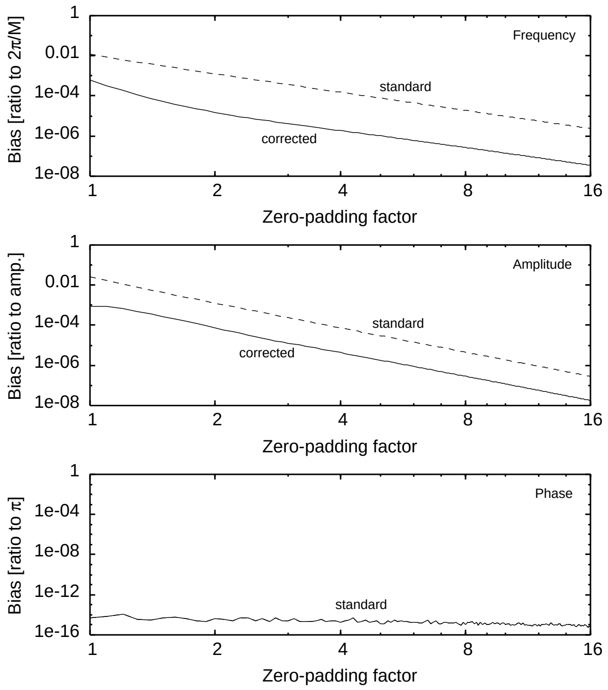

Figure 11: Biases in the CQIFFT and QIFFT with Kaiser-Bessel window of $\alpha = 2.0$.

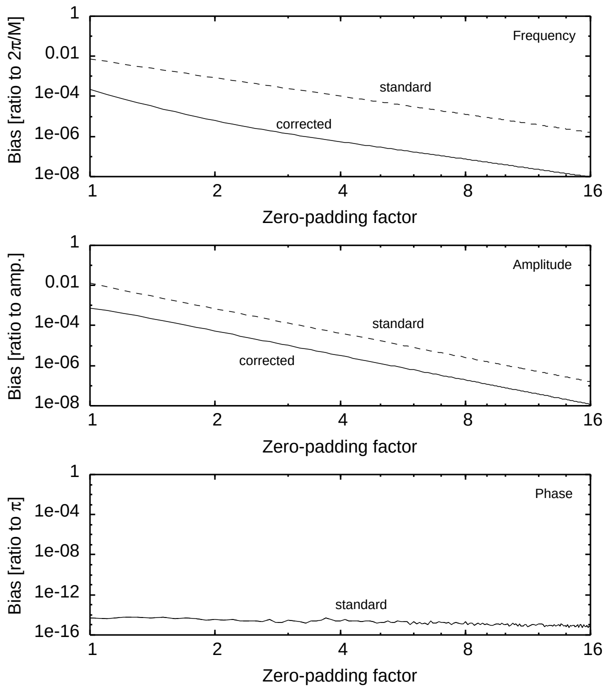

Figure 12: Biases in the CQIFFT and QIFFT with Kaiser-Bessel window of α = 2.5.

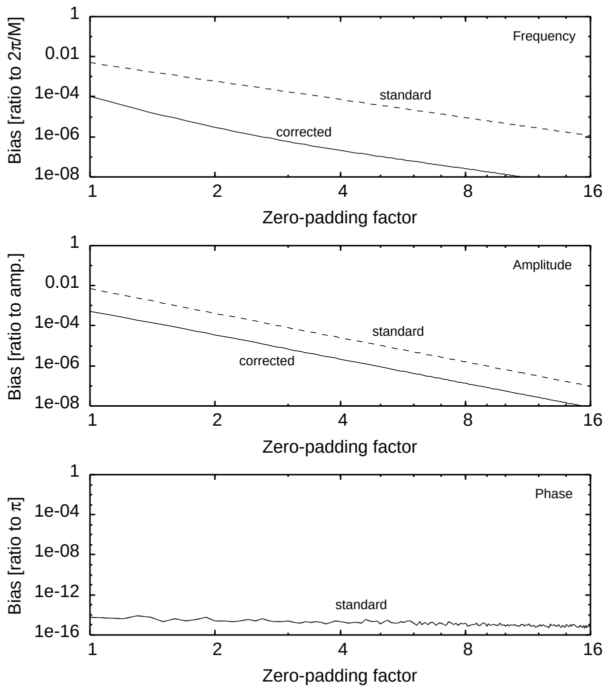

Figure 13: Biases in the CQIFFT and QIFFT with Kaiser-Bessel window of α = 3.0.

Table 2: Maximum biases for given zero-padding factors.

window	Z_p	QIFFT (%)		CQIFFT (%)	
		Frq.	Amp.	Frq.	Amp.
Rect	1.0	19.127	32.119	29.091	107.48
	2.0	1.0360	3.2756	0.0930	0.0820
	3.0	0.2613	0.4572	0.0071	0.0179
	4.0	0.1047	0.1315	0.0021	0.0071
	5.0	0.0526	0.0520	0.0010	0.0036
Hann	1.0	1.5992	3.7933	0.1208	0.0380
	2.0	0.1624	0.1587	0.0029	0.0084
	3.0	0.0467	0.0298	0.0010	0.0022
	4.0	0.0195	0.0093	0.0005	0.0008
	5.0	0.0100	0.0038	0.0003	0.0004
Hamm	1.0	1.6008	4.6495	0.1141	0.0680
	2.0	0.1663	0.1998	0.0027	0.0099
	3.0	0.0479	0.0376	0.0009	0.0026
	4.0	0.0200	0.0117	0.0004	0.0009
	5.0	0.0102	0.0048	0.0003	0.0004
Black	1.0	0.6634	1.0531	0.0175	0.0642
	2.0	0.0767	0.0572	0.0005	0.0047
	3.0	0.0225	0.0111	0.0001	0.0010
	4.0	0.0095	0.0035	0.0001	0.0003
	5.0	0.0049	0.0015	0.0001	0.0002
KB	1.0	2.1744	7.4126	0.2321	0.2473
$\alpha = 1.5$	2.0	0.2094	0.2645	0.0050	0.0189
	3.0	0.0598	0.0490	0.0020	0.0056
	4.0	0.0249	0.0152	0.0010	0.0020
	5.0	0.0127	0.0062	0.0006	0.0009
KB	1.0	1.1728	2.6426	0.0598	0.0892
$\alpha = 2.0$	2.0	0.1270	0.1259	0.0016	0.0075
	3.0	0.0368	0.0240	0.0005	0.0013
	4.0	0.0154	0.0075	0.0002	0.0005
	5.0	0.0079	0.0031	0.0002	0.0002
KB	1.0	0.7394	1.2971	0.0226	0.0728
$\alpha = 2.5$	2.0	0.0844	0.0689	0.0007	0.0054
	3.0	0.0247	0.0133	0.0002	0.0011
	4.0	0.0104	0.0042	0.0001	0.0004
	5.0	0.0053	0.0017	0.0001	0.0002
KB	1.0	0.5110	0.7422	0.0105	0.0506
$\alpha = 3.0$	2.0	0.0600	0.0416	0.0004	0.0036
	3.0	0.0176	0.0081	0.0001	0.0007
	4.0	0.0074	0.0026	0.0001	0.0003
	5.0	0.0038	0.0011	0.0001	0.0001

Table 3: Zero-padding factors required for given bias bounds.

window	max bias	QIFFT		CQIFFT	
		Frq.	Amp.	Frq.	Amp.
Rect	1.00%	2.1	2.6	1.6	1.8
	0.50%	2.5	3.0	1.7	1.9
	0.10%	4.1	4.3	2.0	2.0
	0.01%	8.7	7.5	2.9	3.5
Hann	1.00%	1.2	1.4	1.0	1.0
	0.50%	1.5	1.6	1.0	1.0
	0.10%	2.4	2.3	1.1	1.0
	0.01%	5.0	4.0	1.5	1.9
Hamm	1.00%	1.2	1.4	1.0	1.0
	0.50%	1.5	1.7	1.0	1.0
	0.10%	2.4	2.4	1.1	1.2
	0.01%	5.1	4.2	1.5	2.0
Black	1.00%	1.0	1.1	1.0	1.0
	0.50%	1.1	1.2	1.0	1.0
	0.10%	1.9	1.8	1.0	1.0
	0.01%	4.0	3.1	1.2	1.7
KB	1.00%	1.3	1.5	1.0	1.0
(α = 1.5)	0.50%	1.6	1.8	1.0	1.0
	0.10%	2.6	2.6	1.2	1.3
	0.01%	5.5	4.5	1.7	2.6
KB	1.00%	1.1	1.3	1.0	1.0
(α = 2.0)	0.50%	1.3	1.5	1.0	1.0
	0.10%	2.2	2.2	1.0	1.0
	0.01%	4.7	3.8	1.4	1.9
KB	1.00%	1.0	1.1	1.0	1.0
(α = 2.5)	0.50%	1.2	1.3	1.0	1.0
	0.10%	1.9	1.9	1.0	1.0
	0.01%	4.1	3.3	1.2	1.8
KB	1.00%	1.0	1.0	1.0	1.0
(α = 3.0)	0.50%	1.1	1.1	1.0	1.0
	0.10%	1.7	1.7	1.0	1.0
	0.01%	3.7	2.9	1.1	1.6

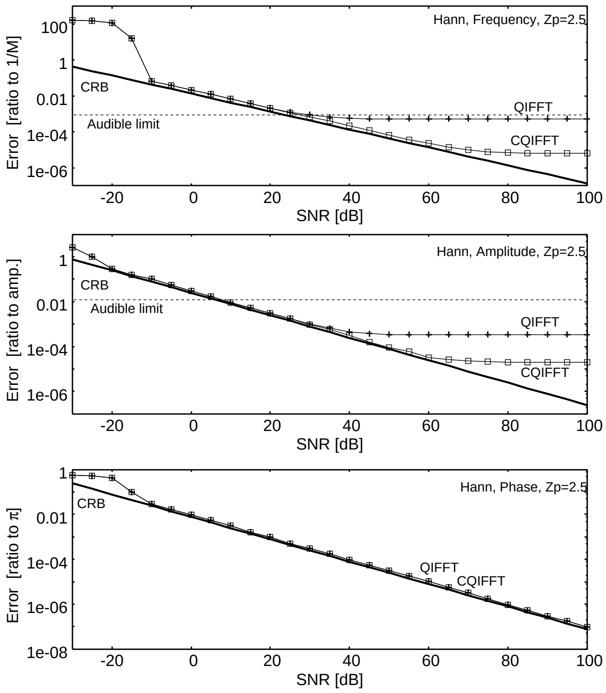

Figure 14: RMS estimation errors for a sinusoid with AWGN (Hann window, $N = 4096$, $Zp = 2.5$): frequency (top), amplitude (middle) and phase (bottom).