CENTER FOR COMPUTER RESEARCH IN MUSIC AND ACOUSTICS
DEPARTMENT OF MUSIC, STANFORD UNIVERSITY
REPORT NO. STAN-M-118

AM/FM Rate Estimation and Bias Correction for Time-Varying Sinusoidal Modeling

October 13, 2004

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

Abstract

Due to its simplicity and accuracy, quadratic peak interpolation in a zero-padded Fast Fourier Transform has been widely used for sinusoidal parameter estimation in audio applications. In this paper, as its natural extension, we propose a method to estimate the first order amplitude and frequency modulation rates of time-varying sinusoidal components, as well as to correct biases in the amplitude, frequency and phase estimates. We first derive exact formulas to estimate the AM/FM rates for Gaussian windows. We then extend it to an approximate method for non-Gaussian windows by introducing a simple window adaptation scheme. Experimental results show the average biases in the estimates of the AM and FM rates with a 30ms Hann window are below 1%, for typical AM/FM rates in a speech.

1 Introduction

Sinusoidal modeling[1, 2, 3] has been widely used to represent the most salient aspects of tonal sound. A key component of sinusoidal modeling is the estimation of the parameters of multiple sinusoids. Among various approximate maximum likelihood (ML) methods[4, 5, 6, 7, 8], quadratic interpolation of magnitude peaks in a Fast Fourier Transform (FFT)[1] has been widely used due to its simplicity and accuracy, which is sufficient for most audio purposes.

However, most of the ML methods assume that the sinusoidal components are stable within their analysis frame. Since the sinusoidal components in natural audio are more or less modulated in both amplitude and frequency, we usually suffer from the well-known trade-off between time and frequency resolutions.

One natural approach to address this problem is to introduce a higher order model of the sinusoids. The simplest extension may be to add the first order AM and FM, which are sometimes referred as decay and chirp rates respectively, to the stable model. Marques[9] and Peeters[10] propose methods to estimate the AM/FM rates with Gaussian windows. Lagrange[11] utilizes empirical functions for the same purpose with non-Gaussian windows. Master and Liu[12] propose a method to estimate the chirp rate based on the derivatives of the phase spectrum of a windowed FFT.

Our approach in this paper is to extend the quadratic interpolation method for the time-varying case. We first derive exact formulas to estimate the AM/FM rates with a Gaussian window. They are based on the first and second derivatives of the peak magnitude and phase spectra, so that they can be easily calculated from the fitted parabolas. Formulas to correct biases in conventional sinusoidal parameters, i.e. amplitude, frequency and phase, are also derived. We then extend them to other types of windows by introducing a simple window adaptation scheme. We experimentally confirm the accuracy of the method for some well-used windows, such as Hann, Hamming and Blackman windows.

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]:

1. Calculate amplitude and phase spectrum of audio data, by using a zero-padded windowed FFT (points in Fig. 1).
2. Find the maximum peak magnitude (u_{k_0}).
3. Quadratically interpolate the log-amplitude of the peak using two neighboring samples (dotted line).
4. Estimate the frequency and amplitude from the interpolation ($\hat{\omega}_0$ and $\hat{\lambda}_0$).
5. Estimate the phase, if needed, by quadratically interpolating the phase spectrum based on the interpolated frequency estimate ($\hat{\phi}_0$).

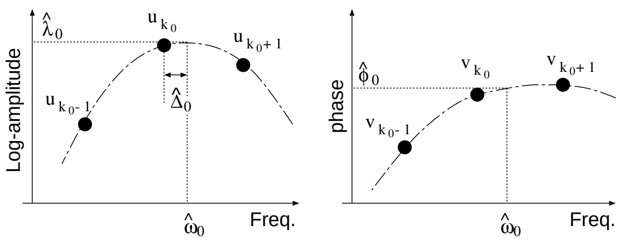

Figure 1: Quadratic interpolation of spectral peak

6. Remove the peak from the FFT data for subsequent processing.
7. Repeat the steps 2-6 above for each peak.

Note that since the log-magnitude and phase spectra of a Gaussian windowed sinusoid are parabolic, the interpolation is exact for a Gaussian window. For other windows, the estimates are biased by the difference between the true peak shape and the fitted parabolas [22].

3 AM/FM rates estimation for Gaussian windows

3.1 Fourier transform of a windowed AM/FM sinusoid

Let a sinusoid with the first order AM and FM be

$$x(t) = e^{\alpha_0 t+\lambda_0} e^{j(\beta_0 t^2+\omega_0 t+\phi_0)}, \tag{1}$$

where

$$\begin{aligned}
\omega_0 &: \text{ instantaneous frequency at } t = 0,\\
\lambda_0 &: \text{ instantaneous log-amplitude at } t = 0,\\
\phi_0 &: \text{ instantaneous phase at } t = 0,\\
\alpha_0 &: \text{ amplitude change rate (ACR)},\\
\beta_0 &: \text{ frequency change rate (FCR)}.
\end{aligned}$$

Let a normalized Gaussian window be

$$w(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}t^2} = \sqrt{\frac{p}{\pi}} e^{-pt^2}, \tag{2}$$

where σ is the standard deviation (= $\sqrt{1/e}$ width) of the Gaussian and $p \triangleq 1/2\sigma^2$. The windowed Fourier transform of the AM/FMed sinusoid can be calculated as[24]

$$X(\omega) = \int w(t)x(t)e^{-j\omega t}dt = e^{u(\omega)+jv(\omega)}, \tag{3}$$

where

$$u(\omega) = \lambda_0 + \frac{\alpha_0^2}{4p} - \frac{1}{4}\log(1+(\frac{\beta_0}{p})^2) - \frac{p}{4(p^2+\beta_0^2)}[\omega-\omega_0-\frac{\alpha_0\beta_0}{p}]^2 \tag{4}$$

is the log-amplitude term and

$$v(\omega) = \phi_0 + \frac{\alpha_0^2}{4\beta_0} + \frac{1}{2}\text{atan}(\frac{\beta_0}{p}) - \frac{\beta_0}{4(p^2+\beta_0^2)}[\omega-\omega_0+\frac{p\alpha_0}{\beta_0}]^2 \tag{5}$$

is the phase term[1]. Note that they are both parabolic functions of the frequency ω. These results are substantially the same as Marques[9] and Peeters[10].

3.2 Bias in amplitude, frequency and phase estimates

The frequency estimate in the QIFFT is the position of the maximum peak in the magnitude spectrum, as

$$\hat{\omega}_0 \triangleq \text{argmax } |X(\omega)| = \omega_0 + \frac{\alpha_0\beta_0}{p}. \tag{6}$$

The log-amplitude estimate is the log-magnitude spectrum at the peak, as

$$\hat{\lambda}_0 \triangleq u(\hat{\omega}_0) = \lambda_0 + \frac{\alpha_0^2}{4p} - \frac{1}{4}\log(1+(\frac{\beta_0}{p})^2), \tag{7}$$

and the phase estimate is the phase spectrum at the magnitude peak, as

$$\hat{\phi}_0 \triangleq v(\hat{\omega}_0) = \phi_0 - \frac{\alpha_0^2\beta_0}{4p^2} + \frac{1}{2}\text{atan}(\frac{\beta_0}{p}). \tag{8}$$

The above equations show that all the estimates in the QIFFT, as well as other methods for sinusoidal parameter estimation using spectral peaks, are biased by the AM and FM.

[1] Although we show here a formula for $\beta_0 \neq 0$, the following results are applicable to $\beta_0 = 0$ case.

3.3 AM/FM rates estimation

From Eqs.(4) and (5), the first and second derivatives of the magnitude and phase spectra at the magnitude peak are calculated as

$$\begin{aligned} v'(\hat{\omega}_0) &= -\frac{\alpha_0}{2p}, & (9)\\ u''(\hat{\omega}_0) &= -\frac{p}{2(p^2+\beta_0^2)}, & (10)\\ v''(\hat{\omega}_0) &= -\frac{\beta_0}{2(p^2+\beta_0^2)}. & (11) \end{aligned}$$

Using these relations, we can estimate α_0 and β_0 from the derivatives, as

$$\begin{aligned} \hat{\hat{\alpha}}_0 &\triangleq -2pv'(\hat{\omega}_0) & (12)\\ \hat{\hat{\beta}}_0 &\triangleq p\frac{v''(\hat{\omega}_0)}{u''(\hat{\omega}_0)}. & (13) \end{aligned}$$

In addition, we can estimate p as

$$\hat{p} \triangleq -\frac{u''(\hat{\omega}_0)}{2(u''^2(\hat{\omega}_0)+v''^2(\hat{\omega}_0))}. \tag{14}$$

This $\hat{p}$ may be used to estimate the second order AM, if it non-negligibly exists. Otherwise, $\hat{p}$ is equivalent to p for the Gaussian window.

3.4 Bias corrected estimation

The biases in the amplitude, frequency and phase estimates can be corrected by using the above $\hat{\hat{\alpha}}_0$ $\hat{\hat{\beta}}_0$ and p (or $\hat{p}$), as

$$\begin{aligned} \hat{\hat{\omega}}_0 &\triangleq \hat{\omega}_0 - \frac{\hat{\hat{\alpha}}_0\hat{\hat{\beta}}_0}{p}, & (15)\\ \hat{\hat{\lambda}}_0 &\triangleq \hat{\lambda}_0 - \frac{\hat{\hat{\alpha}}_0^2}{4p} + \frac{1}{4}\log(1+(\frac{\hat{\hat{\beta}}_0}{p})^2) & (16)\\ \hat{\hat{\phi}}_0 &\triangleq \hat{\phi}_0 + \frac{\hat{\hat{\alpha}}_0^2\hat{\hat{\beta}}_0}{4p} - \frac{1}{2}\mathrm{atan}(\frac{\hat{\hat{\beta}}_0}{p}). & (17) \end{aligned}$$

Note that since all the estimates are based only on up to the second derivatives of $u(\omega)$ and $v(\omega)$ at the magnitude peak, they can be calculated directly from the fitted parabolas in the QIFFT (see appendix A).

4 Extension to non-Gaussian windows

4.1 Direct application of the method for a Gaussian window

The QIFFT can be seen as approximating the nearly parabolic shape of the spectral peak of a non-Gaussian window with the truly parabolic shape of a Gaussian window. Therefore, a simple solution for the AM/FM estimation with non-Gaussian windows may be direct application of the above results. However, since p is an unknown, hypothetical parameter for a non-Gaussian window, we need to find somehow an equivalent value for the window. Here, we simply replace p by $\hat{p}$ which can be obtained from the spectral data themselves. We refer this method as the "direct method".

4.2 Adapted method

As we will see later in the experiments, we can estimate α_0 and β_0 and correct biases in ω_0, λ_0 and ϕ_0 using the direct method to a certain accuracy. However, especially in the estimate of β_0, there exist a large bias even when the β_0 is reasonably small. This is because the difference between Gaussian and non-Gaussian windows affects to the FCR estimate non-negligibly.

To reduce the FCR bias as well as the remaining biases in the other estimates, we propose a simple window adaptation scheme. The idea is to introduce adjusting coefficients (ζ_1, ..., ζ_9) to each term in Eqs.(12),(13), (15), (16)

Table 1: Window adaptation coefficients

	Hann	Hamming	Blackman
ζ_1	0.995354	0.995258	0.997809
ζ_2	0.169257	0.132051	0.103745
ζ_3	1.393056	1.285090	1.210194
ζ_4	0.442406	0.343335	0.230884
ζ_5	-0.717980	-0.779015	-0.826779
ζ_6	-0.251620	-0.234583	-0.246220
ζ_7	0.177511	0.186698	0.202421
ζ_8	0.158120	0.197343	0.183014
ζ_9	-0.503299	-0.502182	-0.499939

For multiple regression analysis, we used 56,000 AM/FMed sinusoids whose parameters are randomly given. Uniform random variables whose ranges are $[0, \pi]$, and $[-\pi, \pi]$ are used for ω_0, ϕ_0, respectively. λ_0 is fixed to 0.0. For α_0 and β_0, Gaussian random variables whose means are 0 and standard deviations are $\sigma_\alpha = 0.3/M$ and $\sigma_\beta = 4.0/M^2$ respectively are used, where M denotes a window length in samples. These coefficients can be used for a wide range of the FFT sizes (≥ 256) and window lengths (≥ 31), regardless of sampling rates.

Table 2: Signal Parameters

parameter	distribution	unit	range
ω_0	uniform	Hz	$[1\text{k}, 15\text{k}]$
$a_0 (= \exp(\lambda_0))$	uniform	—	$[1, 16]$
ϕ_0	uniform	rad	$[-\pi, \pi]$
α_0	Gaussian	s^{-1}	$N(0, 10)$
β_0	Gaussian	rad/s^2	$N(0, 2\pi 1000)$

and (17), as

$$
\begin{aligned}
\check{\alpha}_0 &= \zeta_1\ \hat{\hat{\alpha}}_0 + \zeta_2\ \hat{\Delta}^2 \hat{\hat{\alpha}}_0, & (18)\\
\check{\beta}_0 &= \zeta_3\ \hat{\hat{\beta}}_0 + \zeta_4\ \hat{\Delta}_0 \hat{\hat{\alpha}}_0, & (19)\\
\check{\omega}_0 &= \hat{\omega}_0 + \zeta_5\ \frac{\check{\alpha}_0 \check{\beta}_0}{\hat{\hat{p}}}, & (20)\\
\check{\lambda}_0 &= \hat{\lambda}_0 + \zeta_6\ \frac{\check{\alpha}_0^2}{\hat{\hat{p}}} + \zeta_7\ \log(1 + (\frac{\check{\beta}_0}{\hat{\hat{p}}})^2), & (21)\\
\check{\phi}_0 &= \hat{\phi}_0 + \zeta_8\ \frac{\check{\alpha}_0^2 \check{\beta}_0}{\hat{\hat{p}}} + \zeta_9\ \text{atan}(\frac{\check{\beta}_0}{\hat{\hat{p}}}). & (22)
\end{aligned}
$$

The coefficients are numerically determined by multiple regression analysis, as shown in Table 1. We introduced here a small trick to reduce biases caused by the rough frequency sampling in the FFT. Two additional terms $\hat{\Delta}^2 \hat{\hat{\alpha}}_0$ and $\hat{\Delta}_0 \hat{\hat{\alpha}}_0$ are added to the ACR and FCR estimates respectively, where $\hat{\Delta}_0$ denotes the frequency offset of the peak of the fitted parabola from the nearest FFT bin (Fig. 1). These terms are effective especially when a small (e.g. less than 5) zero-padding factor is used. We refer this method as the "adapted method".

5 Experiments

For the following experiments, we prepared 1000 sinusoids sampled at 44.1kHz whose parameters are randomly given, as shown in Table 2. The ACRs are Gaussian distributed random values whose standard deviation is set to a magnification or decay of 10 times per second. The FCRs are also Gaussian distributed random values whose standard deviation is set to a frequency shift of 1kHz per second. These values are roughly equivalent to the ACRs and FCRs in a human speech[24]. The frequencies, amplitudes and phases are uniformly distributed random values. We evaluate the biases by the ratios to the standard values, which are $\omega_s = 2\pi 100[\text{rad/s}]$, $a_s = a_0$, $\phi_s = \pi[\text{rad}]$, $\alpha_s = 10[\text{s}^{-1}]$, and $\beta_s = 2\pi 1000[\text{rad/s}^2]$.

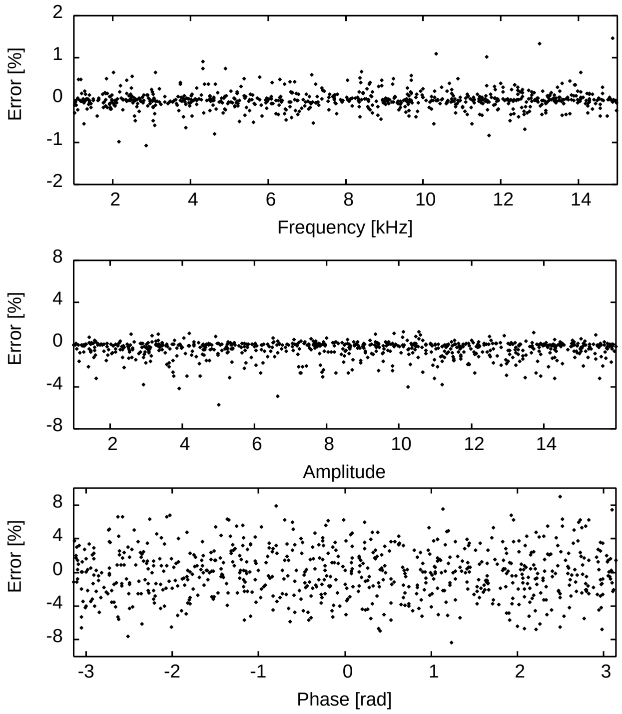

Figure 2: Scatter plots of the biases in the QIFFT: frequency (top), amplitude (middle), and phase (bottom). The horizontal axes show given parameter values and the vertical axes show experimentally obtained biases in the estimates.

5.1 Bias in the standard QIFFT

We plotted the biases in the standard QIFFT with the 30ms Hann window for all the 1000 signals in Fig. 2. We can confirm that maximally a few percents of biases in all the frequency, amplitude and phase estimates exist. We can also confirm that the biases do not depend on the frequency, amplitude and phase themselves.

5.2 Bias in the direct method

The similar plots of the biases in the direct method are shown in Fig. 3. Comparing with Fig. 2, we can confirm that the biases in the frequency, amplitude and phase estimates are reduced to below or around 1%. The ACR bias is around 3%, whereas the FCR bias is maximally over 100%. This is because the FCR bias is correlated to the FCR itself, which can be seen as the slope in Fig. 3(e).

5.3 Bias in the adapted method

The similar plots of the biases in the adapted method are shown in Fig. 4. Comparing with Fig. 2 and Fig. 3, we can confirm the biases in the frequency, amplitude and phase are yet more reduced to below 0.1%. The bias in the ACR is slightly reduced to around 1%, and that in the FCR is greatly reduced to below 1%. The slope in the FCR bias in Fig. 3 is corrected by the multiplicative coefficient to $\hat{\hat{\beta}}_0$, and the widths of the distribution is corrected by the second term.

5.4 Effect of window length

Fig. 5 and Table 3 show the biases in the three methods with Hann window of various window lengths. We can confirm that the adapted method gives the best estimates for most of the parameters and window lengths. We can also see that a longer window in general worsens the biases. This is because we only linearly correct the difference between Gaussian and non-Gaussian windows, while the difference are essentially nonlinear.

If we accept a few percents of estimation errors, the frequency, amplitude and phase estimates in the adapted method can be reliably used up to the window length of 90ms. The ACR and FCR estimates, however, can be used only up to 45ms.

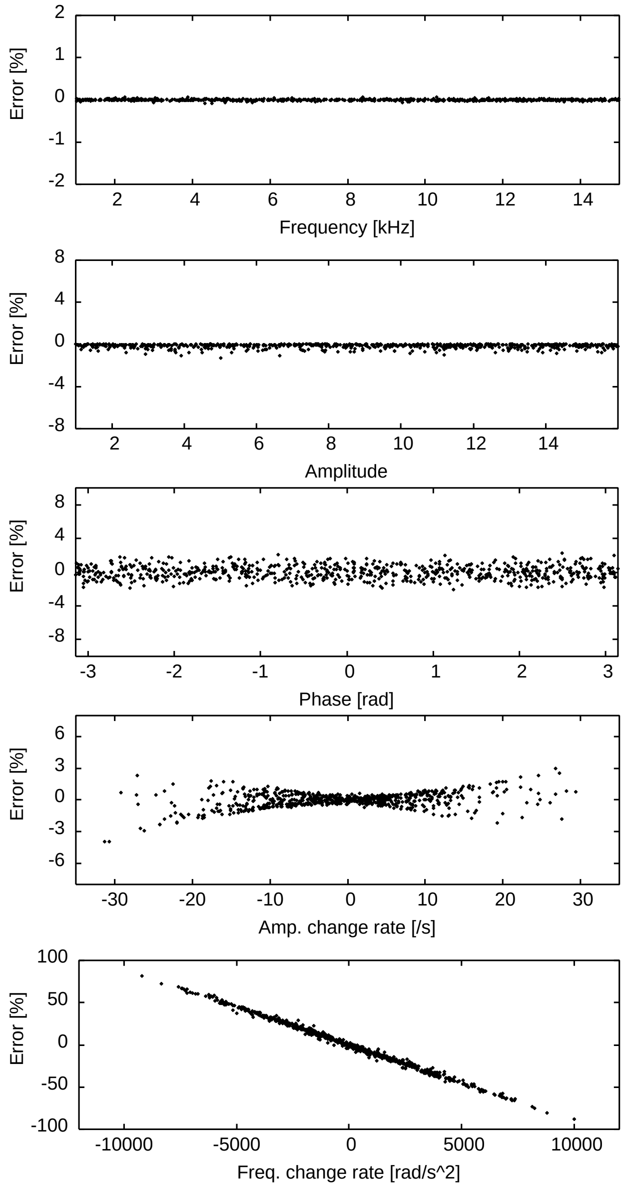

Figure 3: Scatter plots of the biases in the direct method: (a) frequency ($\hat{\omega}_0$), (b) amplitude ($\hat{a}_0 = \exp(\hat{\hat{\lambda}}_0)$), (c) phase ($\hat{\hat{\phi}}_0$). (d) amplitude change rate ($\hat{\alpha}_0$). (e) frequency change rate ($\hat{\hat{\beta}}_0$).

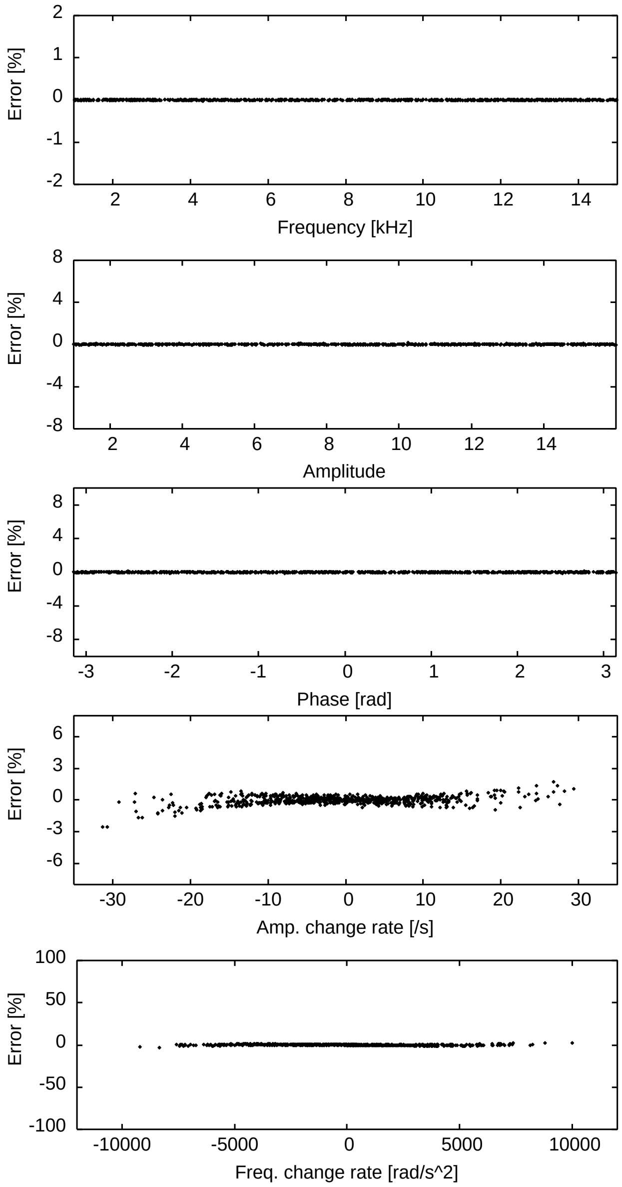

Figure 4: Scatter plots of the biases in the adapted method: (a) frequency ($\check{\omega}_0$), (b) amplitude ($\check{a}_0 = \exp(\check{\lambda}_0)$), (c) phase ($\check{\phi}_0$). (d) amplitude change rate ($\check{\alpha}_0$). (e) frequency change rate ($\check{\beta}_0$).

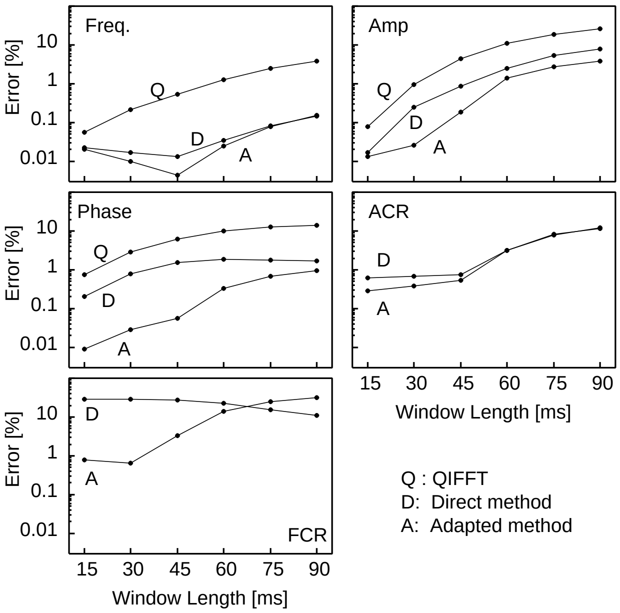

Figure 5: RMS biases with Hann windows of various lengths.

5.5 Comparison between windows

Figure 6 and Table 4 show comparison of the biases for various types of windows of 30ms. We can confirm that the direct and adapted methods work with the other types of windows as well. We can also see the biases in the Blackman window are smallest, whereas those in the Hamming window is the largest. It corresponds to the width of the main lobes of these windows.

6 Discussion and Summary

In this paper, we proposed a method to estimate AM/FM rates and correct biases in frequency, amplitude and phase estimates based on the QIFFT. Major advantages of this method are 1) applicable to well-used, non-Gaussian windows, 2) computationally efficient, and 3) enough accurate for many audio applications.

We show all the biases for the 30ms Hann window are below 1% for typical AM/FM rates in a speech, and the biases in the bias-corrected estimates of the frequency, amplitude and phase are below a few percents for up to the 90ms windows.

For further improvement, nonlinear bias correction functions may be used. Our preliminary experiments show sigmoid-like functions may be well fit to a wide range of the AM/FM biases.

References

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

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

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

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

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

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

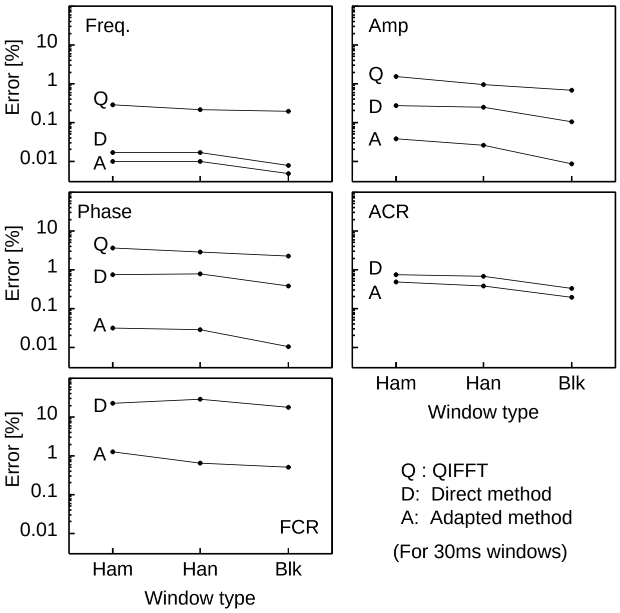

Figure 6: RMS biases with various windows of 30ms.

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

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

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

[10] G. Peeters and X. Rodet: "SINOLA: A New Analysis/Synthesis Method using Spectrum Peak Shape Disortion, Phase and Redassigned Spectrum," Proc. ICMC'99 (1999).

[11] M. Lagrange, S. Marchand and J.-B. Rault: "Sinusoidal Parameter Extraction and Component Selection in a NonStationary Model," Proc. Int. Conf. on Digital Audio Effects (DAFX-02), 59/64 (2002).

[12] A. S. Master and Y. Liu: "Nonstationary Sinusoidal Modeling with Efficient Estimation of Linear Frequency Chirp Parameters," Proc. IEEE ICASSP'03, 656/659 (2003).

[13] 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).

[14] 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).

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

[16] 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).

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

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

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

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

Table 3: Dependency to window sizes (Hann Window) (Q: QIFFT, D: Direct method, A: Adapted method.) The same signals as in Fig. 2 are used.

	Freq.	Amp	Phase	ACR	FCR
M = 661 (15ms), N = 2048					
Q	0.057%	0.076%	0.74%	—	—
D	0.023%	0.016%	0.21%	0.61%	29%
A	0.020%	0.013%	0.0091%	0.28%	0.80%
M = 1323 (30ms), N = 4096					
Q	0.22%	0.97%	2.9%	—	—
D	0.017%	0.25%	0.80%	0.69%	29%
A	0.0098%	0.026%	0.029%	0.39%	0.64%
M = 1985 (45ms), N = 8192					
Q	0.53%	4.48%	6.26%	—	—
D	0.013%	0.86%	1.53%	0.73%	27%
A	0.0045%	0.19%	0.055%	0.52%	3.29%
M = 2645 (60ms), N = 8192					
Q	1.25%	11%	9.8%	—	—
D	0.035%	2.5%	1.84%	3.21%	22%
A	0.025%	1.4%	0.33%	3.17%	14%
M = 3007 (75ms), N = 8192					
Q	2.42%	19%	13%	—	—
D	0.084%	5.4%	1.8%	8.0%	16%
A	0.078%	2.8%	0.68%	8.2%	25%
M = 3969 (90ms), N = 16384					
Q	3.78%	26%	14%	—	—
D	0.15%	7.7%	1.7%	12%	11%
A	0.15%	3.8%	0.97%	12%	31%

[21] A. S. Master: "Sinusoidal Modeling Parameter Estimation Via a Dynamic Channel Vocoder Model," Proc. IEEE ICASSP'02, 1857/1860 (2002).

[22] 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).

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

[24] 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).

[25] 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).

A Estimation from discrete spectra

In the QIFFT, the continuous log-magnitude and phase spectra around a peak are aproximated by quadratic interpolation of discrete FFT spectra. Since all the estimates in this paper depend only on up to the second order derivatives of the spectra, they can be obtained from the fitted parabora. In this appendix, we explicitly show the estimation formulas from the discrete samples.

Let k_0 be the FFT bin number at the maximum peak in FFT magnitude spectrum, u_{k_0} be the log-magnitude at the bin, u_{k_0-1} and u_{k_0+1} be the two neighbours of it, and v_{k_0}, v_{k_0-1}, and v_{k_0+1} be the corresponding phases (Fig. 1).

Then, the continuous log-maginitude and phase spectra are interpolated as

$$\begin{aligned} \hat{u}(k) &= a(k-k_0)^2 + b(k-k_0) + c \\ \hat{v}(k) &= d(k-k_0)^2 + e(k-k_0) + f, \end{aligned} \tag{23}$$

Table 4: Comparison between windows (M = 1323 (30ms), N = 4096).

Hann					
Q	0.22%	0.97%	2.9%	—	—
D	0.017%	0.25%	0.80%	0.69%	29%
A	0.0098%	0.026%	0.029%	0.39%	0.64%
Hamming					
Q	0.29%	1.6%	3.6%	—	—
D	0.017%	0.27%	0.76%	0.75%	23%
A	0.010%	0.038%	0.032%	0.48%	1.3%
Blackman					
Q	0.19%	0.68%	2.3%	—	—
D	0.0080%	0.11%	0.39%	0.33%	18%
A	0.0048%	0.0086%	0.011%	0.19%	0.52%

where k denotes (continuously extended) FFT bin number and

$$\begin{aligned} a &= (u_{k_0+1} - 2u_{k_0} + u_{k_0-1})/2 \\ b &= (u_{k_0+1} - u_{k_0-1})/2 \\ c &= u_{k_0} \\ d &= (v_{k_0+1} - 2v_{k_0} + v_{k_0-1})/2 \\ e &= (v_{k_0+1} - v_{k_0-1})/2 \\ f &= v_{k_0}. \end{aligned} \tag{24}$$

Then Eqs.(6), (7), (8), (14), (12) and (13) can be approximated as

$$\begin{aligned} \hat{p} &= -(\frac{\pi}{N})^2 \frac{d}{a^2 + d^2}, \\ \hat{\Delta}_0 &= -\frac{b}{2a}, \\ \hat{\omega}_0 &= \frac{2\pi}{N}(k_0 + \hat{\Delta}_0), \\ \hat{\lambda}_0 &= a\hat{\Delta}_0^2 + b\hat{\Delta}_0 + c, \\ \hat{\phi}_0 &= d\hat{\Delta}_0^2 + e\hat{\Delta}_0 + f, \\ \hat{\alpha}_0 &= -\frac{N}{\pi} p(2d\hat{\Delta}_0 + e), \\ \hat{\beta}_0 &= p\frac{d}{a}. \end{aligned} \tag{25}$$