CENTER FOR COMPUTER RESEARCH IN MUSIC AND ACOUSTICS
DEPARTMENT OF MUSIC, STANFORD UNIVERSITY
REPORT NO. STAN-M-115

Design Criteria for the Quadratically Interpolated FFT Method (II): Bias due to Interfering Components

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 (FFT) has been widely used for sinusoidal parameter estimation in audio applications. In this series of papers, we investigate the estimation errors associated with various choices of analysis parameters, and provide precise criteria for designing the estimator. In this second paper, we focus on the bias caused by nearby interfering spectral components. We derive precise conditions for estimating the parameters of closely spaced sinusoidal components, confirm the results by numerical experiments, and determine the minimum frequency spacing needed for reliable estimation. The results show, for example, giving a sampling rate of F_s and a length M Hann window, a frequency separation of $2.28F_s/M$ or greater is required, and in this case, the frequency estimation error is bounded by $0.042F_s/M$ for a zero-padding factor of 5. Noise rubustness of the estimator is also investigated. We confirm that the estimator works as well as a ML estimator within a practically important S/N ratio range.

1 Introduction

Among various approximate maximum likelihood (ML) sinusoidal parameter estimators [1, 2, 3, 4], quadratic interpolation of magnitude peaks in a zero-padded Fast Fourier Transform (FFT) [5] has been widely used due to its simplicity and accuracy. Although it works as well as a ML estimator for well-separated sinusoids with a large zero-padding factor, its accuracy in practice is restricted by the choice of its design parameters, such as window type, FFT length and zero-padding factor. In this series of papers, we theoretically predict and numerically measure the estimation error bias associated with the choice of the design parameters, and precise design criteria for the estimator are defined. In the first paper [9], we investigate the bias caused by quadratic interpolation and determine a criterion for choosing a zero-padding factor.

In this second paper, we focus on the bias caused by interference from nearby components. Typically, a larger frequency difference yields more accurate estimates. However, the relation is highly nonlinear due to the main and side lobe structure of the window used. In particular, when the components become closer than a certain frequency separation, the bias sharply rises and the estimates become quite unreliable [6]. An often-used criterion to assure reliable estimation is to set the window length so that the frequency difference is more than one main-lobe width. This is sometimes referred to as "four periods of the lowest frequency included in the analyzed signal" [7] for Hann or Hamming windows. Although it works well as a general criterion, we found that this frequency separation is overly conservative for peak parameter estimation. If we wish to minimize (or find a good compromise for) the window length from the perspective of time resolution, we may need to know more precisely the relation between bias and window length. In Section 3, we derive precise conditions for estimating the parameters of closely spaced sinusoidal components, and the results are confirmed by numerical experiments.

We also discuss noise robustness of the estimator in Section 4. 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 the estimator for noisy signals at various S/N ratios.

The following symbols are used in this paper:

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

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 a zero-padded FFT (points in Fig. 1).

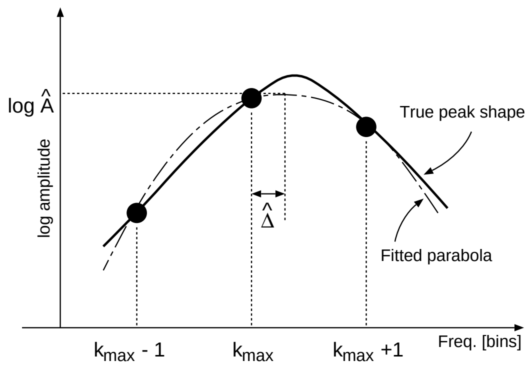

Figure 1: Quadratic interpolation of spectral peak

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[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.

3 Minimum Frequency Separation

3.1 Separation by One Main Lobe

An intuitively complete separation of sinusoidal peaks occurs when they are separated by one main-lobe width, as depicted in Fig. 2(a). Here, a main-lobe width is defined as the frequency difference between the first zero-crossings about the main lobe.[2] We denote this frequency separation as $\Delta\omega_m$. Actual values of this criterion for some often-used windows are shown in Table 1. Note that to obtain numerical values independent of the window length and sampling rate, we use a "window-normalized frequency" which is given by normalized radian frequency divided by 2π over the window length M.

3.2 More Precise Separation Criteria

Because the QIFFT only uses the maximum peak and its two neighbors in the main lobe, the main-lobe separation criterion is overly conservative. We may consider more precise requirements as

1. each peak can be detected (i.e., there is a local maximum), and
2. the peak is not severely distorted by other spectral components.

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

[2]This definition of main-lobe width is not appropriate for all window types, such as those considered in [7], but it is a convenient definition for the window types considered in this paper.

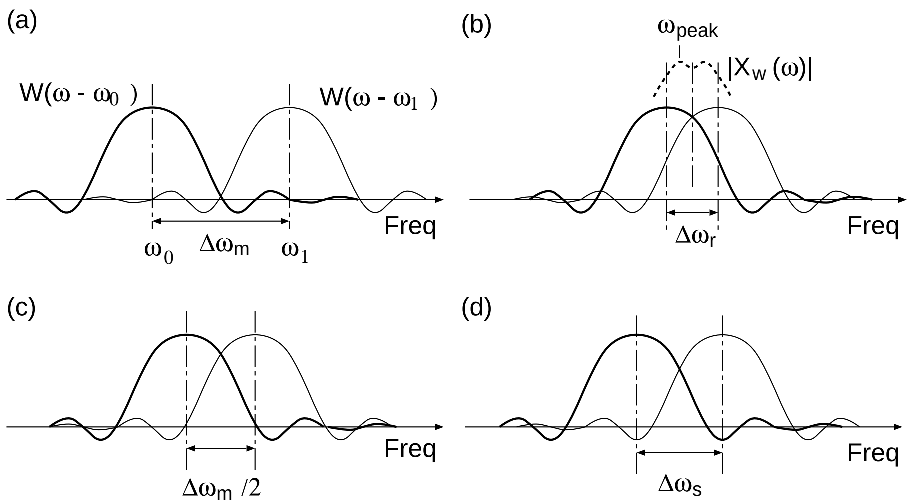

Figure 2: Interference from nearby components of various frequency separations: (a) the two main lobes are separated by one main-lobe width. (b) the two peaks are just resolved, (c) the first side-lobe zeros are placed at the main-lobe peaks, (d) the first side-lobe peaks are placed at the main-lobe peaks (more precise criterion).

To discuss these requirements quantitatively, let us consider a signal composed of two sinusoids, as

$$x(n) = A_0 e^{j(\omega_0 n + \phi_0)} + A_1 e^{j(\omega_1 n + \phi_1)}, \tag{1}$$

where $\{A_0, \omega_0, \phi_0\}$ and $\{A_1, \omega_1, \phi_1\}$ denote the amplitudes, frequencies and phases of the target and interfering components, respectively. Note that since the QIFFT method detects the largest peak first and removes it by linear spectral subtraction before processing the rest of the components, the worst-case distortion is generally from a nearby component having almost the same amplitude.

Denote the windowed Discrete-Time Fourier Transform (DTFT) of the signal $x(n)$ by

$$X_w(\omega) = \sum_{n=0}^{M-1} w(n)x(n)e^{-j\omega n}, \tag{2}$$

where $w(n)$ denotes a window function.

To satisfy the first requirement, the frequency difference between the two components should be large enough so that the spectral value at the center of the two components is below the two peaks (Fig. 2(b)), i.e.,

$$|X_w(\omega)|_{\omega=\omega_{\text{peak}}} > |X_w(\omega)|_{\omega=(\omega_0+\omega_1)/2} \ . \tag{3}$$

Since the exact peak positions in the spectrum cannot be easily determined, we approximate ω_{peak} by the true frequency of the target component, and require

$$|X_w(\omega)|_{\omega=\omega_0} > |X_w(\omega)|_{\omega=(\omega_0+\omega_1)/2} \ . \tag{4}$$

Using Eqs. (1) and (2) and considering the worst-case phases, we obtain the following inequality:

$$Y(\omega) \triangleq W(0) - 2W(\Delta\omega/2) - |W(\Delta\omega)| > 0, \tag{5}$$

where $W(\omega)$ is the DTFT of the window, and $\Delta\omega \triangleq \omega_1 - \omega_0$ denotes the frequency separation. From this inequality, the minimum frequency separation for the first requirement can be described as

$$\Delta\omega_r \triangleq \Delta\omega \text{ s.t. } \mathrm{Y}(\Delta\omega) = 0. \tag{6}$$

Numerically obtained solutions for some windows are tabulated in Table 2.

Table 1: Minimum allowable frequency separation (MAFS) based on the main-lobe separation criterion (in units of window-normalized frequency).

	Rect	Hann	Hamm	Black
$\Delta\omega_m$	2.00	4.00	4.00	6.00
	KB(1.5)	KB(2.0)	KB(2.5)	KB(3.0)
$\Delta\omega_m$	3.61	4.48	5.39	6.33

Table 2: Minimum frequency separation for the peaks to be resolved.

	Rect	Hann	Hamm	Black
$\Delta\omega_r$	1.37	2.00	1.84	2.35
	KB(1.5)	KB(2.0)	KB(2.5)	KB(3.0)
$\Delta\omega_r$	1.78	2.03	2.25	2.45

Table 3: Minimum frequency separation for the peaks not to be severely affected by the interfering components.

	Rect	Hann	Hamm	Black
$\Delta\omega_s$	1.44	2.37	2.22	3.03
	KB(1.5)	KB(2.0)	KB(2.5)	KB(3.0)
$\Delta\omega_s$	2.08	2.46	2.89	3.33

For the second requirement, it intuitively seems sufficient that the two components be separated by *half* a main-lobe width, since in this case, as shown in Fig. 2(c), the two main lobes peaks are well separated. Furthermore, since each main-lobe peak sits atop a zero-crossing of the other's window transform, the two windowed sinusoids are *orthogonal* under the rectangular window at this frequency separation. This "minimum orthogonal frequency separation" is exactly half of the full main-lobe separation spacing $\Delta\omega_m$, i.e., the main lobes overlap by 50%.

It turns out, however, that the minimum orthogonal frequency separation is not optimal for the QIFFT method. The reason is that the frequency estimate depends not only on spectral magnitude but also on the magnitude gradient at ω_0. As can be seen in Fig. 2(c), the steep slope continued from the main-lobe of the interfering component effectively tilts the peak of the target, causing a bias.

A better theoretical prediction can be derived from the first zero of the gradient of the window magnitude-spectrum away from the peak. This condition is depicted in Fig. 2(d). We denote this frequency separation as

$$\Delta\omega_s \triangleq \text{Min}(\omega) \text{ s.t. } \frac{d}{d\omega}W(\omega) = 0 \text{ and } \omega > 0. \tag{7}$$

Actual values for some often-used windows are tabulated in Table 3.

Taking the larger of Eqs. (6) and (7), we obtain the predicted minimum allowable frequency separation (MAFS) as

$$\Delta\omega_{\min} = \text{Max}(\Delta\omega_r, \Delta\omega_s). \tag{8}$$

Comparing Tables 2 and 3, we can confirm that the second requirement is always greater than the first requirement. That is, sufficiently suppressing interference always yields sufficiently separated spectral peaks.

In practice, since the frequency axis is sampled in an FFT, we need to consider the worst-case sampling condition, especially when a small zero-padding factor is used. Since the ambiguity of the frequencies of both the target and interfering components is within 0.5 FFT bins, we can refine Eq. (8) as

$$\Delta\omega_{\min} = \Delta\omega_s + \frac{2\pi}{N} = \Delta\omega_s + \frac{1}{Z_p}(2\pi/M). \tag{9}$$

Table 4: MAFS and maximum error bias.

window		Theory	Experiment			
	Z_p	MAFS	MAFS	Bias (%)		
				Freq	Amp	Pha
Rect	2.0	1.94	1.90	15.9	24.6	5.37
	3.5	1.73	1.39	16.3	23.2	7.14
	5.0	1.64	1.38	16.5	22.1	6.95
Hann	2.0	2.87	2.38	3.89	4.34	1.26
	3.5	2.66	2.30	4.09	2.88	0.91
	5.0	2.57	2.28	4.15	2.74	0.87
Hamm	2.0	2.72	2.35	1.37	2.19	0.60
	3.5	2.51	2.20	1.42	0.87	0.27
	5.0	2.42	2.18	1.45	0.76	0.24
Black	2.0	3.53	3.23	0.39	0.28	0.07
	3.5	3.32	3.05	0.39	0.13	0.04
	5.0	3.23	3.00	0.40	0.13	0.04
KB	2.0	2.58	2.15	3.07	4.49	1.20
$\alpha = 1.5$	3.5	2.37	2.05	3.33	2.16	0.67
	5.0	2.28	2.02	3.40	1.93	0.62
KB	2.0	2.96	2.58	1.14	1.55	0.44
$\alpha = 2.0$	3.5	2.75	2.45	1.27	0.67	0.20
	5.0	2.66	2.43	1.34	0.56	0.18
KB	2.0	3.39	3.07	0.39	0.48	0.13
$\alpha = 2.5$	3.5	3.18	2.90	0.45	0.20	0.06
	5.0	3.09	2.85	0.47	0.16	0.05
KB	2.0	3.83	3.60	0.15	0.14	0.03
$\alpha = 3.0$	3.5	3.62	3.35	0.15	0.07	0.02
	5.0	3.53	3.30	0.16	0.05	0.02

Actual values for some often-used windows with zero-padding factors 2.0, 3.5 and 5.0 are shown in Table 4.

Note that we used sufficient (but not always necessary) conditions to derive the predicted MAFS in the above discussion. Since underestimation of the MAFS may result in a severe estimation error, we believe our slightly conservative prediction is more useful than a tighter but overly optimistic MAFS.

3.3 Simulation Results

This section presents experimentally determined curves showing peak parameter bias as a function of frequency separation between the target peak and an interfering peak.

3.3.1 Test Conditions

The test signal is composed of two complex sinusoids, as

$$x(n) = A_0 e^{j(\omega_0 n + \phi_0)} + A_1 e^{j(\omega_1 n + \phi_1)}. \tag{10}$$

We prepare 1024 signals by setting ω_0 to a uniformly distributed random value in $[0, \pi]$, $\omega_1 = \omega_0 + \Delta\omega$, $A_0 = A_1 = 1.0$, and ϕ_0 and ϕ_1 to uniformly distributed random values in $[-\pi, \pi]$. We estimate the frequency, amplitude and phase of the target sinusoid $(\hat{A}_0, \hat{\omega}_0, \hat{\phi}_0)$ using the QIFFT method and evaluate the biases, as

$$\begin{aligned} \text{Bias}_\omega &= |\hat{\omega}_0 - \omega_0|/(2\pi/M), & (11)\\ \text{Bias}_A &= |\hat{A}_0 - A_0|/A_0, & (12)\\ \text{Bias}_\phi &= |\hat{\phi}_0 - \phi_0|/\pi. & (13) \end{aligned}$$

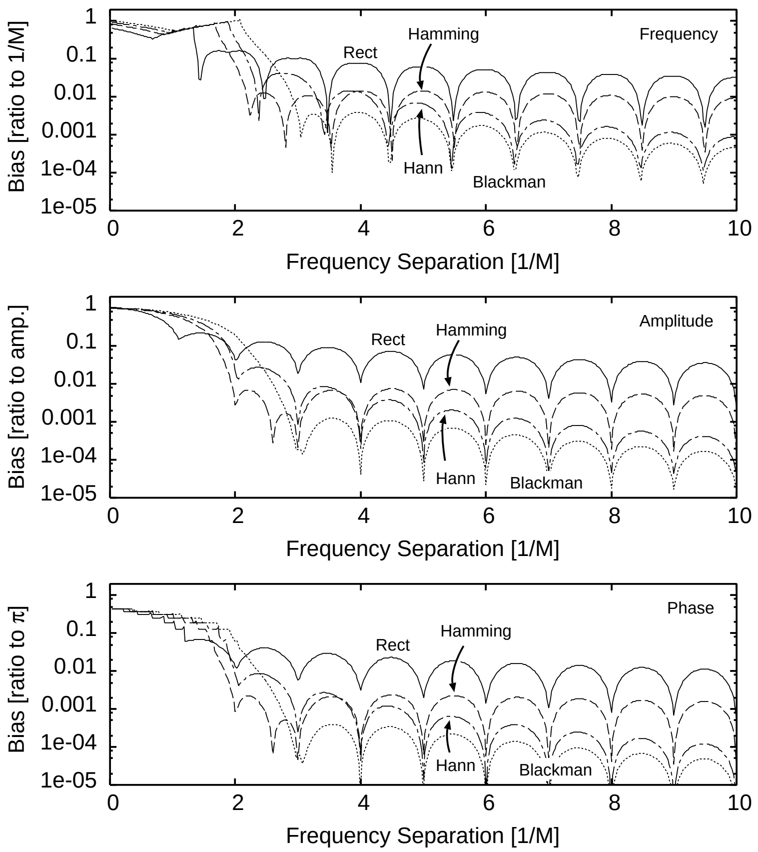

Figure 3: Bias with windows in the Blackman-Harris family ($Z_p = 5.0$): frequency (top), amplitude (middle) and phase (bottom).

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 use 6 FFT sizes from 256 to 8192. By sweeping the frequency difference $\Delta\omega$ from 0.025 to 10.0 with the step of 0.025 (in window-normalized frequency), and by taking the maximum errors out of the 6144 test sets (1024 signals $\times$ 6 FFT sizes) for each frequency separation, we obtain the maximum bias curves. The zero-padding factors 2.0, 3.5 and 5.0 are tested.

3.3.2 Results

The obtained maximum bias curves for the zero-padding factor of 5.0 are shown in Fig. 3 and 5. The curves have many local maxima and minima due to the side-lobe structure of the interfering windowed sinusoid. Note that the positions of the local maxima and minima in the frequency bias are different from those in the amplitude and phase biases. This is because the amplitude and phase estimates are mainly determined by the spectral values themselves, whereas the frequency estimate is highly dependent on the slope of the magnitude spectrum at the target frequency. Note, in particular, the amplitude and phase bias exhibit local minima at frequency-spacings for which a zero-crossing of the interfering side lobes is aligned with the target peak, as shown in Fig. 2(c), while the peak-frequency bias, on the other hand, is lower when an interfering side-lobe extremum is aligned with the target peak, as shown in Fig. 2(d).

Since the exact frequency spacing between a target sinusoid and an interfering component is generally not known, it is useful to filter out each local minimum based on the local maximum to its left. I.e, by taking the maximum value to the right at each frequency separation, we obtain the filtered results shown in Figs. 4 and 6. These curves indicate the maximum biases for all frequency separations larger than a given horizontal axis value.

In the small frequency separation range, the frequency biases are about 100% for all the windows. In this range, the main lobes of the two sinusoids overlap severely so that the peaks are not resolved. In the moderate frequency

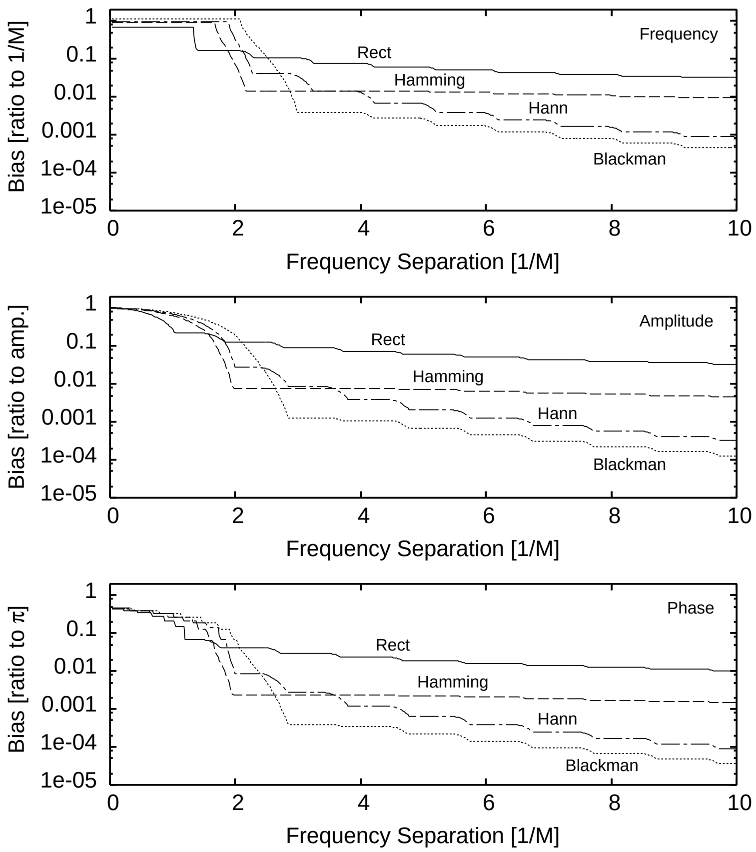

Figure 4: Filtered bias with windows in Blackman-Harris family ($Z_p = 5.0$): frequency (top), amplitude (middle) and phase (bottom).

separation range, the bias curves fall off sharply. In this range, though the peaks are resolved into two local peaks, the target peak is still severely degraded by the main lobe of the interfering component. The steep slope reflects the shape of the main lobe.

Above certain frequency separations, the bias envelopes become fairly flat as a function of additional frequency separation. This condition occurs when the main lobes are sufficiently separated so that the target peak is only disturbed by side lobes of the interfering component, and the biases become determined by the window side-lobe level. The small negative slope in the flat region is a function of the side-lobe roll-off rate for the window used. We may consider the "knee" point at which flatness begins as the measured MAFS for each window. For the amplitude and phase biases, similar phenomena are observed.

The MAFSs and the maximum biases for the zero-padding factors of 2.0, 3.5 and 5.0 are shown in Table 4 on page 6. As predicted from theory in Eq. (8), the MAFS slightly depends on the zero-padding factors. A larger zero-padding factor decreases the MAFS in general, but the improvement is quite small for zero-padding factors greater than 3.

Comparing to the theoretically predicted values, we can confirm that all the experimentally obtained MAFSs are smaller to a varying extent. This is because the theoretical predictions are derived from sufficient conditions, as mentioned in the previous section. Though the theoretical predictions overestimate the MAFSs, they are tighter than the main-lobe-separation criterion shown in Table 1, and they never underestimate the experimentally obtained MAFSs.

3.4 Minimum Window Length

The minimum window length needed for a given frequency separation can be calculated as

$$M \geq 2\pi(\text{MAFS})/\Delta\omega, \tag{14}$$

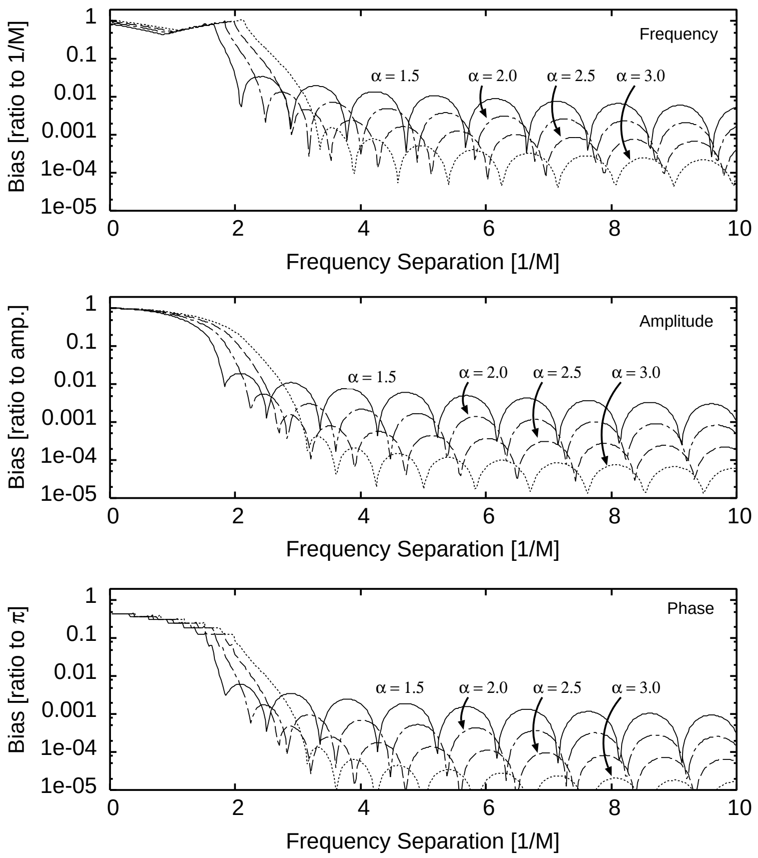

Figure 5: Bias with Kaiser-Bessel windows (Z_p = 5.0): frequency (top), amplitude (middle) and phase (bottom).

where the MAFS can be obtained from Figs. 4, 6, Table 4, or Eq. (9).

For example, if we are to use a Hann window with a zero-padding factor of 5.0 and an expected minimum frequency separation of 50 Hz, the window length should be at least 2.28/50 = 45.6 ms.

4 Noise Robustness

Since the QIFFT is an approximate spectral peak estimator, it classifies as an approximate ML estimator for sinusoidal parameters. In this section, in order to evaluate the QIFFT method 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 a fixed FFT size of 4096, and zero-padding factors of 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 frequency separation is set to the MAFS of each window.

The results are shown in Figs.7 and 8. As references, we also show the CRBs for a single sinusoid 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 [1]. 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 sections.

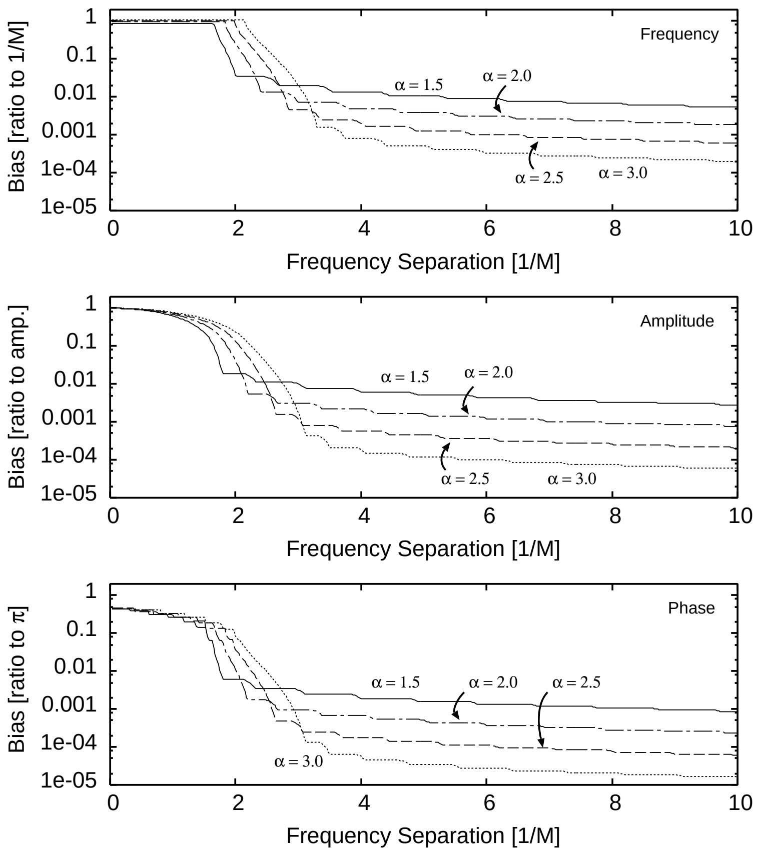

Figure 6: Filtered bias with Kaiser-Bessel windows (Z_p = 5.0): frequency (top), amplitude (middle) and phase (bottom).

5 Discussion and Summary

In this paper, we theoretically and numerically investigated the relation between estimation bias and frequency spacing of multiple components, and provided a criterion to reliably estimate parameters of multiple sinusoids. The results show, for example, that given a sampling rate of F_s and a length M Hann window, a frequency separation of $2.28F_s/M$ or greater is required, and in this case, the frequency estimation error is bounded by $0.042F_s/M$ for a zero-padding factor of 5. We also confirmed that the QIFFT works as an approximate ML estimator within a middle S/N ratio range. This range can be controled through the use of longer windows.

It may be important to mention that, except slight dependency on zero-padding factors, the results in this paper can be applied to other ML estimators based on sinusoidal peak detection.

For further extension, some iterative methods to reduce the interference bias have been proposed [7, 8]. Combining the QIFFT with these methods may reduce the bias in exchange for some computational complexity.

Another important cause of bias is time-variation in the underlying signal parameters. Amplitude and/or frequency modulations in sinusoidal components generally impose an upper limit on the window length. Criteria for this bias are discussed in our following paper [10].

References

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

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

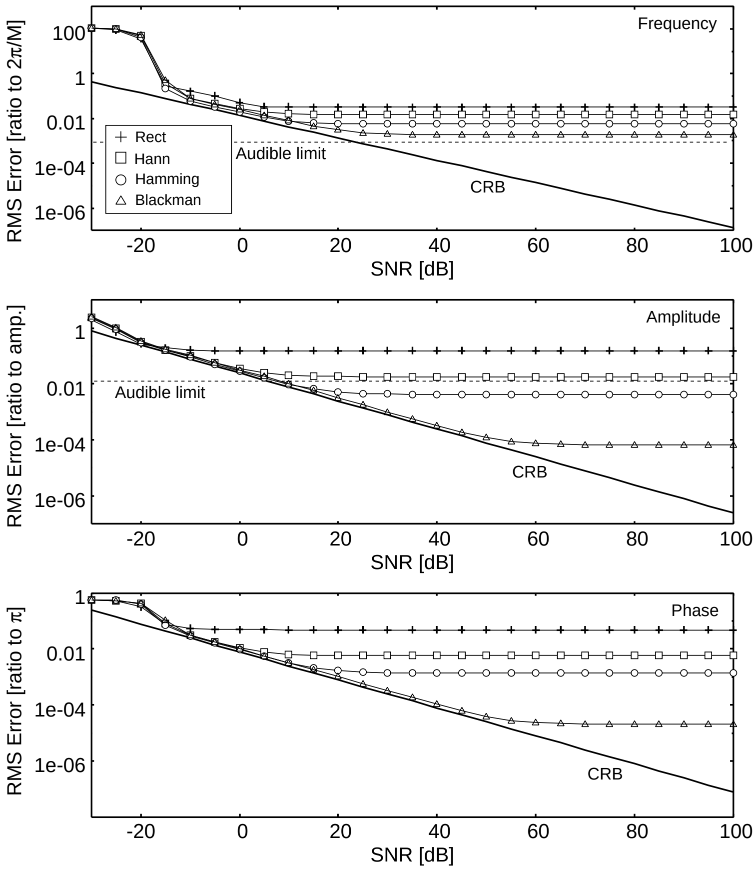

Figure 7: RMS errors for two sinusoids with AWGN, shown with the CRBs for a single sinusoid (Blackman-Harris family): frequency (top), amplitude (middle) and phase (bottom).

[3] 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).

[4] 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).

[5] 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.

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

[7] 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).

[8] J. S. Marques and L. B. Almeida: "Frequency-Varying Sinusoidal Modeling of Speech," IEEE Trans. Acoust., Speech, Signal Processing, 37, 5, 763/765 (1989).

[9] 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).

[10] 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).

[11] 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, October, (2004).

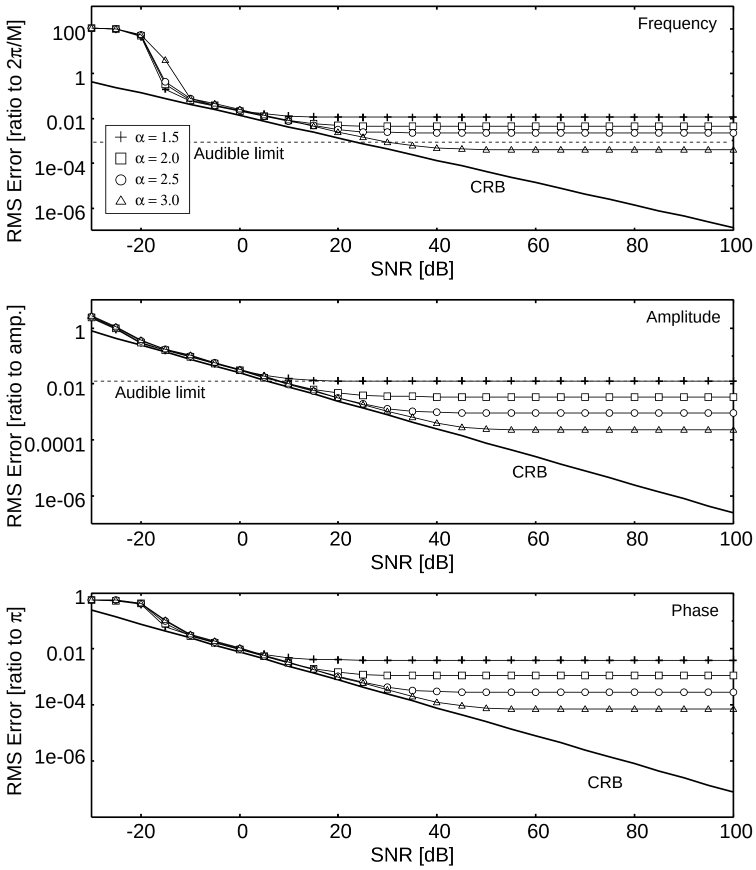

Figure 8: RMS errors for two sinusoids with AWGN, shown with the Cramer-Rao bounds for a single sinusoid (Kaiser-Bessel windows): frequency (top), amplitude (middle) and phase (bottom).

[12] 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).