Audio Engineering Society

Convention Paper

Presented at the 117th Convention
2004 October 28–31 San Francisco, CA, USA

Design Criteria for Simple Sinusoidal Parameter Estimation based on Quadratic Interpolation of FFT Magnitude Peaks

Mototsugu Abe[1], and Julius O. Smith III[2]

[1]*SONY Corporation, 6-7-35 Kita-Shinagawa, Shinagawa-ku Tokyo 141-0001, Japan*

[2]*CCRMA, Music Department, Stanford University, CA 94305. U.S.A.*

Correspondence should be addressed to Mototsugu Abe (abe@av.crl.sony.co.jp)

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. In this paper, we theoretically predict and numerically measure the errors associated with various choices of analysis parameters, and provide precise criteria for designing the estimator. In particular, we determine 1) the minimum zero-padding factor needed for a given error bound in quadratic peak interpolation, and 2) the minimum allowable frequency separation for a given window length, for various window types.

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 may wish to minimize zero

padding for computational efficiency, and we may need to contend with interference from nearby signal components. The zero-padding factor determines the maximum error bias caused by the quadratic interpolation. The window type and length determine the maximum bias caused by the interfering components, as well as the minimum frequency separation between sinusoids for reliable estimation. 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 paper, we theoretically predict and numerically measure the estimation error bias associated with the choice of the window type, length, and zero-padding factor, and precise design criteria for the estimator are defined.

Firstly, we focus on the bias caused by quadratic interpolation. This 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.

Secondly, we focus on the bias caused by nearby interfering spectral 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 [16]. 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" [17] 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 4, we derive precise conditions for estimating the parameters of closely spaced sinusoidal components, and the results are confirmed by numerical experiments.

Finally, we discuss noise robustness of the estimator in Section 5. 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.

Based on these results, the design of a peak parameter estimator becomes quite trivial in terms of the maximum allowable error bound and the minimum frequency spacing of sinusoids to be estimated.

The following symbols are used in this paper:

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

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

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

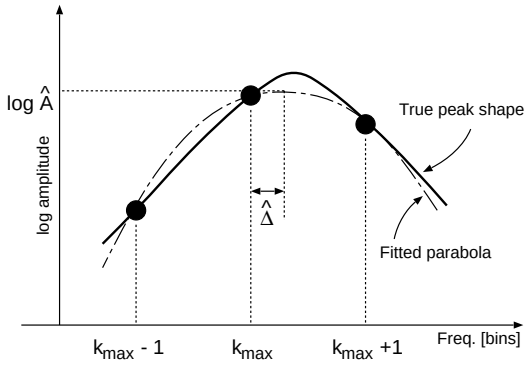

Fig. 1: Quadratic interpolation of spectral peak.

2. THE QIFFT METHOD FOR SINUSOIDAL PARAMETER ESTIMATION

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.

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

This paper 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

As shown in Fig. 1, the estimation error bias caused by quadratic interpolation 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. Though theoretical discussion is possible to some extent, the results are generally nonlinear and complicated. It is more convenient to utilize numerical simulation for this problem.

3.1. Conditions for Numerical Experiments

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

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

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

$$\begin{aligned} \text{Bias}_\omega &= |\hat{\omega}_0 - \omega_0|/(2\pi/M), & (2)\\ \text{Bias}_A &= |\hat{A}_0 - A_0|/A_0, & (3)\\ \text{Bias}_\phi &= |\hat{\phi}_0 - \phi_0|/\pi. & (4) \end{aligned}$$

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. For a rectangular window and all members of the Blackman-Harris window family [18, 19] (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 $N =$

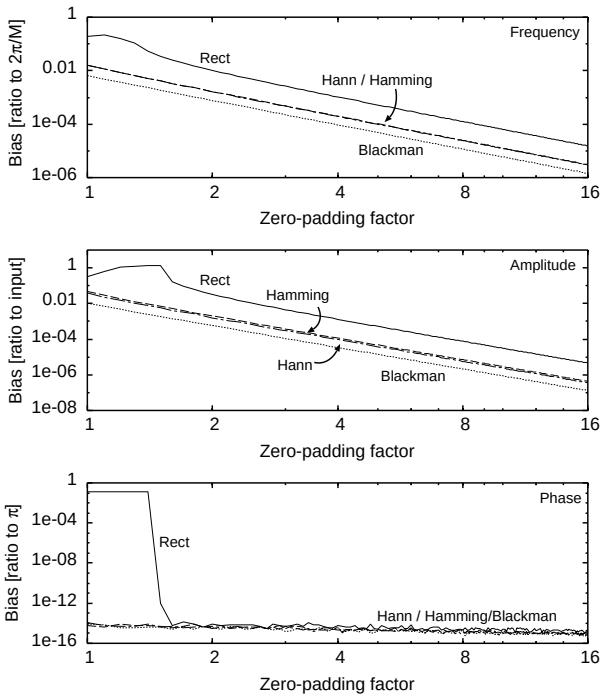

Fig. 2: Maximum-bias curves for Blackman-Harris windows: frequency (top), amplitude (middle) and phase (bottom).

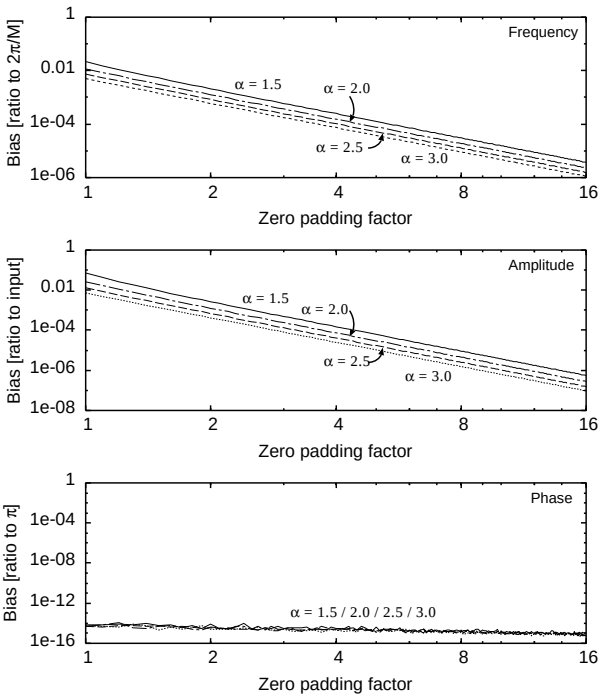

Fig. 3: Maximum-bias curves for Kaiser-Bessel windows: frequency (top), amplitude (middle) and phase (bottom).

$\{64, 128, ..., 8192\}$, 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.

As we mentioned before, the results for the rectangular windows 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 type of window (the bottom figures in Figs. 2 and 3). This implies that the phase estimates in practice can be regarded as completely unbiased. This 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 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 figures in Figs. 2 and 3). We can also see that when zero-padding factors are the same, a window having a larger main-lobe width is generally less biased. For example, the biases with a Blackman window are less 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

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%

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 zero-padding factor for any prescribed bias bounds as

$$Z_p \geq c_0 \mathrm{Bias}_\omega^{\rho_0} \quad \text{(frequency)}, \tag{5}$$

$$Z_p \geq c_1 \mathrm{Bias}_A^{\rho_1} \quad \text{(amplitude)}, \tag{6}$$

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

$$Z_p \geq 0.25\ \mathrm{Bias}_\omega^{-0.33} \quad \text{(frequency)}, \tag{7}$$

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

$$Z_p \geq 0.41\ \mathrm{Bias}_A^{-0.25} \quad \text{(amplitude)}. \tag{8}$$

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

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

is needed.

Table 5: Coefficients for approximate maximum 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.

4. MINIMUM FREQUENCY SEPARATION

4.1. 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. 4(a). Here, a main-lobe width is defined as the frequency difference between the first zero-crossings about the main lobe.[3] We denote this frequency separation as $\Delta\omega_m$. Actual values of this criterion for some often-used windows are shown in Table 6. 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.

4.2. 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.

[3]This definition of main-lobe width is not appropriate for all window types, such as those considered in [17], but it is a convenient definition for the window types considered in this paper.

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

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

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

$$X_w(\omega) = \sum_{n=0}^{M-1} w(n)x(n)e^{-j\omega n}, \qquad (11)$$

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. 4(b)), i.e.,

$$|X_w(\omega)|_{\omega=\text{peak}} > |X_w(\omega)|_{\omega=(\omega_0+\omega_1)/2}\ . \qquad (12)$$

Since the exact peak positions in the spectrum cannot be easily determined, we approximate ω_0 by the true frequency of the target component, and require

$$|X_w(\omega)|_{\omega=\omega_0} > |X_w(\omega)|_{\omega=(\omega_0+\omega_1)/2}\ . \qquad (13)$$

Using Eqs. (10) and (11) and considering the worst-case phases, we obtain the following inequality:

$$Y(\omega) \stackrel{\triangle}{=} W(0) - 2W(\Delta\omega/2) - |W(\Delta\omega)| > 0, \quad (14)$$

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

$$\Delta\omega_r \stackrel{\triangle}{=} \Delta\omega \text{ s.t. } \mathrm{Y}(\Delta\omega) = 0. \qquad (15)$$

Numerically obtained solutions for some windows are tabulated in Table 7.

For the second requirement, it intuitively seems sufficient that the two components be separated by *half*

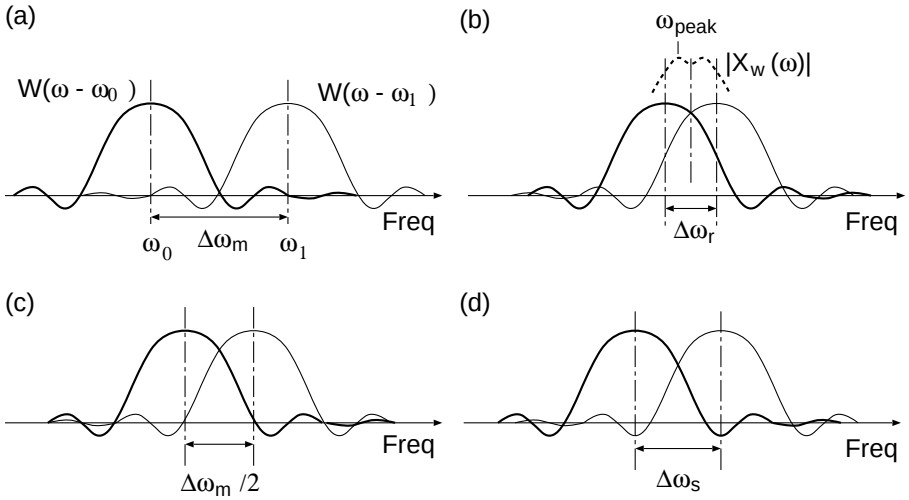

Fig. 4: Interference from nearby components of various frequency separations: (a) the two main lobes are separated by one main-lobe width. (b) the two peaks are just resolved, (c) the first side-lobe zeros are placed at the main-lobe peaks, (d) the first side-lobe peaks are placed at the main-lobe peaks (more precise criterion).

a main-lobe width, since in this case, as shown in Fig. 4(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. 4(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. 4(d). We denote this frequency separation as

$$\Delta\omega_s \stackrel{\triangle}{=} \text{Min}(\omega) \text{ s.t. } \frac{d}{d\omega}W(\omega) = 0 \text{ and } \omega > 0. \quad (16)$$

Actual values for some often-used windows are tabulated in Table 8.

Taking the larger of Eqs. (15) and (16), we obtain the predicted minimum allowable frequency separation (MAFS) as

$$\Delta\omega_{\min} = \text{Max}(\Delta\omega_r, \Delta\omega_s). \quad (17)$$

Comparing Tables 7 and 8, 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. (17) as

$$\Delta\omega_{\min} = \Delta\omega_s + \frac{2\pi}{N} = \Delta\omega_s + \frac{1}{Z_p}(2\pi/M). \quad (18)$$

Actual values for some often-used windows with zero-padding factors 2.0, 3.5 and 5.0 are shown in Table 9.

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

Table 6: 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 7: 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 8: 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

Table 9: 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

believe our slightly conservative prediction is more useful than a tighter but overly optimistic MAFS.

4.3. 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.

4.3.1. Test Conditions

The test signal is composed of two complex sinusoids, as

$$x(n) = A_0 e^{j(\omega_0 n + \phi_0)} + A_1 e^{j(\omega_1 n + \phi_1)}. \qquad (19)$$

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 (A_0, ω_0, ϕ_0) using the QIFFT method. 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.

4.3.2. Results

The obtained maximum bias curves for the zero-padding factor of 5 are shown in Fig. 5 and 7. 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

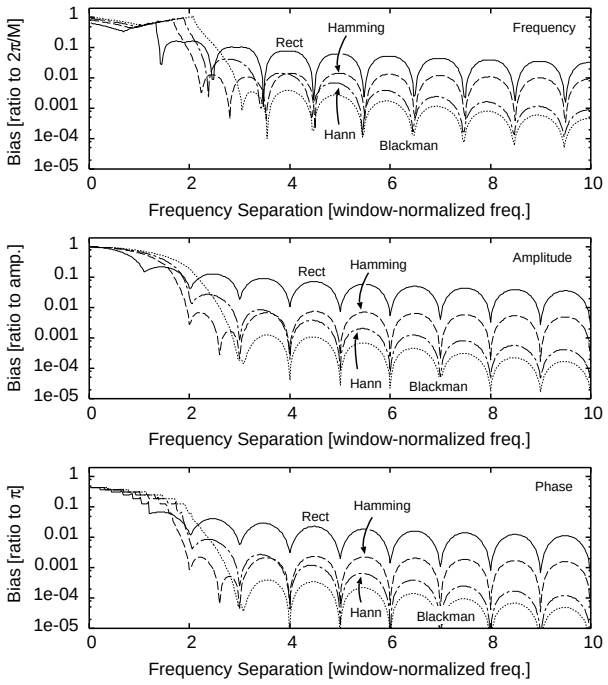

Fig. 5: Bias with windows in the Blackman-Harris family ($Z_p = 5.0$): frequency (top), amplitude (middle) and phase (bottom).

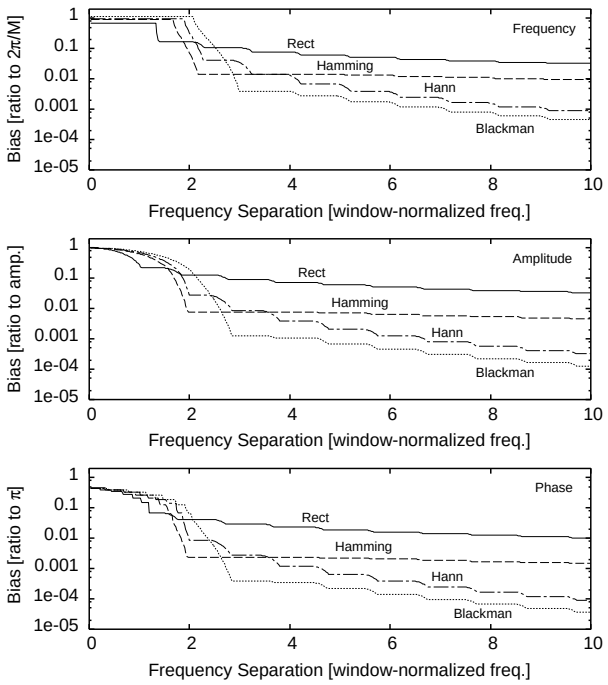

Fig. 6: Filtered bias with windows in Blackman-Harris family ($Z_p = 5.0$): frequency (top), amplitude (middle) and phase (bottom).

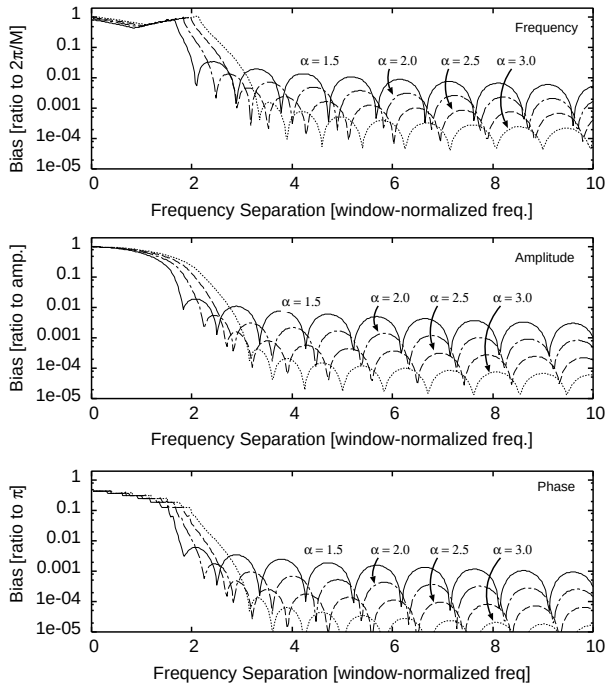

Fig. 7: Bias with Kaiser-Bessel windows ($Z_p = 5.0$): frequency (top), amplitude (middle) and phase (bottom).

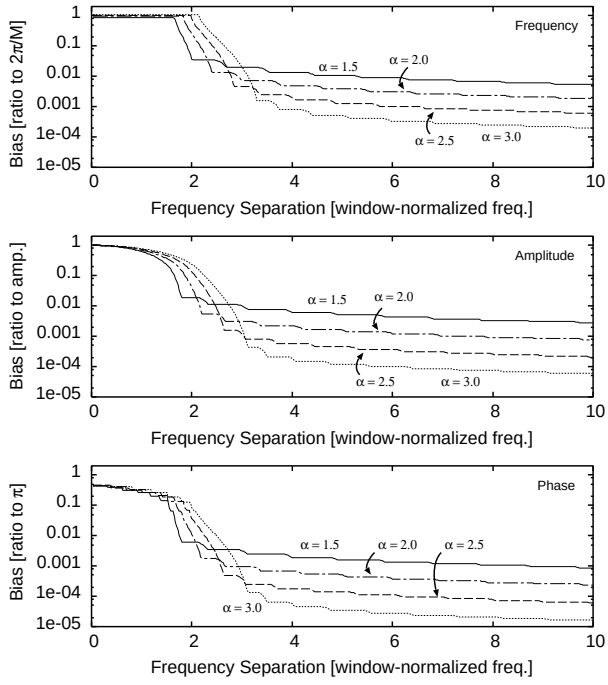

Fig. 8: Filtered bias with Kaiser-Bessel windows ($Z_p = 5.0$): frequency (top), amplitude (middle) and phase (bottom).

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. 4(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. 4(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. 6 and 8. 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 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 9 on page 8. As predicted from theory in Eq. (17), 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 6, and they never underestimate the experimentally obtained MAFSs.

4.4. Minimum Window Length

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

$$M \geq 2\pi(\text{MAFS})/\Delta\omega, \tag{20}$$

where the MAFS can be obtained from Figs. 6, 8, Table 9, or Eq. (18).

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.

5. NOISE ROBUSTNESS

Rife and Boorstyn [8, 16] 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.

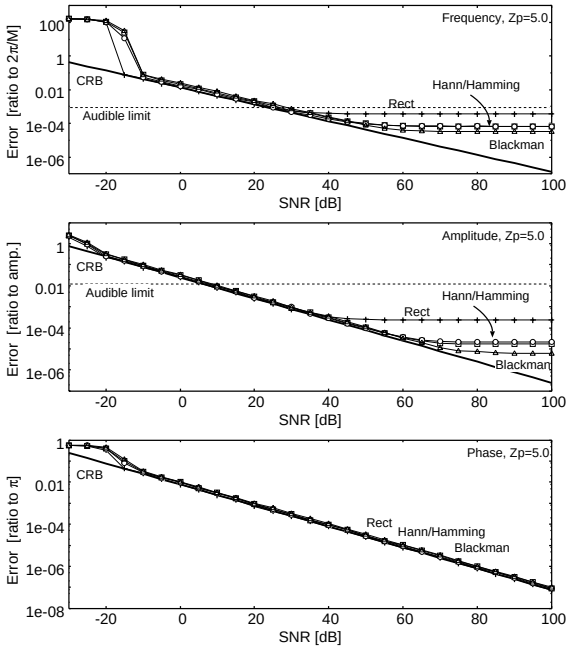

Fig. 9: RMS errors for a single sinusoid in AWGN (Blackman-Harris family): frequency (top), amplitude (middle) and phase (bottom).

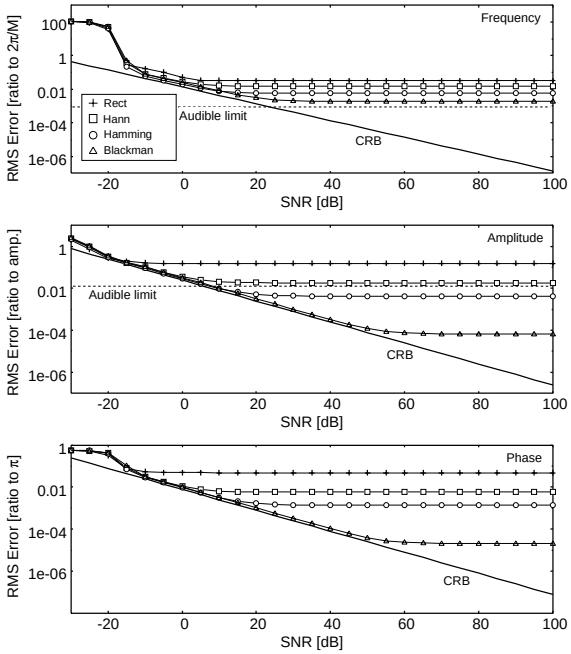

Fig. 11: RMS errors for two sinusoids with AWGN, shown with the CRBs for a single sinusoid (Blackman-Harris family): frequency (top), amplitude (middle) and phase (bottom).

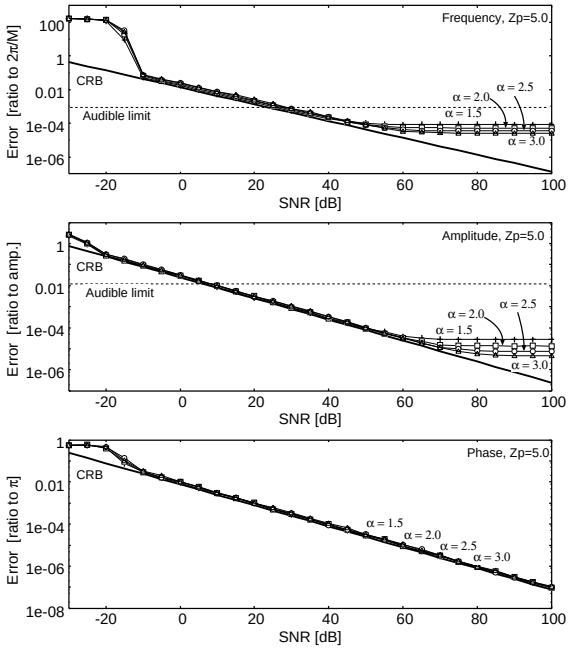

Fig. 10: RMS errors for a sinusoid with AWGN (Kaiser-Bessel windows): frequency (top), amplitude (middle) and phase (bottom).

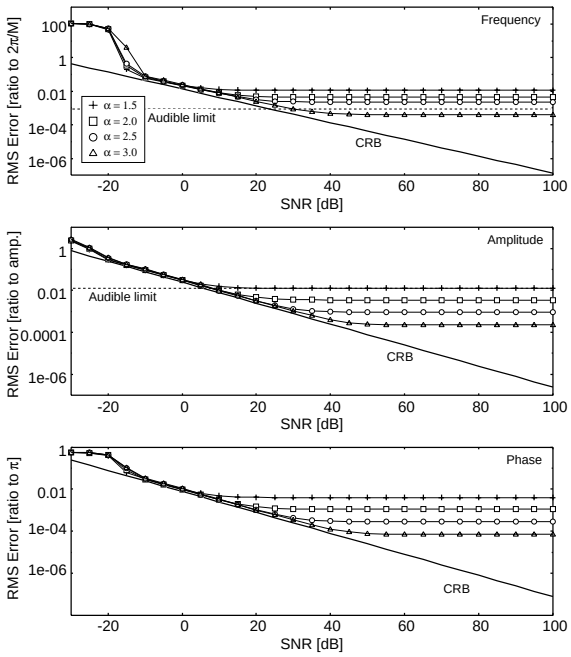

Fig. 12: RMS errors for two sinusoids with AWGN, shown with the Cramer-Rao bounds for a single sinusoid (Kaiser-Bessel windows): frequency (top), amplitude (middle) and phase (bottom).

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 (which depends on the zero-padding factor) and interfering side lobes (which depends on the choice of window type). 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 chosen to be AWGN. Specifically, we use a fixed FFT size of 4096 and a zero-padding factor of 5.0—a value 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 sections. In case of multiple sinusoids, the frequency separation is set to the MAFS of each window.

The results are shown in Figs. 9 and 10 for a single sinusoid, and in Figs. 11 and 12 for multiple sinusoids. 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 [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 sections. It should perhaps be noted that a 0.1% peak-frequency bias cannot be heard in most audio applications.

6. DISCUSSION AND 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 examined the relation between estimation bias and frequency spacing of multiple components, and provided a criterion for reliable estimation. 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 confirmed that the QIFFT works as an approximate ML estimator within a middle S/N ratio range as well. This range can be extended arbitrarily high through the use of more zero padding, e.g., to the limits of human error perception.

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. We have reported our preliminary investigation in [23] and [25].

For further extension, to reduce the interpolation bias, a simple bias-correction method utilizing cubic and quadratic functions has been proposed [24]. To reduce the interference bias, some iterative methods to reduce the interference bias have been proposed [17, 20]. Combining the QIFFT with these methods may reduce the bias in exchange for some computational complexity.

7. 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://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, Vol.34, No.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 Processing, 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. 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] D. C. Rife and R. R. Boorstyn: "Multiple Tone Parameter Estimation from Discrete-Time Observations," Bell System Technical Journal, 55, 9, 1389/1410 (1976).

[17] 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).

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

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

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

[21] 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, August, (2004).

[22] 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).

[23] 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).

[24] 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).

[25] 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).