CENTER FOR COMPUTER RESEARCH IN MUSIC AND ACOUSTICS
DEPARTMENT OF MUSIC, STANFORD UNIVERSITY
REPORT NO. STAN-M-114

Design Criteria for the Quadratically Interpolated FFT Method (I): Bias due to Interpolation

October 13, 2004

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

Abstract

Due to its simplicity and accuracy, quadratic peak interpolation in a zero-padded Fast Fourier Transform (FFT) has been widely used for sinusoidal parameter estimation in audio applications. While general criteria can guide the choice of window type, FFT length, and zero-padding factor, it is sometimes desirable in practice to know more precisely the requirements for achieving a prescribed error bound. The purpose of this series of papers is to investigate the estimation error bias associated with the choice of analysis parameters, and provide precise criteria for designing the optimum estimator. In this first paper, we focus on the bias caused by quadratic interpolation. We numerically determine the maximum error bias in the estimation of the frequency, amplitude and phase for various zero-padding factors, and determine the minimum zero-padding factors needed for achieving prescribed error bounds. The results show, for example, zero-padding factors of {4.1, 2.4, 1.9} are sufficient for achieving a 0.1% frequency-error bound when using {rectangular, Hann or Hamming, Blackman} windows, respectively. Noise robustness of the estimator is also investigated. We confirm the estimator works as well as a ML estimator within a certain S/N ratio range.

1 Introduction

Sinusoidal modeling [1, 2, 3] has been widely used to represent the most salient aspects of tonal sound. Applications include analysis, synthesis, modification, coding, and the like [4, 5, 6, 7]. 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 [8, 9, 10, 11, 12], quadratic interpolation of magnitude peaks in a Fast Fourier Transform [1] (referred to 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 [13, 14]. Indeed, it works nearly as well as a ML estimator for well resolved sinusoids when a large FFT zero-padding factor is used.

In practice, however, we are shackled with various restrictions, such as 1) we may wish to minimize zero padding for computational efficiency, 2) we may need to consider interference from nearby signal components, and 3) we may have to deal with time-varying sinusoids. The zero-padding factor determines the maximum error bias caused by the quadratic interpolation. The window type and length determines the maximum error bias caused by the interfering components and the time-variation of the sinusoidal components. When it is desired to minimize computational cost given a prescribed error tolerance, more accurate design formulas are needed than what appear to be available in the literature to date.

In this series of papers, we theoretically predict and numerically measure the estimation error bias associated with the choice of design parameters, such as the window type, FFT length and zero-padding factor. The goal is to provide precise criteria for designing the optimum estimator.

In this first paper, we focus on the bias caused by quadratic interpolation. The bias is due to the difference between the true peak shape and the fitted quadratic polynomial. Since zero padding in the time domain accurately interpolates the frequency sampling points of the FFT, this bias can be reduced by using a larger zero-padding factor. As for quantitative guidelines, Smith [1] shows that a zero-padding factor of more than 5 can bound the maximum frequency-bias below 0.1% of the distance from a sinc maximum to its first zero crossing for rectangular windows. Brown [15] proved that the frequency bias is up to 5.3% of the FFT bin spacing with a non-zero-padded Hann window.[1] However, if we wish to minimize the zero-padding factor, we need to know more precise relations between the maximum bias and the zero-padding factors. In Section 3, we numerically determine the maximum error bias in the estimation of the frequency, amplitude and phase for various zero-padding factors, and determine the minimum zero-padding factors needed for achieving prescribed error bounds. Noise robustness of the estimator is also discussed. Although the estimation error variance approaches the Cramer-Rao bound (CRB) as the FFT zero-padding factor is increased, for well resolved sinusoids, the performance at high signal-to-noise (S/N) ratios is restricted by the above mentioned bias. We determine numerically the accuracy of a family of QIFFT estimators for noisy signals at various S/N ratios in Section 4.

The following symbols are used in this paper:

$$N \triangleq \text{FFT length (samples)}$$

[1]This result is for quadratic interpolation in a (linear) magnitude spectrum. Use of log-magnitude spectrum generally yields more accurate estimates [1].

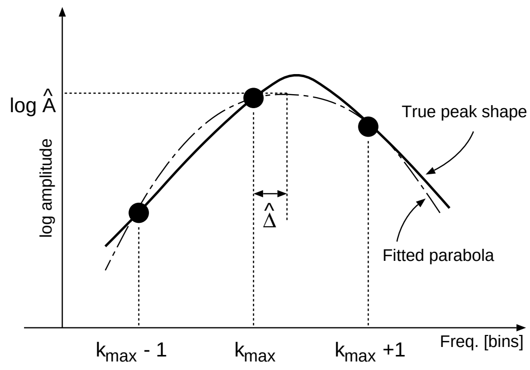

Figure 1: Quadratic interpolation of spectral peak.

$$M \triangleq \text{window length,}$$
$$Z_p \triangleq N/M = \text{zero-padding factor, and}$$
$$F_s \triangleq \text{sampling frequency (Hz).}$$

Normalized radian frequency ω is in units of radians per sample and taken to lie in the interval $\omega \in [-\pi, \pi)$.

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.
4. Estimate the peak frequency in bins as $k_{\max} + \hat{\Delta}$, and define the estimated peak amplitude $\hat{A}$ 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[2] 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.

This series of papers is concerned primarily with finding the "appropriate" window type, window length, and zero-padding factor. 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 Minimum Zero-padding factor

3.1 Conditions for numerical experiments

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

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

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

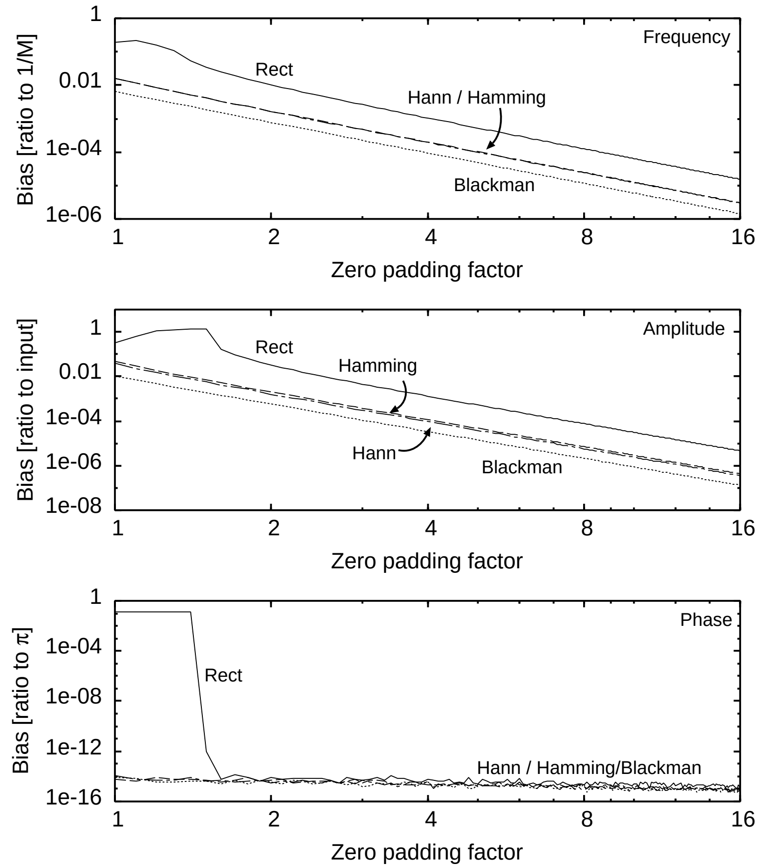

Figure 2: Maximum bias curves for Blackman-Harris windows.

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

$$\text{Bias}_\omega = |\hat{\omega}_0 - \omega_0|/(2\pi/M), \quad (2)$$

$$\text{Bias}_A = |\hat{A}_0 - A_0|/A_0, \quad (3)$$

$$\text{Bias}_\phi = |\hat{\phi}_0 - \phi_0|/\pi. \quad (4)$$

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. For a rectangular window and all members of the Blackman-Harris window family [16, 17] (including Hann, Hamming, Blackman, and so on), Ω_M is the side-lobe zero-crossing interval in radians per sample.

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. We test 8 types of windows, which are rectangular, Hann, Hamming, Blackman windows and Kaiser-Bessel windows with $\alpha = 1.5, 2.0, 2.5$, and 3.0.

3.2 Maximum bias curves

The experimentally obtained maximum bias curves are shown in Figs. 2 and 3. Some precise values of the frequency and amplitude biases are tabulated in Tables 1 and 2.

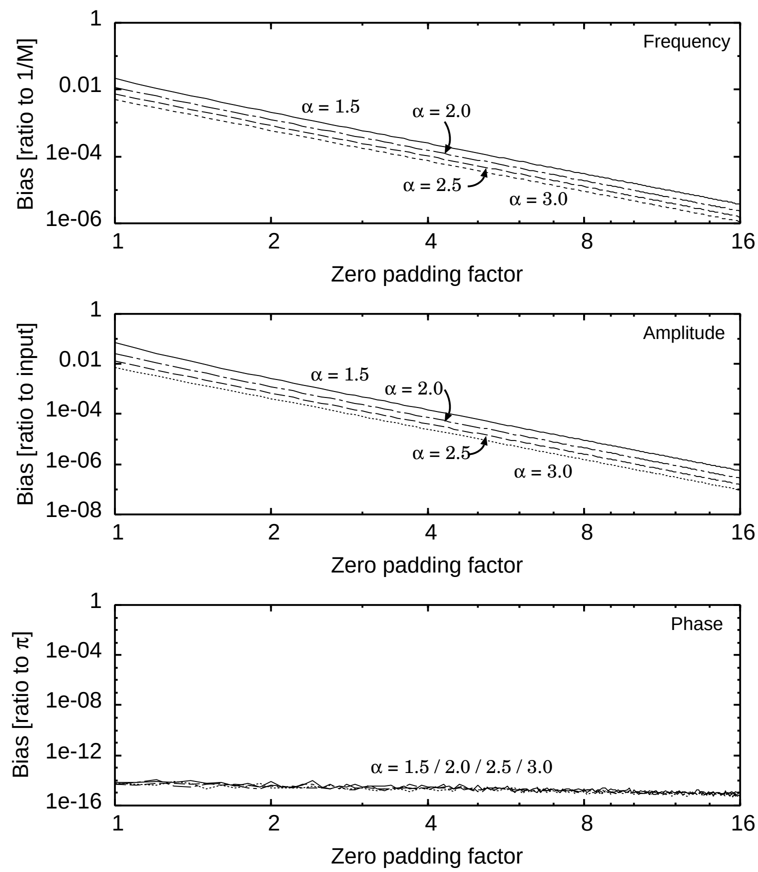

Figure 3: Maximum bias curves for Kaiser-Bessel windows.

As we mentioned before, the results for the rectangular window with a zero-padding factor below 1.5 are not reliable. Indeed, exceptionally large biases can be seen in Fig. 2. For the other cases, the maximum biases in phase are below 10^{-12}% for any types of windows (the bottom graphs in Figs. 2 and 3). This implies that the phase estimates can be regarded as completely unbiased. The result is quite reasonable because the phase response of a symmetric filter is linear, so that linear (or higher order polynomial) interpolation is exact. We can see that the maximum frequency and amplitude biases decrease approximately linear with the zero-padding factor in the log-log representation (the top and middle graphs in Figs. 2 and 3). We can also see that when zero-padding factors are the same, a window having larger main-lobe width is generally less biased. For example, the biases with a Blackman window are greater than those with a Hann or Hamming window. This can be expected since a wider main lobe is flatter at the top and therefore better approximated by a second order polynomial (i.e., the coefficients of the Taylor series expansion at the peak generally decay faster as a function of order).

3.3 Minimum Zero-Padding Factors

The minimum zero-padding factors needed for some typical error bounds are shown in Tables 3 and 4. For example, the results show that zero-padding factors of {2.1, 1.2, 1.0} are sufficient for achieving a 1% error bound when using {rectangular, Hann or Hamming, Blackman} windows, respectively. Similarly, zero-padding factors {4.1, 2.4, 1.9} ensure a 0.1% bound for these respective window types.

In order to obtain the minimum zero-padding factor for any given bias bound, it is convenient to approximate the maximum bias curves by simple analytic expressions. We see that the maximum biases are approximately linearly related to the zero-padding factors in the log-log representation. This suggests that we can approximate the dependencies by simple exponential functions. Using least-squares approximation, we obtain functions which give the minimum

Table 1: Maximum frequency biases for various zero-padding factors (normalized by $2\pi/M$).

Z_p	Rect	Hann	Hamm	Black
1	19.127%	1.600%	1.601%	0.664%
2	1.036%	0.163%	0.167%	0.077%
3	0.262%	0.047%	0.048%	0.023%
4	0.105%	0.020%	0.020%	0.010%
5	0.053%	0.010%	0.011%	0.005%

Table 2: Maximum amplitude biases for various zero-padding factors (normalized by A_0).

Z_p	Rect	Hann	Hamm	Black
1	32.119%	3.794%	4.650%	1.053%
2	3.276%	0.159%	0.200%	0.058%
3	0.457%	0.030%	0.038%	0.011%
4	0.132%	0.010%	0.012%	0.004%
5	0.052%	0.004%	0.005%	0.002%

zero-padding factor for any prescribed bias bounds as

$$Z_p \geq c_0 \text{Bias}_\omega^{\rho_0} \quad \text{(frequency)}, \tag{5}$$
$$Z_p \geq c_1 \text{Bias}_A^{\rho_1} \quad \text{(amplitude)}, \tag{6}$$

where c_0, ρ_0, c_1 and ρ_1 are window-dependent coefficients shown in Tables 5.
For example, the functions for a Hann window are

$$Z_p \geq 0.25\ \text{Bias}_\omega^{-0.33} \quad \text{(frequency)}, \tag{7}$$
$$Z_p \geq 0.41\ \text{Bias}_A^{-0.25} \quad \text{(amplitude)}. \tag{8}$$

If we are to bound the frequency-bias below 0.1Hz with a 30ms window, a zero-padding factor larger than

$$0.25 \times (0.1[\text{Hz}] \times 0.03[\text{s}])^{-0.33} \approx 1.70 \tag{9}$$

is needed.

4 Noise robustness

Rife and Boorstyn [8, 18] have established the following results:

1) The exact spectral peak estimator (obtained in principle when the zero-padding factor approaches infinity) is the Maximum Likelihood (ML) estimator for the parameters of a single complex sinusoid in Additive White Gaussian Noise (AWGN), when a rectangular window is used. (ML estimators asymptotically achieve the Cramer-Rao Bound (CRB) for the estimation variance.)

2) Use of non-rectangular windows can greatly reduce the interference bias for multiple sinusoids, while slightly increasing the estimation variance.

Since the QIFFT is an approximate spectral peak estimator, it classifies as an approximate ML estimator for sinusoidal parameters. We now wish to evaluate the QIFFT method as an approximate ML estimator by studying the effects of bias due to quadratic interpolation. To accomplish this, we will 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

Table 3: Minimum zero-padding factors for frequency bias bounds.

Blackman-Harris				
Max Bias	Rect	Hann	Hamm	Black
1.00%	2.1	1.2	1.2	1.0
0.50%	2.5	1.5	1.5	1.1
0.10%	4.1	2.4	2.4	1.9
0.01%	8.7	5.0	5.1	4.0
Kaiser-Bessel				
Max Bias	$\alpha = 1.5$	$\alpha = 2.0$	$\alpha = 2.5$	$\alpha = 3.0$
1.00%	1.3	1.1	1.0	1.0
0.50%	1.6	1.3	1.2	1.1
0.10%	2.6	2.2	1.9	1.7
0.01%	5.4	4.7	4.1	3.7

Table 4: Minimum zero-padding factors for amplitude bias bounds.

Blackman-Harris				
Max Bias	Rect	Hann	Hamm	Black
1.00%	2.6	1.4	1.4	1.1
0.50%	3.0	1.6	1.7	1.2
0.10%	4.3	2.3	2.4	1.8
0.01%	7.5	4.0	4.2	3.1
Kaiser-Bessel				
Max Bias	$\alpha = 1.5$	$\alpha = 2.0$	$\alpha = 2.5$	$\alpha = 3.0$
1.00%	1.5	1.3	1.1	1.0
0.50%	1.8	1.5	1.3	1.1
0.10%	2.6	2.2	1.9	1.7
0.01%	4.5	3.8	3.3	2.9

chosen to be AWGN. Specifically, we use FFT sizes of 2048 and 4096, and zero-padding factors of 2.5 and 5.0 — values commonly used with the QIFFT method when measuring audio spectral peaks. The sinusoidal parameters are randomly given, as in the numerical simulations in the previous section.

The results are shown in Figs. 4 to 7. As references, we also show the CRBs and audible error limits which we used 0.1% in frequency and 0.1dB ($\approx$ 1.2%) in amplitude.

We can see from the figures that all the results reveal roughly the same trend. At low S/N ratios, 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 [8]. At moderate S/N ratios, which are most important in audio processing, we find that the QIFFT works essentially as well as the ML estimator. Note that for non-rectangular windows, the errors lie a bit above the CRBs, as expected. At high S/N ratios, the errors are dominated by the biases. We can control this bias as desired using the criteria described in the previous section. It should perhaps be noted that a 0.1% peak-frequency bias cannot be heard in most audio applications.

5 Summary

In this paper, we clarified the relation between estimation bias and zero padding, and provided a design criterion for achieving a prescribed error bound. The results show, for example, that zero-padding factors of {4.1, 2.4, 1.9} are sufficient for achieving a 0.1% frequency-error bound when using {rectangular, Hann or Hamming, Blackman} windows, respectively. We also confirmed that the QIFFT works as an approximate ML estimator within a middle S/N

Table 5: Coefficients for approximate bias curves

Window	c_0	ρ_0	c_1	ρ_1
Rect	0.4467	–0.3218	0.8560	–0.2366
Hann	0.2436	–0.3288	0.4149	–0.2456
Hamming	0.2456	–0.3282	0.4381	–0.2451
Blackman	0.1868	–0.3307	0.3156	–0.2475
KB (1.5)*	0.2672	–0.3270	0.4747	–0.2437
KB (2.0)	0.2231	–0.3292	0.3881	–0.2461
KB (2.5)	0.1932	–0.3304	0.3312	–0.2472
KB (3.0)	0.1718	–0.3310	0.2907	–0.2479

(*) KB(α): Kaiser-Bessel window with the indicated α value.

ratio range. This range can be extended arbitrarily high through the use of more zero padding, e.g., to the limits of human error perception.

For further extension to reduce the interpolation bias, a simple bias-correction method utilizing cubic and quadratic functions has been proposed [21].

Other important causes of bias are interference from nearby sinusoidal components and time-variation in the underlying signal parameters. The former generally restricts a lower limit on the window length, whereas the latter impose an upper limit. Criteria for these biases are discussed in our following papers [19], [20].

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 Trans. Acoust., Speech, Signal Processing, 34, 4, 744/754 (1986).

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

[4] K. N. Hamdy, M.Ali and A. H. Tewfik: "Low Bit Rate High Quality Audio Coding with Combined Harmonic and Wavelet Representations," Proc. IEEE ICASSP'96, 1045/1048 (1996).

[5] S. N. Levine and J. O. Smith III: "A Switched Parametric & Transform Audio Coder," Proc. IEEE ICASSP'99, 985/988 (1999).

[6] T. S. Verma and T. H. Y. Meng: "Sinusoidal Modeling using Frame-Based Perceptually Weighted Matching Pursuits," Proc. IEEE ICASSP'99, 981/984 (1999).

[7] T. Virtanen and A Klapuri: "Separation of Harmonic Sound Sources Using Sinusoidal Modeling," Proc. IEEE ICASSP'00, 765/768 (2000).

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

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

[10] 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).

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

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

[13] 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).

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

[15] J. C. Brown and M. S. Puckette: "A High Resolution Fundamental Frequency Determination Based on Phase Changes of the Fourier Transform," J. Acoust. Soc. Am., 94, 2, 662/667 (1993).

[16] F. J. Harris: "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," Proc. of the IEEE, 66, 1, 51/83 (1978).

[17] A. H. Nuttall: "Some Windows with Very Good Sidelobe Behavior," IEEE Trans. Acoust., Speech, Signal Processing, 29, 1 84/91 (1981).

[18] D. C. Rife and R. R. Boorstyn: "Multiple Tone Parameter Estimation from Discrete-Time Observations," Bell System Technical Journal, 55, 9, 1389/1410 (1976).

[19] 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, August, (2004).

[20] 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, August, (2004).

[21] M. Abe and J. O. Smith III: "Correcting Bias in a Sinusoidal Parameter Estimator based on Quadratic Interpolation of FFT Magnitude Peaks," Technical Report STAN-M-117, Dept. of Music, Stanford University, August, (2004).

[22] 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, August, (2004).

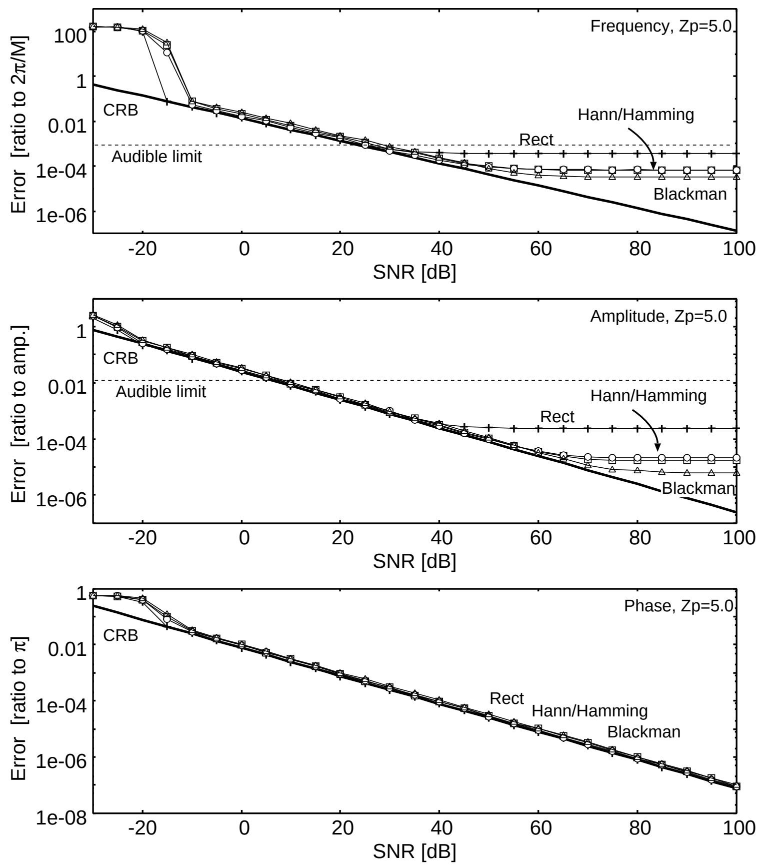

Figure 4: RMS errors for a sinusoid with noise and the Cramer-Rao bounds (Blackman-Harris family, $N = 4096$, $Z_p = 5.0$): (a) frequency, (b) amplitude, (c) phase.

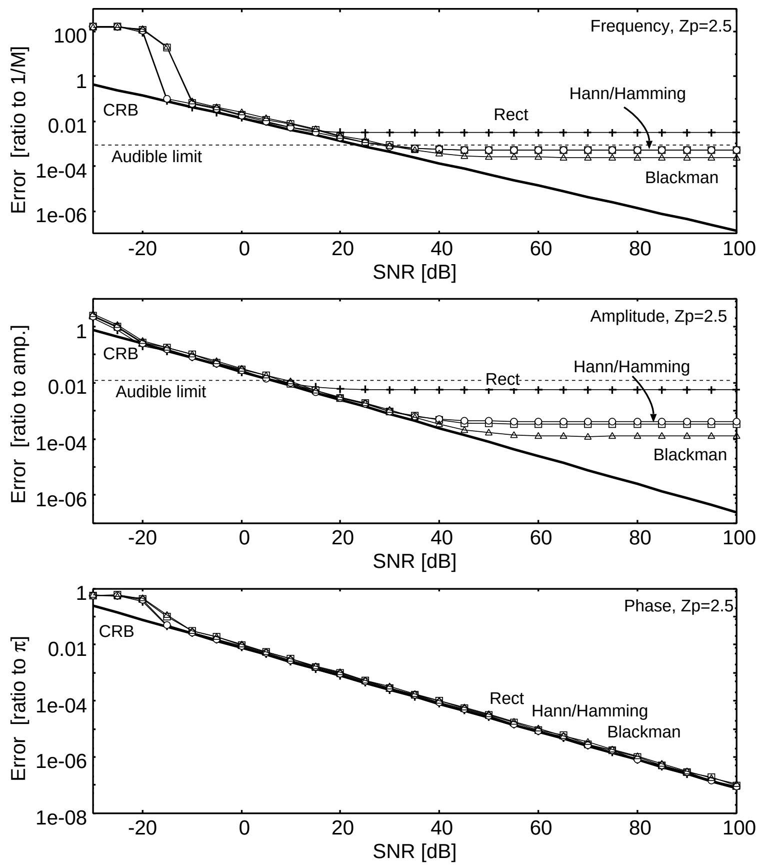

Figure 5: RMS errors for a sinusoid with noise and the Cramer-Rao bounds (Blackman-Harris family, $N = 2048$, $Z_p = 2.5$): (a) frequency, (b) amplitude, (c) phase.

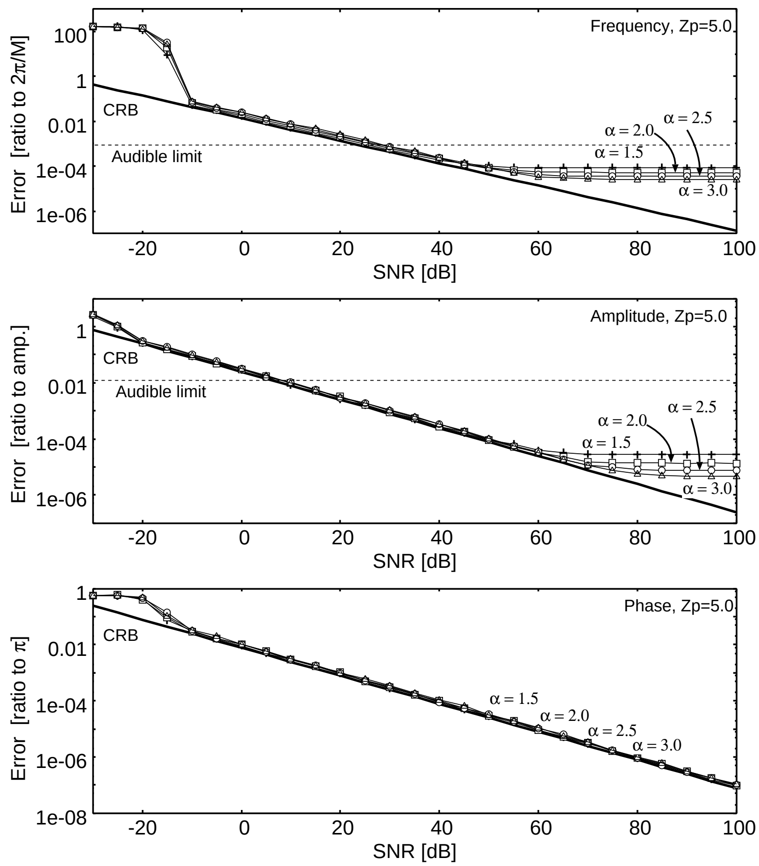

Figure 6: RMS errors for a sinusoid with AWGN (Kaiser-Bessel windows, $N = 4096$, $Z_p = 5.0$): (a) frequency, (b) amplitude, (c) phase.

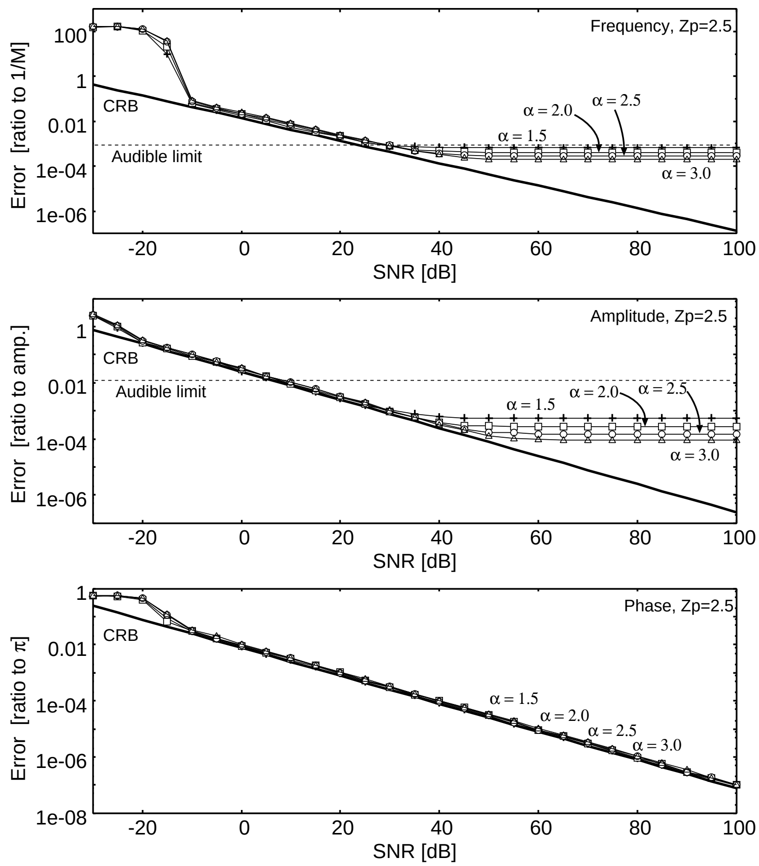

Figure 7: RMS errors for a sinusoid with AWGN (Kaiser-Bessel windows, $N = 2048$, $Z_p = 2.5$): (a) frequency, (b) amplitude, (c) phase.