AM/FM RATE ESTIMATION FOR TIME-VARYING SINUSOIDAL MODELING

Mototsugu Abe

SONY Corporation
abe@av.crl.sony.co.jp

Julius O. Smith III

CCRMA, Music Department, Stanford University
jos@ccrma.stanford.edu

ABSTRACT

Due to its simplicity and accuracy, quadratic peak interpolation in a zero-padded Fast Fourier Transform (FFT) has been widely used for sinusoidal parameter estimation in audio applications. 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 conventional amplitude, frequency and phase estimates. We derive exact formulas for Gaussian windows and obtain approximate formulas for often-used windows by introducing a simple window adaptation scheme. Experimental results show the average estimation biases of the AM and FM rates with a 30ms Hann window are below 1% for typical AM/FM rates in speech.

1. INTRODUCTION

Sinusoidal modeling [1] 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 [2], 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 approximate ML methods, including quadratic interpolation, generally assume stability of sinusoidal components within their analysis frame. Since sinusoidal components in natural audio are more or less modulated in both amplitude and frequency, we usually suffer from a well-known trade-off between time and frequency resolutions. One natural approach to address this problem is to introduce a higher order sinusoidal model. The simplest extension may be to add the first order AM and FM terms to a stable model. For example, Marques [3] and Peeters [4] propose methods to estimate the AM/FM rates with Gaussian windows. Lagrange [5] proposes an empirical method with non-Gaussian windows. Although their methods are effective for some limited applications, a more general, mathematically consistent and computationally efficient method seems to be needed.

Our approach is based on the quadratic interpolation method and extends it to the time-varying case. We first derive exact formulas for AM/FM rate estimation using Gaussian windows. They are obtained from the first and second derivatives of the magnitude and phase spectra of the FFT, so that they can be easily calculated from the fitted quadratic polynomials. Bias correction functions for the conventional sinusoidal parameters, i.e. amplitude, frequency and phase, are also derived at the same time. We then extend them to other often-used windows by introducing a simple window adaptation scheme. We experimentally confirm the accuracy of the method with Hann windows.

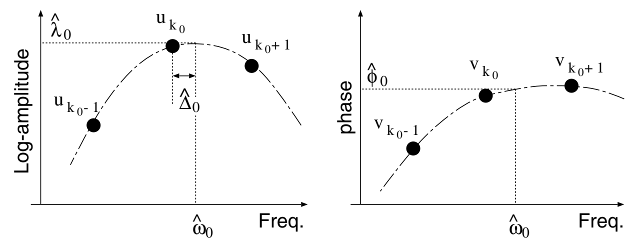

Fig. 1. Quadratic interpolation of spectral peak

2. QIFFT METHOD

The Quadratically Interpolated FFT (QIFFT) method for estimating sinusoidal parameters from peaks in spectral magnitude data can be summarized as follows:

1. Calculate 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 (log-)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 ($\hat{\phi}_0$).
6. Remove the peak from the FFT data for subsequent processing.
7. Repeat the steps 2-6 above for each peak.

3. AM/FM RATE ESTIMATION: GAUSSIAN WINDOWS

3.1. Fourier transform of a windowed AM/FM sinusoid

Let a sinusoid with first-order AM and FM be written as

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

where

ω_0	:	instantaneous frequency at $t = 0$,
λ_0	:	instantaneous log-amplitude at $t = 0$,
ϕ_0	:	instantaneous phase at $t = 0$,
α_0	:	amplitude change rate (ACR),
β_0	:	frequency change rate (FCR).

Let a normalized Gaussian window be written as

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

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

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

where

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

is the log-amplitude term and

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

is the phase term. These results are substantially the same as in Marques [3] or Peeters [4]. Note that since the log-amplitude and phase are both parabolic functions of the frequency ω, quadratic interpolation is *exact* for a Gaussian window.

3.2. Biases in amplitude, frequency and phase

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

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

the log-amplitude estimate is the log-magnitude value at the peak,

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

and the phase estimate is the phase value at the magnitude peak,

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

(The single hats denote the estimates from the original QIFFT.) The second and third terms are non-zero when AM and FM exist, which means all three QIFFT estimates are biased.

3.3. Estimating AM/FM rates

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

$$v'(\hat{\omega}_0) = -\frac{\alpha_0}{2p}, \tag{9}$$

$$u''(\hat{\omega}_0) = -\frac{p}{2(p^2+\beta_0^2)}, \tag{10}$$

$$v''(\hat{\omega}_0) = -\frac{\beta_0}{2(p^2+\beta_0^2)}. \tag{11}$$

Using these relations, α_0 and β_0 can be estimated as

$$\hat{\hat{\alpha}}_0 \triangleq -2pv'(\hat{\omega}_0) \tag{12}$$

$$\hat{\hat{\beta}}_0 \triangleq p\frac{v''(\hat{\omega}_0)}{u''(\hat{\omega}_0)}. \tag{13}$$

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

(The double hats denote the estimates of the AM/FM model.) In addition, we can estimate p as

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

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

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 $\hat{\hat{p}}$), as

$$\hat{\hat{\omega}}_0 \triangleq \hat{\omega}_0 - \frac{\hat{\hat{\alpha}}_0\hat{\hat{\beta}}_0}{p}, \tag{15}$$

$$\hat{\hat{\lambda}}_0 \triangleq \hat{\lambda}_0 - \frac{\hat{\hat{\alpha}}_0^2}{4p} + \frac{1}{4}\log\left[1 + \left(\frac{\hat{\hat{\beta}}_0}{p}\right)^2\right], \tag{16}$$

$$\hat{\hat{\phi}}_0 \triangleq \hat{\phi}_0 + \frac{\hat{\hat{\alpha}}_0^2\hat{\hat{\beta}}_0}{4p} - \frac{1}{2}\mathrm{atan}\left(\frac{\hat{\hat{\beta}}_0}{p}\right). \tag{17}$$

Note that since all the estimates are based on at most the second derivatives of $u(\omega)$ and $v(\omega)$ at the magnitude peak, they can be easily calculated from the fitted quadratic polynomial.

4. EXTENSION TO NON-GAUSSIAN WINDOWS

4.1. Direct method

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 the p by $\hat{\hat{p}}$ which can be obtained from the spectral data. We refer to this method as the "direct method".

4.2. Adapted method

As we will see later in the experiments, using the direct method, we can estimate α_0 and β_0 and correct biases in ω_0, λ_0 and ϕ_0 to certain accuracies. However, especially in the estimate of β_0, there

Table 2. Signal Parameters for experiments.

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	$N(0, 2\pi 1000)$

exists a large bias even when the β_0 is reasonably small. This is because the difference between Gaussian and non-Gaussian windows strongly influences the estimates. To reduce the FCR bias as well as remaining biases in the other estimates, we introduce adjusting coefficients (ζ_1, ..., ζ_9) to each term in the estimates by the direct method, as

$$\check{\alpha}_0 = \zeta_1 \hat{\hat{\alpha}}_0 + \zeta_2 \hat{\Delta}^2 \hat{\hat{\alpha}}_0, \tag{18}$$

$$\check{\beta}_0 = \zeta_3 \hat{\hat{\beta}}_0 + \zeta_4 \hat{\Delta}_0 \hat{\hat{\alpha}}_0, \tag{19}$$

$$\check{\omega}_0 = \hat{\omega}_0 + \zeta_5 \frac{\check{\alpha}_0 \check{\beta}_0}{\hat{\hat{p}}}, \tag{20}$$

$$\check{\lambda}_0 = \hat{\lambda}_0 + \zeta_6 \frac{\check{\alpha}_0^2}{\hat{\hat{p}}} + \zeta_7 \log\left[1 + \left(\frac{\check{\beta}_0}{\hat{\hat{p}}}\right)^2\right], \tag{21}$$

$$\check{\phi}_0 = \hat{\phi}_0 + \zeta_8 \frac{\check{\alpha}_0^2 \check{\beta}_0}{\hat{\hat{p}}} + \zeta_9 \operatorname{atan}\left(\frac{\check{\beta}_0}{\hat{\hat{p}}}\right). \tag{22}$$

The coefficients ζ_i are numerically determined by multiple regression analysis (MRA), as shown in Table 1. For the MRA, we use 56,000 AM/FMed sinusoids whose parameters are randomly given. For ω_0 and ϕ_0, we use uniform random values whose ranges are $[0, \pi]$ and $[-\pi, \pi]$ respectively. λ_0 is fixed to 0.0. For α_0 and β_0, we use Gaussian random values whose means are 0 and standard deviations are $\sigma_\alpha = 0.3/M$ and $\sigma_\beta = 4.0/M^2$, respectively, where M denotes a window length in samples.

In order to reduce biases due to coarse frequency sampling in the FFT, two extra 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 especially effective when a small (e.g. less than 5) zero-padding factor is used. We refer to this method as the "adapted method".

5. EXPERIMENTS

For the following experiments, we prepare 1000 sinusoids sampled at 44.1kHz whose parameters are randomly given. The ACRs and FCRs are Gaussian random values and the frequencies, amplitudes and phases are uniform random values whose ranges are shown in Table 2. These values are roughly equivalent to the ACRs and FCRs in human speech [7]. To be intuitively intelligible, the biases are normalized by their standard values, which are $\omega_s = 2\pi 100$[rad/s], $a_s = a_0$, $\phi_s = \pi$[rad], $\alpha_s = 10$[s^{-1}], and $\beta_s = 2\pi 1000$[rad/s^2].

5.1. Bias in the standard QIFFT

We plot the frequency bias in the standard QIFFT with a 30ms Hann window and 186ms FFT (8192 points) for all the 1000 signals in Fig. 2. We can confirm that the bias is on the order of a percent or less, and the bias does not depend on the sinusoidal frequency itself. Note that for a stable sinusoidal component, interpolation bias with this zero-padding factor would be below 0.01% [6]. Increased zero padding does not eliminate this AM/FM bias. The amplitude and phase biases reveal similar trends.

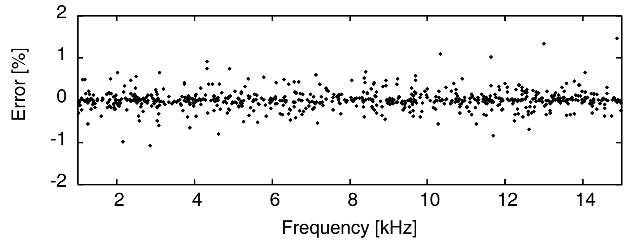

Fig. 2. Scatter plots of the frequency bias ($\hat{\omega}_0$) in the QIFFT.

5.2. Bias in the direct method

The biases in the frequency, ACR and FCR estimates by the direct method are shown in Fig. 3. Comparing with Fig. 2, we can confirm that the maximum bias in the frequency is reduced to below 0.1%. 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 (bottom).

5.3. Bias in the adapted method

The same plots for the adapted method are shown in Fig. 4. We can confirm that the frequency bias is further reduced to below 0.02%, the ACR bias is slightly reduced to around 2%, and the FCR bias is greatly reduced to around 2%. The slope in the FCR bias in Fig. 3 is corrected by the multiplicative coefficient to $\hat{\hat{\beta}}_0$. The width of the distribution of the ACR and FCR biases are narrowed by the extra terms in Eqs.(18) and (19).

5.4. Comparison of different window lengths

Figure 5 shows the biases in the three methods with Hann windows of various lengths. The FFT sizes are set to the minimum power-of-twos greater than 5 times the window lengths. We can confirm that the adapted method gives the best estimates for most of the sinusoidal parameters and window lengths. We can also see that a longer window in general worsens the biases. This is because we linearly correct the difference between Gaussian and non-Gaussian windows, while the differences are essentially nonlinear. For the same reason, we see that FCR estimates by the direct method surpass those by the adapted method for windows longer than 75ms for this FFT length.

If we accept on the order of one percent error, the frequency, amplitude and phase estimates in the adapted method can be reliably used up to window lengths of 90ms. The ACR and FCR estimates, however, can be used only up to 45ms.

6. 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

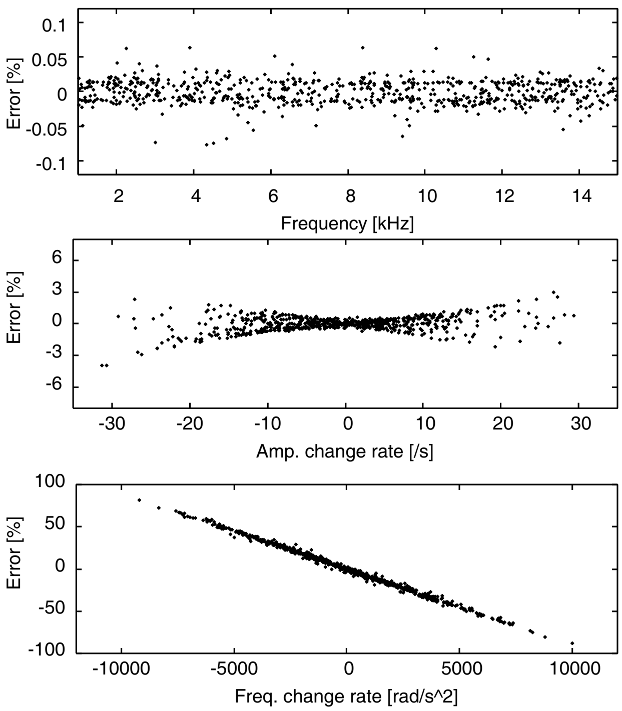

Fig. 3. Scatter plots of the biases in the direct method: frequency ($\hat{\omega}_0$) (top), ACR ($\hat{\alpha}_0$) (middle), FCR ($\hat{\hat{\beta}}_0$) (bottom). Note that the vertical scale in the top figure is different from Fig. 2.

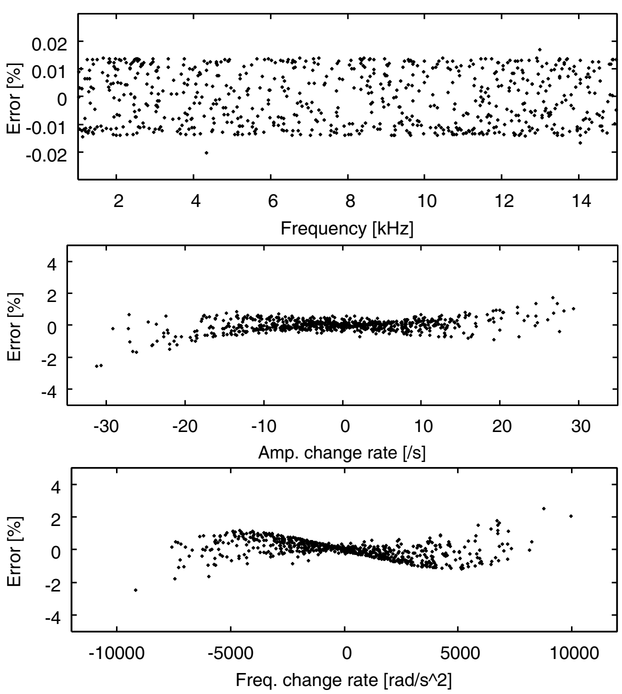

Fig. 4. Scatter plots of the biases in the adapted method: frequency ($\hat{\omega}_0$) (top), ACR ($\hat{\alpha}_0$) (middle), FCR ($\hat{\hat{\beta}}_0$) (bottom). Note that the vertical scales are different from Fig. 3.

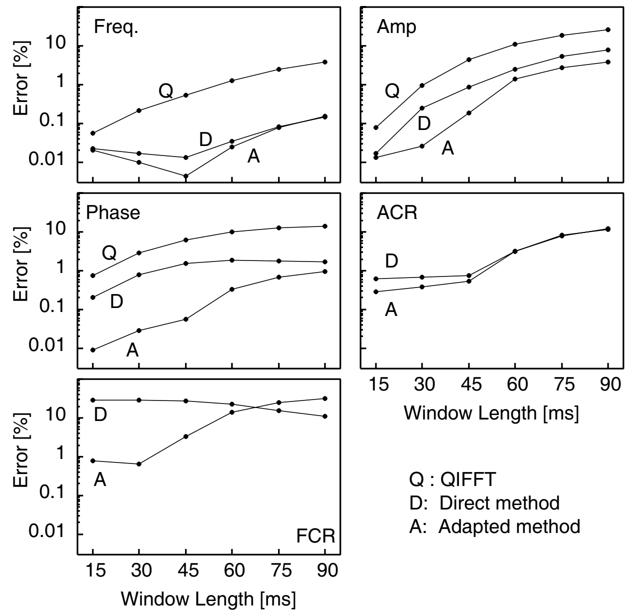

Fig. 5. RMS biases with Hann windows of various lengths.

to commonly used (non-Gaussian) windows, 2) computationally efficient (since no iterative operation is needed), and 3) accurate enough for many audio applications. For further improvement, nonlinear bias correction functions may be used. Our preliminary experiments show sigmoid-like functions may be used to fit a wide range of the AM/FM biases.

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

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

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

[5] M. Lagrange et. al.: "Sinusoidal Parameter Extraction and Component Selection in a Non Stationary Model," Proc. Int. Conf. on Digital Audio Effects (DAFX-02), 59/64 (2002).

[6] M. Abe and J. O. Smith III: "Design Criteria for Simple Sinusoidal Parameter Estimation based on Quadratic Interpolation of FFT Magnitude Peaks," Proc. AES 117th Convention, October, (2004).

[7] 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, (2004).