



**QUEEN'S
UNIVERSITY
BELFAST**

the XQIFFT: Increasing the Accuracy of Quadratic Interpolation of Spectral Peaks via Exponential Magnitude Spectrum Weighting

Werner, K. J. (2015). the XQIFFT: Increasing the Accuracy of Quadratic Interpolation of Spectral Peaks via Exponential Magnitude Spectrum Weighting. In *Proceedings of the International Computer Music Conference* (pp. 326–333). <http://quod.lib.umich.edu/cgi/p/pod/dod-idx/xqiff-increasing-the-accuracy-of-quadratic-interpolation.pdf?c=icmc;idno=bbp2372.2015.066>

Published in:
Proceedings of the International Computer Music Conference

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

© 2015 The Authors.

This is an open access article published under a Creative Commons Attribution License (<https://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

The XQIFFT: Increasing the Accuracy of Quadratic Interpolation of Spectral Peaks via Exponential Magnitude Spectrum Weighting

Kurt James Werner

Center for Computer Research in Music and Acoustics (CCRMA), Stanford University
Stanford, CA, USA 94305

kwerner@ccrma.stanford.edu

ABSTRACT

In spectral audio, accurate estimation of the frequency and magnitude of spectral peaks is a central task. One particularly attractive method for increasing the interpolation density of a spectrum is both inexpensive and accurate: quadratic interpolation over magnitude spectrum peaks (MQIFFT). Although it is known that using this method on a logarithmically-weighted magnitude spectrum (LQIFFT) can increase its accuracy, other weighting functions are under-explored. In this paper, we introduce an exponentially-weighted version of the QIFFT: the XQIFFT. The accuracy of each method is estimated experimentally; for estimation of stationary sinusoid parameters, the XQIFFT is much more accurate than other known QIFFT-derived methods. This increased accuracy is not accompanied by an increase in runtime computational cost, but it does require a one-time tuning procedure which chooses the best exponent by brute-force evaluation of a range of candidates. Correction functions (CQIFFT) are known to unbiased some of the systematic errors of these methods. We apply similar functions to improve XQIFFT estimation. We informally study noise sensitivity properties of the XQIFFT and CXQIFFT; they are more sensitive to noise than known QIFFT-derived methods, but still perform better across a wide range of noise levels.

1. INTRODUCTION

Peak estimation is a central task in spectral audio [1]. As an essential part of sinusoidal analysis/synthesis methods [2–6], pre-processing step in music information retrieval (MIR) applications, etc., the accuracy of peak estimation can be very important. However, peak estimation accuracy is limited to ± 0.5 spectral bins by the resolution of the discrete Fourier transform (DFT). This lack of accuracy can be problematic, especially if accuracy is needed in a regime that matches human perception, i.e. logarithmic scales. In this paper, we’ll call the method of estimating the true peak location by just

taking the peak bin in the magnitude spectrum the “nearest bin method.”

One way to find peaks more accurately is to increase the interpolation density of the DFT by zero padding [7]¹. However, zero padding is not computationally efficient.² There are other exotic techniques for processing an entire spectrum to increase peak-estimation accuracy [8–13], but these tend to be expensive. For many spectral audio applications, efficiency is important. This is especially true for real-time applications and applications based on processing huge data sets, e.g. MIR.

One particularly well-known technique for estimating the frequency and magnitude of stationary complex sinusoids is both simple and inexpensive [14–17].³ In the audio DSP literature, this technique is called the QIFFT: the “Quadratically-Interpolated FFT” [19]. In the QIFFT, each peak bin in a logarithmically-weighted magnitude spectrum and its left and right neighbors are chosen, a parabola is fit exactly to these three points, and then the vertex of this parabola is chosen as a refined estimate of the true spectral peak.⁴

In addition to the nearest bin method for peak interpolation, this paper considers a family of QIFFT-derived methods. These include the MQIFFT (a version of the QIFFT calculated on an un-weighted magnitude spectrum); the standard logarithmically-weighted version, which we’ll call the LQIFFT; and an exponentially-weighted version which we’ll introduce in this paper, the XQIFFT. We’ll also consider “corrected” versions of these: the CMQIFFT, the CLQIFFT, and the CXQIFFT.

The LQIFFT and its cousins are very attractive, since they (a) can reduce error significantly; (b) are inexpensive (requiring only a few multiplies to find the refined peak frequency

¹ https://ccrma.stanford.edu/~jos/dft/Zero_Padding_Theorem_Spectral.html

² For a length- N signal, the computational complexity of the Fast Fourier Transform (FFT) algorithm for calculating the DFT is $\mathcal{O}(N \log N)$. So, zero padding a signal by a factor of D up to a length ND raises the cost by a factor $D(\log_N D + 1)$. Zero padding by a factor D increases the cost of the algorithm by a factor that is greater than D —this is very expensive.

³ The roots of parabolic interpolation as a general mathematical concept stretch many centuries earlier than these signal-processing references [18].

⁴ A variation on this technique is to compute a least-squares fit of the three magnitude spectrum bins to a parabola with a fixed shape, rather than the unique parabola which fits the three bins exactly [20]. Parabolic interpolation techniques are also used for improving estimates of local extrema in time-domain methods (correlation, squared-difference, etc.) [21–24].

and amplitude estimates); (c) can be used with any window type⁵; and (d) can even be combined with zero padding. Even with the inclusion of the CLQIFFT’s correction factor, some error remains. A question asked by Smith and Serra in 1985 [2] still remains open: “... what is the optimum non-linear compression of the magnitude spectrum when quadratically interpolating it to estimate peak locations?” In this paper, we investigate this question, adding to the family of QIFFT methods in the process.

The goal of this paper is to introduce *exponential* nonlinear weighting of the magnitude spectrum as a method in the spectral audio toolbox. We’ll call this technique XQIFFT for “eXponentially-weighted QIFFT.” The XQIFFT is merely a QIFFT on a magnitude spectrum that has been raised to a certain power p . We will show that the XQIFFT has huge potential for improving upon existing QIFFT-derived methods. By adding compensation to the XQIFFT to produce a CXQIFFT—a “Corrected XQIFFT”—the estimate can be refined even further.

One caveat of the XQIFFT is that p must be chosen carefully. When it is chosen correctly, it can increase the accuracy substantially. But if p is chosen incorrectly, it can decrease estimation accuracy back to that of the nearest bin method. It is possible that analytic solutions or clever iterative methods exist for finding the p which most benefits the XQIFFT, but in this paper we’ll just use brute force search methods. For a given window type and window length, searching for the best p is a one-time cost only. By merely using a brute force search across reasonable values, we’ll get the right answer to high precision in only a few minutes.

Since we won’t be using particularly refined methods to find p , we won’t make any claims of optimality. Rather, we hope to demonstrate that the XQIFFT and CXQIFFT produce substantial improvements on the order of $\times 50$ over their highest-performing QIFFT cousins (LQIFFT and CLQIFFT) in a typical situation, nearly for free, and can be expected to provide at least some improvement when used on any window.⁶

The proper choice of p depends heavily on the length and type of the analysis window. In this paper, we’ll be testing length-4096 Hann windows. So, the results (estimate of the best choice of p) will be valid for length-4096 Hann windows only. However, the testing framework we introduce will be valid for *any* window length and type—each window just has to be run through the testing procedures we’ll define in this paper.

The structure of the remainder of this paper is as follows. §2 considers “uncorrected methods” (the MQIFFT, LQIFFT, and XQIFFT) and contains a review of parabolic interpolation and the nonlinear weighting factors which yield the LQIFFT and XQIFFT from the MQIFFT. §3 contains an analysis of the error of the uncorrected methods, and shows how estim-

ates of this error can yield refined “corrected” methods: the CMQIFFT, CLQIFFT, and CXQIFFT. The desirable properties of the XQIFFT are highlighted. §4 briefly studies the noise sensitivity properties of all considered methods. §5 concludes the paper and contains speculation about potential future work.

2. UNCORRECTED METHODS

In this section, we’ll define the goals of peak estimation and explain how the performance of a method can be predicted experimentally. We’ll also review nearest bin (§2.1), QIFFT (§2.2), and LQIFFT (§2.3.1) methods and introduce our novel XQIFFT method (§2.3.2). Along the way, we will introduce a unified notation for nonlinear weightings of the magnitude spectrum in the QIFFT family (§2.3) and discuss experimental tests of the performance of each method (§2.4).

The DFT of a length- N signal $x(n)$ is defined as [7]⁷:

$$X(k) \triangleq \sum_{n=0}^{N-1} x(n)e^{-j2\pi nk/N}, \quad k = 0, 1, 2, \dots, N-1. \quad (1)$$

Complex sinusoids [7]⁸ in the time domain will show up as impulses in the DFT magnitude spectrum $|X(k)|$, provided that two criteria are met: (a) the frequency of the complex sinusoid is exactly on a bin center; and (b) the signal is infinitely long. In general, neither of these criteria will be met. The signal $x(n)$ will be finite-length, and multiplied by a windowing function in the time domain. By the dual of the convolution theorem [7]⁹, multiplication in the time domain corresponds to convolution in the frequency domain. So, multiplication by a window function in the time domain corresponds to smoothing in the frequency domain by the window’s transform—the spectrum we be the sum of the window transform shifted to the location and scaled to the magnitude of each sinusoid present in $x(n)$.

The goal of the nearest bin method and the QIFFT family is to estimate the bin \hat{k}_t and the magnitude $|\hat{X}_t|$ which approximate the true (not necessarily integer) bin k_t and magnitude $|X_t|$ of the sinusoid which is contributing a particular spectral peak in $X(k)$. In this notation, a hat over a value indicates that it is an estimate of that value, and the subscript t will always refer to the true peak.

In the rest of the paper, we’ll be talking about the “performance” of various methods. For an individual peak estimate, we define the errors e_k and $e_{|X|}$ of the bin and magnitude

⁵ The only exception is the non-zero-padded rectangular window—since it has a main lobe width less than 3 bins, it will require a zero-padding factor of $\times 1.5$ for QIFFT methods to be reliable [25].

⁶ We won’t expect serious gains for the Gaussian window, however, since logarithmic weighting of the Gaussian window transform already produces a main lobe that is nearly perfectly parabolic.

⁷ https://ccrma.stanford.edu/~jos/mdft/Mathematics_DFT.html

⁸ https://ccrma.stanford.edu/~jos/mdft/Complex_Sinusoids.html

⁹ https://ccrma.stanford.edu/~jos/mdft/Dual_Convolution_Theorem.html

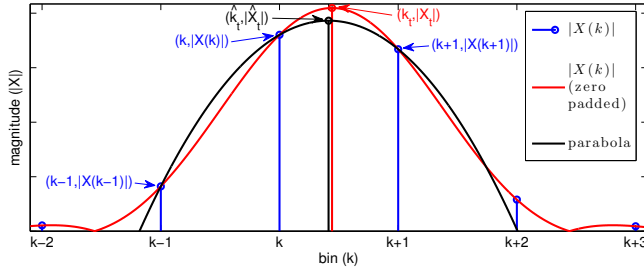


Figure 1. Parabolic fit to magnitude spectrum peaks, finding \hat{k}_t and $|\hat{X}_t|$ (MQIFFT).

estimates as¹⁰

$$e_k = \hat{k}_t - k_t \quad (2)$$

$$e_{|X|} = \frac{|\hat{X}_t| - |X_t|}{|X_t|}. \quad (3)$$

Along with individual estimate errors e_k and $e_{|X|}$, we define corresponding *worst-case* errors E_k and $E_{|X|}$. Since we don't have analytic expressions for E_k or $E_{|X|}$, we will have to estimate them experimentally. In each case, estimates \hat{E}_k and $\hat{E}_{|X|}$ of the worst-case bin and magnitude errors are found by testing 1000 randomly-generated complex sinusoids¹¹ and choosing the worst (i.e. maximum $|e_k|$ or $|e_{|X|}|$, respectively). When we talk about good performance, we mean minimizing \hat{E}_k or $\hat{E}_{|X|}$.

2.1 Nearest Bin Method

The nearest bin method for peak estimation involves finding peaks in the magnitude spectrum, i.e. discrete magnitude spectrum bins which are larger than both their lower and upper neighbors. In the nearest bin method, this bin's index k and magnitude $|X(k)|$ are considered to be estimates of the true peak, which is at bin k_t and magnitude $|X_t|$, i.e.

$$\hat{k}_t = k \quad (4)$$

$$|\hat{X}_t| = |X(k)|. \quad (5)$$

2.2 MQIFFT

The MQIFFT method works by considering the same bin found by the nearest bin method as well as its lower and upper neighbors, i.e. for each peak, the MQIFFT method considers three spectral bins with indices $k-1$, k , and $k+1$ and their

¹⁰ Note well that we are defining bin error e_k in the obvious way, as the difference between the estimate and the ground truth. Magnitude error $e_{|X|}$ is defined proportionally, as the ratio between $|\hat{X}_t| - |X_t|$ and $|X_t|$.

¹¹ It may seem that complex sinusoids are an odd choice to test the QIFFT methods. However, since sine and cosine functions produce both positive and negative frequency components, it is actually difficult to control for the effect of interference, i.e. for a sinusoid, the sidelobes of positive frequency components near DC or Nyquist will "wrap around" and affect their negative-frequency counterparts (and vice versa). Rather than restrict the range of test frequencies to try to control for this, we'll just test individual complex sinusoids, with their single frequency component.

method	$f(\Theta)$	$f^{-1}(\Phi)$
MQIFFT	Θ	Φ
LQIFFT	$\ln \Theta$	e^Φ
XQIFFT	Θ^p	$\Phi^{\frac{1}{p}}$

Table 1. Magnitude spectrum weighting and unweighting functions for the QIFFT family.

corresponding entries in the magnitude spectrum $|X(k-1)|$, $|X(k)|$, and $|X(k+1)|$. From now on, we'll refer to the peak magnitude spectrum bin and its two neighbors with a substitution (to clean up the notation a little, and to echo [2]):

$$\alpha = |X(k-1)| \quad (6)$$

$$\beta = |X(k)| \quad (7)$$

$$\gamma = |X(k+1)|. \quad (8)$$

As with the nearest bin method, the goal of the MQIFFT is to find an estimate \hat{k}_t of the true peak bin k_t and an estimate $|\hat{X}_t|$ of the true peak magnitude $|X_t|$. The MQIFFT accomplishes this by fitting a unique parabola to the three points $(k-1, \alpha)$, (k, β) , and $(k+1, \gamma)$ that represent the DFT peak and its two neighbors. This can be done by writing up the second-order Lagrange interpolating polynomial [1,26]¹² that fits the three known points, rearranging it into the vertex form of a parabola ($|X(\kappa)| = a(\kappa - \hat{k}_t)^2 + |\hat{X}_t|$), and then considering the vertex of the parabola $(\hat{k}_t, |\hat{X}_t|)$ to be an estimate of the true peak. This is shown in Figure 1.

The equations for \hat{k}_t and $|\hat{X}_t|$ end up being closed-form expressions in terms of k , α , β , and γ :

$$\hat{k}_t = k + \left(\frac{1}{2}\right) \frac{\alpha - \gamma}{\alpha - 2\beta + \gamma} \quad (9)$$

$$|\hat{X}_t| = \beta - \left(\frac{1}{8}\right) \frac{(\alpha - \gamma)^2}{\alpha - 2\beta + \gamma}. \quad (10)$$

2.3 Nonlinear-weighted QIFFTs

The MQIFFT works on the magnitude spectrum directly. Cousins of the MQIFFT work on nonlinearly-weighted versions of the magnitude spectrum. One well-known technique in this category is the LQIFFT; the XQIFFT method is another. Here, we introduce a unified notation, which is a variation on (9)–(10), for referring to these nonlinear weightings:

$$\hat{k}_t = k + \left(\frac{1}{2}\right) \frac{f(\alpha) - f(\gamma)}{f(\alpha) - 2f(\beta) + f(\gamma)} \quad (11)$$

$$|\hat{X}_t| = f^{-1} \left(f(\beta) - \left(\frac{1}{8}\right) \frac{(f(\alpha) - f(\gamma))^2}{f(\alpha) - 2f(\beta) + f(\gamma)} \right). \quad (12)$$

Note that in addition to replacing the α , β , and γ from (9)–(10) with $f(\alpha)$, $f(\beta)$, and $f(\gamma)$, we also need to unweight

¹² https://ccrma.stanford.edu/~jos/pasp/Lagrange_Interpolation.html

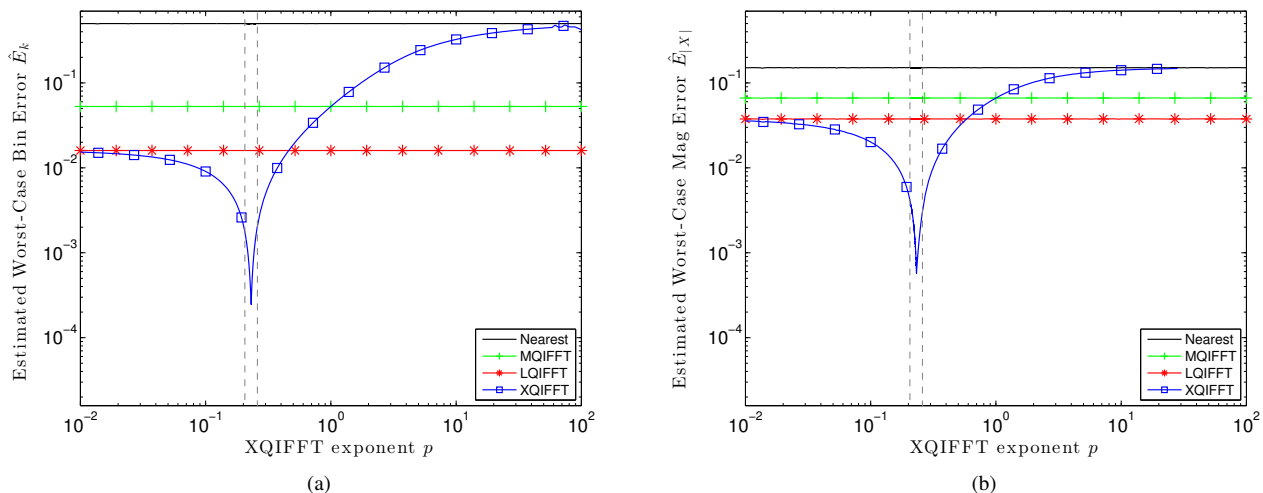


Figure 2. Estimated worst-case errors for nearest bin, QIFFT, LQIFFT, and XQIFFT methods: (a) estimated worst-case bin error \hat{E}_k ; and (b) estimated worst-case magnitude error $\hat{E}_{|X|}$. The dotted gray box shows the region that is shown in greater detail (and with the inclusion of corrected methods) in Figure 4.

equation (12) with a f^{-1} , since the quadratic fit is now estimating a *weighted* magnitude.

Using (11)–(12), each of the methods from the QIFFT family is implemented by choosing appropriate weighting and unweighting functions $f(\Theta)$ and $f^{-1}(\Phi)$. In general, the combination of a weighting function and its corresponding unweighting function is an identity function, i.e. $f(f^{-1}(\Phi)) = \Phi$. The only unweighted method in the family of QIFFT methods is the MQIFFT—both its weighting and unweighting functions are the identity function; they are only mentioned at all so that we can use a common notation for the entire family.

Nonlinear weighting and unweighting functions $f(\Theta)$ and $f^{-1}(\Phi)$ for each method are given in Table 1. The nonlinear weighting and unweighting functions for the LQIFFT and the XQIFFT are discussed below.

2.3.1 LQIFFT

The LQIFFT method, already discussed in e.g. [2, 19], is identical to the MQIFFT method, except that the quadratic fit is done on a *logarithmically-weighted* version of the magnitude spectrum, rather than the magnitude spectrum itself. Each of the three bin magnitudes α , β , and γ is weighted by $f(\Theta) = \ln \Theta$ and the magnitude estimate is unweighted by the inverse of that function $f^{-1}(\Phi) = e^\Phi$.

2.3.2 XQIFFT

The XQIFFT method is identical to the MQIFFT method, except that the quadratic fit is done on an *exponentially-weighted* version of the magnitude spectrum, rather than the magnitude spectrum itself. Each of the three bin magnitudes α , β , and γ is weighted by $f(\Theta) = \Theta^p$, and the magnitude estimate is unweighted by the inverse of that function $f^{-1}(\Phi) = \Phi^{\frac{1}{p}}$. Compared to the QIFFT and the LQIFFT, the XQIFFT is unique in that it has a tuning parameter p that controls the weighting

and unweighting functions. The performance of the XQIFFT depends heavily on choosing this exponential factor p carefully. We’ll look into how to choose p in the next section, which will also compare the performance of the methods.

2.4 Performance Test

Figure 2 shows plots of the estimated worst-case bin error \hat{E}_k and estimated worst-case magnitude error $\hat{E}_{|X|}$ for all of the peak estimation methods discussed so far: the nearest bin method, MQIFFT, LQIFFT, and XQIFFT. As described at the beginning of this section, since we don’t have analytic expressions for worst-case error, we *estimate* worst-case error by testing 1000 different complex sinusoids, and taking the worst estimate from that batch as the estimated worst-case.

Numerical results from Figure 2 are shown in Table 2. Two slightly different values of p (0.2308 and 0.2318) minimize \hat{E}_k or $\hat{E}_{|X|}$ respectively for the XQIFFT method. For either value, both \hat{E}_k and $\hat{E}_{|X|}$ are almost two orders of magnitude lower than for the existing best uncorrected method, the LQIFFT. Assuming that logarithm and power functions have the same computational cost, this accuracy is approximately free at runtime—the one-time up-front cost of finding the proper p through this testing procedure takes a few minutes by brute-force search on a modest laptop computer.

3. CORRECTED METHODS

In this section, we’ll show how to use “corrected” methods to further refine the estimates we found in §2. Versions of these correction methods have been used previously on the LQIFFT [25]. We extend these methods to our novel XQIFFT method and introduce error estimation factors \hat{e}_k and $\hat{e}_{|X|}$ which are both more accurate than existing polynomial func-

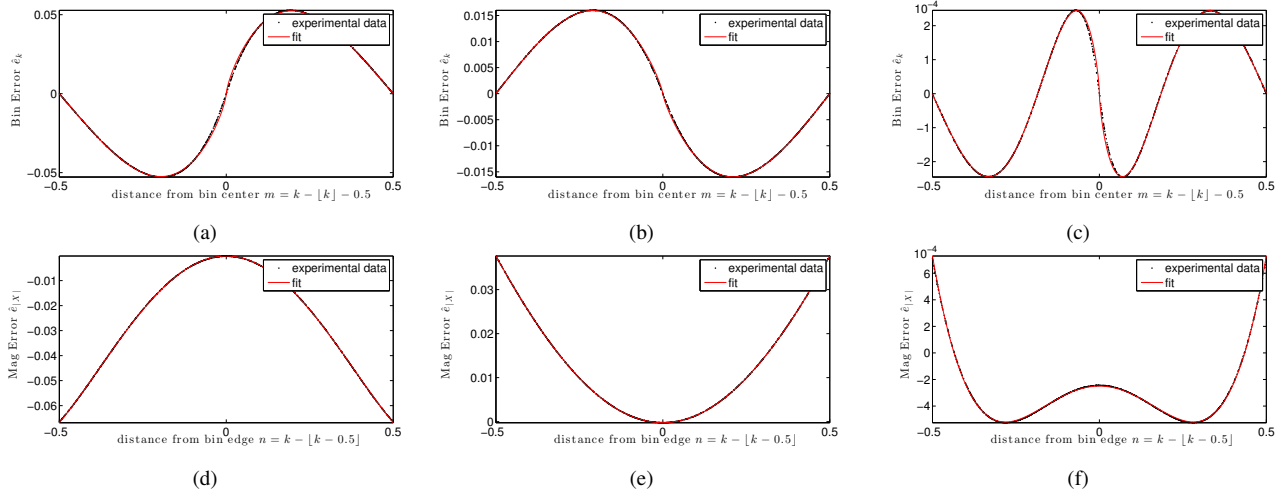


Figure 3. Experimental errors: (a) bin error of MQIFFT; (b) bin error of LQIFFT; (c) bin error of XQIFFT; (d) magnitude error of MQIFFT; (e) magnitude error of LQIFFT; and (f) magnitude error of XQIFFT. Each plot contains both experimental data (black dots) and curve fits (red lines). Bin errors are plotted against $m = \hat{k}_t - \lfloor \hat{k}_t \rfloor - 0.5$, the offset of \hat{k}_t from the closest bin center. Magnitude errors are plotted against $n = \hat{k}_t - \lfloor \hat{k}_t + 0.5 \rfloor$, the offset of \hat{k}_t from the closest bin edge.

method	p	\hat{E}_k	$\hat{E}_{ X }$
nearest bin	—	$4.995 \cdot 10^{-1}$	$1.510 \cdot 10^{-1}$
MQIFFT	—	$5.276 \cdot 10^{-2}$	$6.639 \cdot 10^{-2}$
LQIFFT	—	$1.600 \cdot 10^{-2}$	$3.761 \cdot 10^{-2}$
XQIFFT	0.2308	$2.453 \cdot 10^{-4}$	$6.947 \cdot 10^{-4}$
XQIFFT	0.2318	$2.975 \cdot 10^{-4}$	$5.760 \cdot 10^{-4}$

Table 2. Table of estimated worst-case bin error \hat{E}_k and estimated worst-case magnitude errors $\hat{E}_{|X|}$ for uncorrected methods only, as determined by tests on 1000 randomly-generated sinusoids. Values are given to 4 significant figures. “—” indicates that \hat{E}_k and $\hat{E}_{|X|}$ are p -independent and averaged over all trials.

tions¹³ and appropriate for all three of the MQIFFT, LQIFFT, and XQIFFT methods. The goal of the CMQIFFT, CLQIFFT, and CXQIFFT methods is to find refined peak estimates \tilde{k}_t and $|\tilde{X}_t|$ which incorporate these two correction factors.

In §3.1 we give a justification for the family of corrected methods. In §3.2 we explain the process for experimentally determining error estimation functions $\hat{e}_k(\hat{k}_t)$ and $\hat{e}_{|X|}(\hat{k}_t)$. In §3.3 we compare the performance of corrected methods.

3.1 Refined Estimation via Error Estimate

The true peak bin and magnitude estimates \hat{k}_t and $|\hat{X}_t|$ from the uncorrected methods are good, but they are still not perfect.

Recall (2)–(3) from §2. Rearranging to find the true peak bin and magnitudes k_t and $|X_t|$ in terms of the estimates and errors gives:

$$k_t = \hat{k}_t - e_k \quad (13)$$

$$|X_t| = \frac{1}{e_{|X|} + 1} |\hat{X}_t|. \quad (14)$$

So, if we knew the errors e_k and $e_{|X|}$ alongside our estimates \hat{k}_t and $|\hat{X}_t|$, we would be able to find the true peak bin and magnitudes exactly. We’ll soon find (in §3.2) that although we don’t know the errors exactly, we will be able to form good estimates \hat{e}_k and $\hat{e}_{|X|}$ of the errors as functions of \hat{k}_t . So, we can rewrite versions of (13)–(14) as *refined estimates* of the true values, incorporating error estimates \hat{e}_k and $\hat{e}_{|X|}$:

$$\tilde{k}_t = \hat{k}_t - \hat{e}_k \quad (15)$$

$$|\tilde{X}_t| = \frac{1}{\hat{e}_{|X|} + 1} |\hat{X}_t|. \quad (16)$$

where a tilde over a value indicates that it is a refined estimate of that value.

3.2 Forming Error Estimate Functions

In §3.1, we showed how estimates \hat{e}_k and $\hat{e}_{|X|}$ of the errors e_k and $e_{|X|}$ of true peak bin and magnitudes estimates \hat{k}_t and $|\hat{X}_t|$ could yield refined estimates \tilde{k}_t and $|\tilde{X}_t|$. To actually use this insight, we need a way of finding the error estimates \hat{e}_k and $\hat{e}_{|X|}$.

Figure 3 shows the errors (systematic bias) of \hat{e}_k and $\hat{e}_{|X|}$ for the MQIFFT, LQIFFT, and XQIFFT (at the value of p which minimizes \hat{E}_k , $p = 0.2308$).¹⁴ Luckily, these biases are structured so that they form smooth functions against m and n , where

$$m = \hat{k}_t - \lfloor \hat{k}_t \rfloor - 0.5, \quad (17)$$

the offset of the bin estimate \hat{k}_t from the closest bin center, and

$$n = \hat{k}_t - \lfloor \hat{k}_t + 0.5 \rfloor, \quad (18)$$

¹³ albeit very slightly more expensive

¹⁴ These plots were generated from the same tests used in the previous section

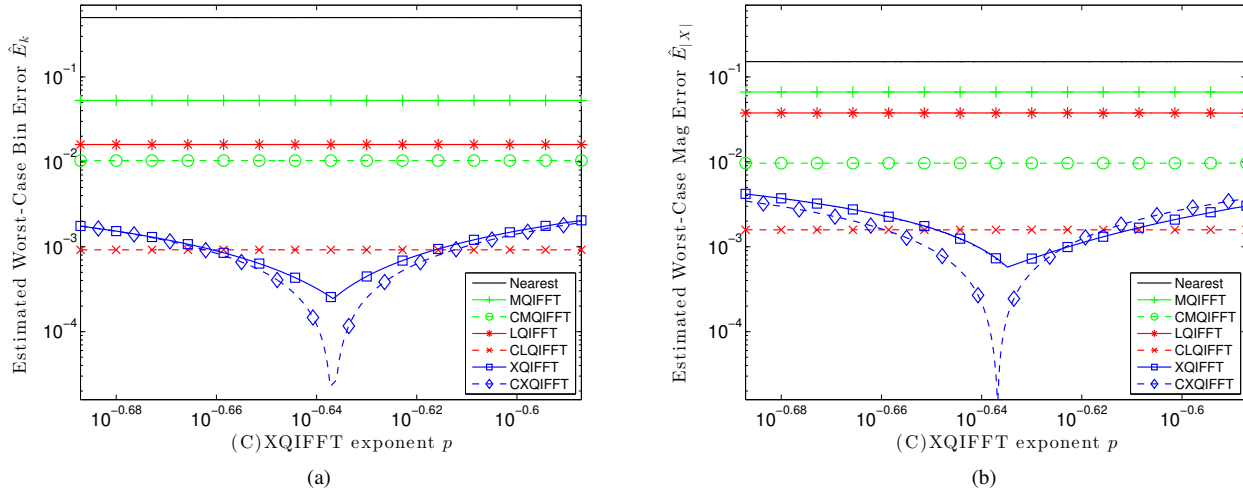


Figure 4. Detail of estimated worst-case errors for nearest bin, MQIFFT, CMQIFFT, LQIFFT, CLQIFFT, XQIFFT, and CXQIFFT methods: (a) estimated worst-case bin error \hat{E}_k ; and (b) estimated worst-case magnitude error $\hat{E}_{|X|}$.

the offset from the closest bin edge.

In [25], polynomial functions are fit to the systematic errors of the LQIFFT to estimate this error. In this work, we fit a nonlinear function to the measured data for \hat{e}_k , and we fit a biquadratic polynomial to $\hat{e}_{|X|}$. These are very slightly more expensive than the polynomials used in [25], but they have the advantage of being more accurate and also the advantage of being appropriate for all three of the MQIFFT, LQIFFT, and XQIFFT methods. These estimates are given by:

$$\hat{e}_k = \text{sgn}(m)c_0 \sin(c_1|m|^{c_2}) \quad (19)$$

$$\hat{e}_{|X|} = c_3n^4 + c_4n^2 + c_5, \quad (20)$$

where coefficients c_0 – c_5 are found using Matlab’s `cftool`.

Figure 3 shows the fits against the measured data—the fits are very good. In the next section, we will show how these correction factors improve peak estimation performance.

method	p	\hat{E}_k	$\hat{E}_{ X }$
nearest bin	—	$4.995 \cdot 10^{-1}$	$1.510 \cdot 10^{-1}$
MQIFFT	—	$5.276 \cdot 10^{-2}$	$6.639 \cdot 10^{-2}$
CMQIFFT	—	$1.033 \cdot 10^{-2}$	$9.643 \cdot 10^{-3}$
LQIFFT	—	$1.600 \cdot 10^{-2}$	$3.761 \cdot 10^{-2}$
CLQIFFT	—	$9.206 \cdot 10^{-4}$	$1.581 \cdot 10^{-3}$
XQIFFT	0.2308	$2.453 \cdot 10^{-4}$	$6.947 \cdot 10^{-4}$
XQIFFT	0.2318	$2.975 \cdot 10^{-4}$	$5.760 \cdot 10^{-4}$
CXQIFFT	0.2305	$2.268 \cdot 10^{-5}$	$3.606 \cdot 10^{-5}$
CXQIFFT	0.2308	$2.399 \cdot 10^{-5}$	$1.370 \cdot 10^{-5}$

Table 3. Table of estimated worst-case bin error \hat{E}_k and estimated worst-case magnitude errors $\hat{E}_{|X|}$ for uncorrected and corrected methods, as determined by tests on 1000 randomly-generated sinusoids. Values are given to 4 significant figures. “—” indicates that \hat{E}_k and $\hat{E}_{|X|}$ are p -independent and averaged over all trials.

3.3 Performance Test (Including Corrected Methods)

Figure 4 shows plots of the estimated worst-case bin error \hat{E}_k and estimated worst-case magnitude error $\hat{E}_{|X|}$ for all of the peak estimation methods discussed so far: the nearest bin method, MQIFFT, CMQIFFT, LQIFFT, CLQIFFT, XQIFFT, and CXQIFFT. As in §2.4, we estimate worst-case error by testing 1000 different complex sinusoids, and taking the worst estimate from that batch as the estimated worst-case. Figure 4 zooms in on the neighborhood of p values that minimize \hat{E}_k in the XQIFFT to show detail.

Numerical results from Figure 4 are shown in Table 3. Two slightly different values of p (0.2305 and 0.2308) minimize \hat{E}_k or $\hat{E}_{|X|}$ respectively for the CXQIFFT method.¹⁵ For either value, both \hat{E}_k and $\hat{E}_{|X|}$ are almost two orders of magnitude lower than for the existing best corrected method, the CLQIFFT. Again, the CXQIFFT performs much better than existing methods without incurring extra cost, i.e. correction factors with the same computational complexity are used for all methods. The CXQIFFT performs approximately one order of magnitude better than the XQIFFT.

4. NOISE SENSITIVITY

So far we have been testing the performance of these algorithms on individual complex sinusoids. In this context, the XQIFFT far out-performs existing uncorrected methods, and the CXQIFFT far outperforms existing corrected methods. Neither is more expensive than the existing best methods, the LQIFFT and the CLQIFFT. Although this seems very promising, we need to be sure that this increase in performance is not only available for a single complex sinusoid. Practical situations may involve many sources of estimation error, including

¹⁵ Note that when p is very small the performance of the XQIFFT approaches that of the LQIFFT, and when p is very large the performance of the XQIFFT approaches that of the nearest bin method.

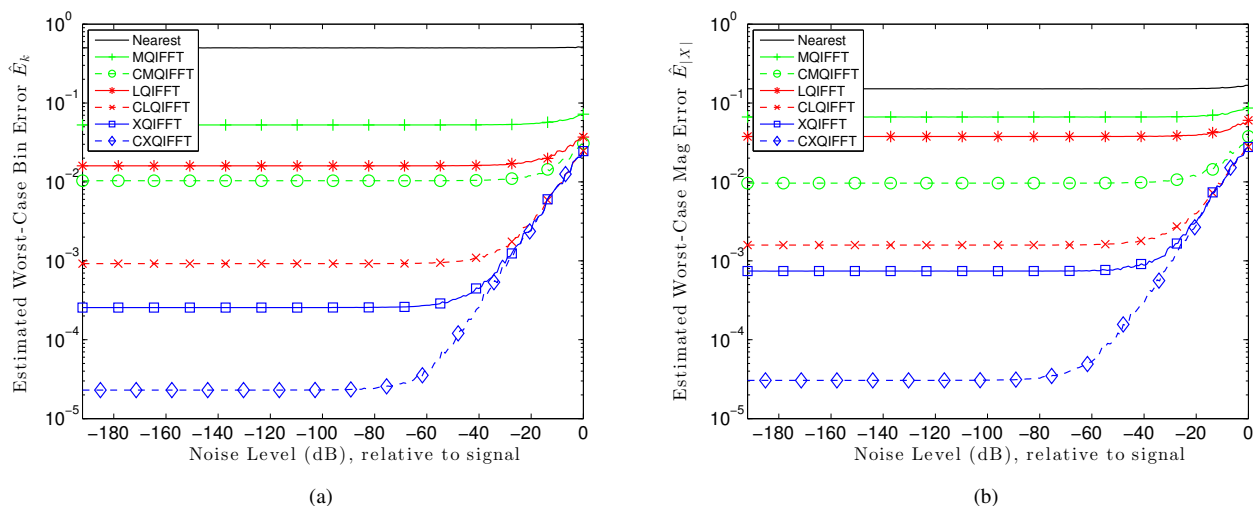


Figure 5. Estimate worst-case error for nearest bin, MQIFFT, CMQIFFT, LQIFFT, CLQIFFT, XQIFFT, and CXQIFFT methods with noise: (a) estimated worst-case bin error \hat{E}_k ; and (b) estimated worst-case magnitude error $\hat{E}_{|X|}$. A range of worst-case errors are tested under a range of noise levels $\in [-192, 0]$

nearby frequency components (interference), non-stationarity of the sinusoids (amplitude or frequency modulation), and noise. Here, we’ll test single sinusoids in noise as a way of starting to investigate the behavior in the presence of error sources.

Each method is tested on a single sinusoid in the presence of uniform noise, with noise levels (relative to sinusoid level) at 1 dB increments between -96 to 0 . As in the previous section, the XQIFFT and CXQIFFT use the value of p which minimized \hat{E}_k for the QIFFT, $p = 0.2308$. The results of this test are shown in Figure 5. Although the performance of each method degrades as the noise level rises—i.e. \hat{E}_k and $\hat{E}_{|X|}$ increase—the hierarchy across method performance remains the same. The XQIFFT always outperforms other uncorrected methods and the CXQIFFT always outperforms other corrected methods, although the performance gap narrows as noise level increases.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we introduced the concept of using a magnitude spectrum weighted exponentially in the context of quadratic interpolation of magnitude spectrum peaks: the XQIFFT. We have shown that choosing the right exponent p in this weighting for a particular window length and type is very important. If p is chosen correctly, as the result of a brute-force search over a range of candidates, we can expect the XQIFFT to perform significantly better than the LQIFFT as an estimator of peak frequencies and magnitudes of stationary sinusoids.

A “corrected” version of the XQIFFT, the CXQIFFT, attempts to mitigate systematic bias in the XQIFFT. This improves estimation by approximately another order of magnitude, and outperforms the CLQIFFT.

Of course, in practical situations, sources of recording noise, interferers, etc. will necessarily keep the accuracy of any existing or future member of the QIFFT family below the theo-

retical limit. We informally studied the noise sensitivity of the XQIFFT/CXQIFFT, showing that although its performance degrades in the presence of interferers, it is still expected to perform better than the LQIFFT/CLQIFFT. A more formal analysis of sources of errors [27, 28] should be the subject of future work.

Late in this research, we became aware of a QIFFT-like technique [16] in particle physics, and an XQIFFT-like refinement [17, 29]. In the latter of these works, data acquisition is done at 10 bits, so gains in accuracy past the theoretical limit of the XQIFFT would be pointless, i.e. even using the XQIFFT, errors are dominated by quantization error, not interpolation error. In audio, 16-bit data acquisition is standard and 24-/32-bit architectures are not uncommon. Even using the CXQIFFT, the meager error that remains is totally dominated by interpolation error, not quantization error. This means that, in audio, there is still scope for improvement of QIFFT methods, even upon the CXQIFFT. The results shown in this paper show obvious improvement in reducing peak estimation bias. Putting formal bounds on this improvement should be the subject of current research.

Certainly, more elegant solutions for finding a good p , even an optimal one, may exist. However, we stress that the brute-source search is a one-time up-front expense, is very reliable, and can get to arbitrary precision by recursively searching smaller and smaller regions.

Since they are not available in closed form, the methods presented in this paper are inherently justified only by experimental results. Although the results we have presented are compelling, they are not analytic proofs of the method’s viability. However, we hope they are convincing enough to show that nonlinear magnitude weightings have an untapped power to improve spectral peak estimation techniques based on parabolic interpolation. We make no claims of optimality, but point out that even non-optimal methods for estimating p

can lead to significant increases in peak estimation accuracy. Future research will hopefully study formal bounds on the algorithm's accuracy, comparing it to, e.g., the Cramér–Rao bound.

We hope that future work will consider new types of non-linear weightings, focus on methods for finding optimal non-linear weighting parameters analytically, and show improvements in real-world applications based on improved peak estimation.

Acknowledgments

Thanks to professor Julius O. Smith III for his feedback—this paper grew out of work in his spectral audio signal processing seminar [1]. Thanks to Melissa Kagen and John Granzow for their editorial feedback and to Tim O'Brien for helpful discussions on spectral audio. Thanks to the anonymous reviewers who gave truly helpful advice which I hope to be able to explore in the future.

6. REFERENCES

- [1] J. O. Smith III, *Spectral Audio Signal Processing*, 2010 ed., online book, <http://ccrma.stanford.edu/~jos/sasp/>.
- [2] J. O. Smith III and X. Serra, "PARSHL: A program for the analysis/synthesis of inharmonic sounds based on a sinusoidal representation," in *Proceedings of the International Computer Music Conference*, Champaign–Urbana, IL, 1987, technical report version available in [14].
- [3] X. Serra, "A system for sound analysis / transformation / synthesis based on a deterministic plus stochastic decomposition," Ph.D. dissertation, Stanford University, Stanford, CA, October 1989.
- [4] X. Serra and J. O. Smith III, "Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, 1990.
- [5] X. Serra, *Musical sound modeling with sinusoids plus noise*. Routledge, 1997, pp. 91–122, appears in [30].
- [6] R. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 4, pp. 744–754, August 1986.
- [7] J. O. Smith III, *Mathematics of the Discrete Fourier Transform (DFT) with Audio Applications, Second Edition*, 2007 ed., online book, <https://ccrma.stanford.edu/~jos/mdft/mdft.html>.
- [8] A. Degani, M. Dalai, R. Leonardi, and P. Miglorati, "Time-frequency analysis of musical signals using the phase coherence," in *Proceedings of the International Conference on Digital Audio Effects (DAFx-13)*, vol. 16, Maynooth, Ireland, September 2–5 2013.
- [9] J. Nam, G. J. Mysore, J. Ganseman, K. Lee, and J. S. Abel, "A super-resolution spectrogram using coupled PLCA," in *Proceedings of the Conference of the International Speech Communication Association (Interspeech)*, vol. 11, Makuhari, Japan, September 26–30 2010, pp. 1696–1699.
- [10] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Transactions on Signal Processing*, vol. 43, no. 5, pp. 1068–1089, May 1995.
- [11] K. R. Fitz, "The reassigned bandwidth-enhanced method of additive synthesis," Ph.D. dissertation, University of Illinois at Urbana–Champaign, Urbana, IL, 1999.
- [12] K. Fitz and L. Haken, "On the use of time-frequency reassignment in additive sound modeling," *Journal of the Audio Engineering Society*, vol. 50, no. 11, pp. 879–893, November 2002.
- [13] F. Keiler and S. Marchand, "Survey on extraction of sinusoids in stationary sounds," in *Proceedings of the International Conference on Digital Audio Effects (DAFx-02)*, vol. 5, Hamburg, Germany, September 26–28 2002, pp. 51–58.
- [14] J. O. Smith III and X. Serra, "PARSHL: An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation," CCRMA, Department of Music, Stanford University, Stanford, CA, STAN-M 43, 1985.
- [15] Y. Goto, "Highly accurate frequency interpolation of apodized FFT magnitude-mode spectra," *Applied spectroscopy*, vol. 52, no. 1, pp. 134–138, 1998.
- [16] A. Chapman-Hatchett, V. Chohan, and T. E. d'Amico, "Tune measurement for the CERN proton synchrotron booster rings using DSP in VME," in *Proceedings of the Particle Accelerator Conference*, vol. 3, New York, NY, March 27 1999, pp. 2202–2204.
- [17] M. Gasior and J. L. Gonzalez, "Improving FFT frequency measurement resolution by parabolic and gaussian spectrum interpolation," in *Proceedings of the Beam Instrumentation Workshop*, vol. 11, Knoxville, TN, May 3–6 2004, pp. 276–285, also available as a technical note [31].
- [18] E. Meijering, "A chronology of interpolation: from ancient astronomy to modern signal and image processing," *Proceedings of the IEEE*, vol. 90, no. 3, pp. 319–342, March 2002.
- [19] M. Abe and J. O. Smith III, "design criteria for the quadratically interpolated FFT method (i): bias due to interpolation," CCRMA, Department of Music, Stanford University, Stanford, CA, STAN-M 114, October 13 2004.
- [20] M. C. McIntyre and D. A. Dermott, "A new fine-frequency estimation algorithm based on parabolic regression," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, San Francisco, CA, March 23–26 1992, pp. 541–544.
- [21] R. E. Boucher and J. C. Hassab, "Analysis of discrete implementation of generalized cross correlator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 3, pp. 609–611, June 1981.
- [22] A. De Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, 2002.
- [23] P. McLeod and G. Wyvill, "A smarter way to find pitch," in *Proceedings of International Computer Music Conference (ICMC)*, 2005, pp. 138–141.
- [24] P. McLeod, "Fast, accurate pitch detection tools for music analysis," Ph.D. dissertation, University of Otago, Dunedin, New Zealand, May 30 2008.
- [25] M. Abe and J. O. Smith III, "CQIFFT: Correcting bias in a sinusoidal parameter estimator based on quadratic interpolation of FFT magnitude peaks," CCRMA, Department of Music, Stanford University, Stanford, CA, STAN-M 117, October 13 2004.
- [26] B. Archer and E. W. Weisstein, "Lagrange interpolating polynomial," MathWorld—A Wolfram Web Resource, <http://mathworld.wolfram.com/LagrangeInterpolatingPolynomial.html>.
- [27] M. Abe and J. O. Smith III, "Criteria for the quadratically interpolated FFT method (ii): bias due to interfering components," CCRMA, Department of Music, Stanford University, Stanford, CA, STAN-M 115, October 13 2004.
- [28] —, "Design criteria for the quadratically interpolated FFT method (iii): Bias due to amplitude and frequency modulation," CCRMA, Department of Music, Stanford University, Stanford, CA, STAN-M 116, October 13 2004.
- [29] M. Gasior, "Improving frequency resolution of discrete spectra," Ph.D. dissertation, AGH University of Science and Technology, Kraków, Poland, 2006.
- [30] C. Roads, S. Pope, A. Picialli, and G. D. Poli, Eds., *Musical Signal Processing*. Routledge, January 1997.
- [31] M. Gasior and J. L. Gonzalez, "Improving FFT frequency measurement resolution by parabolic and gaussian interpolation," CERN—AB Division, Geneva, Switzerland, AB-Note 2004-021 BDI, February 2004.