Estimation of partial parameters for non stationary sinusoids

Axel Roebel
IRCAM-CNRS-STMS, Analysis/Synthesis Team
roebel@ircam.fr

Abstract

The following paper deals with the estimation of partial parameters for non stationary sinusoids. First the existing bias for the analysis of non stationary sinusoids in a standard estimator is discussed. Then a new approach to bias reduction is proposed that consists of frequency slope estimation and demodulation to reduce the bias of the standard parameter estimator. The new approach does not require the use of Gaussian analysis windows. We present an experimental evaluation that compares the new parameter estimation scheme with previously existing methods. The results demonstrate that the bias is significantly reduced to a level that is similar or lower than the bias that exists for Gaussian analysis windows. The parameter range for which significant bias reduction can be achieved is increased.

1 Introduction

The sinusoidal or additive signal model is widely used for signal analysis and/or signal transformation of speech and music sound signals (Amatriain, Bonada, Loscos, and Serra 2002). The estimation of the sinusoidal parameters (amplitude, frequency and phase) from the spectral peaks of the Fourier transform of the signal is more or less straightforward as long as the sinusoids are stationary. For non stationary sinusoids the parameter estimation is significantly more difficult. It is known that the standard algorithm for stationary sinusoids provides results with non negligible bias and a number of algorithms have been proposed to solve this problem. Nearly all of these algorithms rely on the fact that the analysis window is approximately Gaussian such that analytic investigation becomes tractable (Marques and Almeida 1986; Peeters and Rodet 1999; Abe and Smith 2005). For non Gaussian windows (Abe and Smith 2005) proposes a linear adaptation of the bias correction functions that originally were derived for a Gaussian window. Complete non stationary parameter trajectories can be obtained with an adaptive algorithm (Röbel 2006). Unfortunately the adaptive procedure is extremely costly.

The following article investigates into a new strategy to reduce the estimation bias of the standard estimator. Following the argumentation below the frequency slope is the key to bias removal. If the frequency slope is properly estimated it

can be used to demodulate the spectral peak under investigation such that the standard estimator can be applied. Therefore, we propose to use the approximate frequency slope estimator of (Abe and Smith 2005) to estimate the frequency slope and use the demodulation scheme and then the standard sinusoidal estimator to find the fundamental sinusoidal parameters. Note, that for a sinusoidal signal model, the slope parameters are generally not required, because the parameter variations are automatically created by means of interpolation of the parameters of subsequent frames.

The organization of the article is as follows. In section **2** we will argue that the bias of the standard estimators is related to the frequency slope, only. In section **3** we will describe the frequency slope estimator to be used as well as the demodulation scheme. In section **4** we present the experimental results and in section **5** we conclude with an outlook on further improvements.

2 Estimation bias

The signal model that will be used in the following assumes a linear evolution for amplitude and frequency trajectories. Accordingly a discrete time sinusoid can be represented as

$$s(n) = (A + an)\exp(i(\phi + \omega n + Dn^2)). \qquad (1)$$

Here A is the mean amplitude of the signal and a is the amplitude slope. ϕ is the phase of the sinusoid at time $n = 0$, ω is its mean frequency and D is the frequency slope.

Assume that the analysis window is positioned such that its center is located at time 0. The ideal estimator should provide (A, ω, ϕ) as estimates for amplitude frequency and phase. As mentioned already we will not be concerned with the estimation of the amplitude slope because in a piecewise linear amplitude trajectory model the frame center amplitude values are sufficient to completely describe the amplitude trajectory.

As standard sinusoidal parameter estimator we consider the quadratically interpolating (or QIFFT) method summarized in (Abe and Smith 2005). This standard estimator has two sources of bias, first the use of a second order model for interpolating the spectral bins is systematically wrong for all but Gaussian windows and exponential amplitude evolution. This interpolation bias can be reduced by means of zero padding

the analysis window. The second source of bias is related to the fact that for the sinusoidal model eq. (1) the selection of the amplitude maximum of the second order amplitude interpolation will create biased estimates for phase and amplitude whenever the frequency slope $D \neq 0$. The frequency estimate will be biased only if frequency and amplitude slope are non zero. This result has been shown for Gaussian windows in (Peeters and Rodet 1999). By means of simple symmetry considerations we may prove that similar relations hold true for all symmetric analysis windows.

Consider the use of the continuous Fourier transform (FT) and the case $D = 0$. For this case it can be easily seen that due to the constant frequency the peak maximum will always be located at the frequency ω of the sinusoid. The sinusoid in eq. (1) can now be split into two sinusoids one with constant amplitude ($a = 0$) and the other with average amplitude being zero ($A = 0$). The FT of the first part provides the desired values A and ϕ when evaluated at frequency ω. The FT of the second part, due to anti-symmetry of the amplitude, will be zero. Therefore the FT provides correct estimates for all parameters as long as $D = 0$. For additive modeling the bias related to phase an frequency seems particularly important because as long as phase an frequency are properly selected the DFT amplitude will produce a minimum energy residual even if the amplitude evolution does not match the model. A real match between amplitude model and the signal will rarely be the case for real world audio signals.

As a consequence we can conclude that we may apply the standard estimator whenever $D = 0$. If $D \neq 0$ the arguments no longer apply and in this case the standard estimator is biased.

3 Reducing the bias

From the preceding section we conclude that a conceptually simple approach to the estimation of non stationary parameters can be performed using two steps:

1. estimate the frequency slope related to a spectral peak

2. demodulate the sinusoid that is related to the spectral peak and use the QIFFT estimator to find the amplitude, phase, and frequency parameters.

Frequency slope estimation: Most of the available frequency slope estimators assume that the analysis window is Gaussian. To our knowledge the simplest existing frequency slope estimator that can be used for signals with amplitude modulation and that can be adapted to work for all window types is the frequency slope estimator proposed in (Abe and Smith 2005). Therefore, we will use this slope estimator to obtain the frequency slope.

The estimator is based on a closed form mathematical analysis of the DFT of a sinusoid with linear FM and exponential AM using a Gaussian window. In the original paper it is shown that both, the resulting log amplitude spectrum and

the phase spectrum have a quadratic form. If the quadratic form of the unwrapped phase is

$$P(w) = a_p w^2 + b_p w + c_p \tag{2}$$

and the quadratic form of the log amplitude spectrum is

$$A(w) = a_a w^2 + b_a w + c_a \tag{3}$$

then the frequency slope estimate is simply

$$\hat{D} = \frac{a_p}{2(a_a^2 + a_p^2)}, \tag{4}$$

where $\hat{()}$ denotes an estimated quantity. The quadratic forms are simply obtained from the QIFFT model. It is interesting to note that while the above estimator has been derived for exponential amplitude evolution the same equations have been obtained in (Peeters 2001) by means of second order approximation of the log amplitude and phase spectrum for linear amplitude modulation as in eq. (1). Because the estimator does not change after a significant change of the model it seems to be a good choice for real world applications where a proper amplitude evolution cannot be expected.

	Hanning	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

Table 1: Correction factors for the different analysis windows (extracted from (Abe and Smith 2005)).

To obtain an estimator for other window types Abe and Smith propose to adapt the estimator taking into account the interpolation error due to the limited DFT sampling and the estimated amplitude decay factor $\hat{\alpha}$ of the exponential amplitude model. If we denote by Δ the frequency offset between the maximum of the quadratic form of the log amplitude spectrum and the next bin position in rad then the adapted frequency slope estimation formula is

$$\hat{\alpha} = -\frac{a_a b_p}{(a_a^2 + a_p^2)} \tag{5}$$

$$\hat{\hat{\alpha}} = \hat{\alpha}(\zeta_1 + \zeta_2 \Delta^2) \tag{6}$$

$$\hat{\hat{D}} = \zeta_3 \hat{D} + \zeta_4 \Delta \hat{\hat{\alpha}}. \tag{7}$$

Note, that the samplerate used to calculate Δ needs to be normalized. We only reproduce the adapted estimation formulas for the frequency slope here, for the complete set of formulas we refer the reader to Abe and Smiths original article. The factors ζ_i, that have been given by Abe and Smith are repeated

in table (1). They have been found by means of multiple regression analysis using randomly selected sinusoidal parameters using distributions that are said to reasonably match parameters for audio signals. The limiting value of the exponential amplitude rate achieves about 30% of amplitude change within a window duration, which can be considered sufficient. With respect to the frequency slope the selection has been a bit more restrictive. With the analysis parameters that have been used in (Abe and Smith 2005) the maximum frequency variation over an analysis window of length M is confined to about $\approx \frac{1}{M}$. Taking into account that the observed frequency slope increases linearly with the partial order this limit is not sufficient to model higher order partials in audio signals. Due to the nonlinear relations between the optimal estimators for the different types of windows this limit cannot be extended without an improved modification scheme in eq. (5-7).

Demodulation: Having obtained an estimate of the frequency slope D we may construct a demodulator signal as follows

$$y(n) = \exp(-Dn^2). \quad (8)$$

Multiplication of the signal in eq. (8) with the signal will remove the frequency evolution. Note, however, that the deconvolution of the whole signal is neither required nor helpful because the signal we are interested in is observable only in the mainlobe of the spectral peak.

The demodulation can be performed in the frequency domain using as sources the spectral peak to be analyzed and the mainlobe of the deconvolution signal. The deconvolution will in general not be exact because we can access only the part of the sinusoid that is represented with its main spectral peak - or even less due to the background noise. To compensate the missing part of the observed peak we normalize the result of the convolution by means of multiplication with

$$Z = \frac{\sum_{k=0}^{K-1} |Y(k)|^2}{|Y(0)|^2 + \sum_{k=1}^{(B_l-1)/2} |Y(k)|^2}. \quad (9)$$

Here $Y(k)$ is the DFT spectrum of the demodulator signal and B_l is the minimum of the widths of the extracted peak and the demodulator mainlobe. This normalization exactly compensates for the spectral parts that are missing due to background noise. As a further means to increase efficiency we pre-calculate demodulation kernels for a fixed grid of slope factor and linearly interpolate these kernels to obtain the demodulation kernel that is required for the actual situation. Note, that due to the fact that the observed spectrum of the sinusoidal component is always confined to the main lobe the convolution operation is rather cheap because we have to convolve to complex signals of the width of the spectral main lobe only.

4 Experimental evaluation

The proposed parameter estimation procedure will be evaluated in comparing it to the existing methods based on a Gaussian window (Peeters and Rodet 1999) (denoted as *PR*) as well as the estimator using a Gaussian and a Hanning window as presented in (Abe and Smith 2005) (denoted as *AS*). As mentioned above the *AS* and *PR* frequency slope estimators are the same. All the other parameter estimators are different because of the different amplitude models involved. We denote with *DE* the demodulating estimator that uses the same frequency slope estimator as the two other methods. The window type used will be indicated by adding the letter *G* for Gaussian or *H* for Hanning to the estimator shortcut. The Gaussian window that will be used is cut such that it has a length of 8σ with σ being the standard deviation of the Gaussian. The results of the QIFFT estimator are shown as reference as well as the Cramer Rao bounds for second order polynomial phase estimation that have been presented in (Ristic and Boashash 1998). Note, however, that these bounds have been found for constant amplitude polynomial phase signals, such that they can not used to claim efficiency.

In the following experiments we use synthetic test signals with a single sinusoid according to eq. (1) with $A = 1$, ω (normalized frequency) randomly sampled from a uniform distribution over the frequency range of the central bin of the spectrum, ϕ randomly chosen from a uniform distribution between $[-\pi, \pi]$, and varying slopes a and D. The analysis window contains $M = 1001$ samples and the size of the DFT is $N = 4096$.

For the first experiment displayed in the left group in fig. 1 we randomly sample amplitude slopes from a uniform distribution over the range $a = [0, 1/(2M)]$ and we use frequency slopes randomly sampled from a uniform distribution over the interval $[-0.5\pi/M^2, 0.5\pi/M^2]$. These settings approximately reflect the range of values that have been used to select the correction parameters in table (1). For the two pictures in the right box in fig. 1 we explore the limits of the method by setting $a = 2/M$ and $D = 4\pi/M$ using fixed amplitude and random phase and frequency as before.

For brevity we summarize the results. For the constrained parameter variation:

- amplitude and phase bias is largest for the ASH estimator followed by the ASG estimator.

- QIFFT has lower bias then ASG and ASH because for low frequency slope the bias due to the exponential amplitude model used in AS is larger then the bias due to the frequency slope in the QIFFT..

- The DEH and DEG have significantly lower bias than QIFFT. Here the bias due to $D \neq 0$ is nearly completely removed

- PRG is best for phase estimation and marginally better the DEG for amplitude estimation.

- for frequency estimation QIFFT and ASH have largest bias followed by DEH then DEG. The lowest bias is present in PRG and ASG.

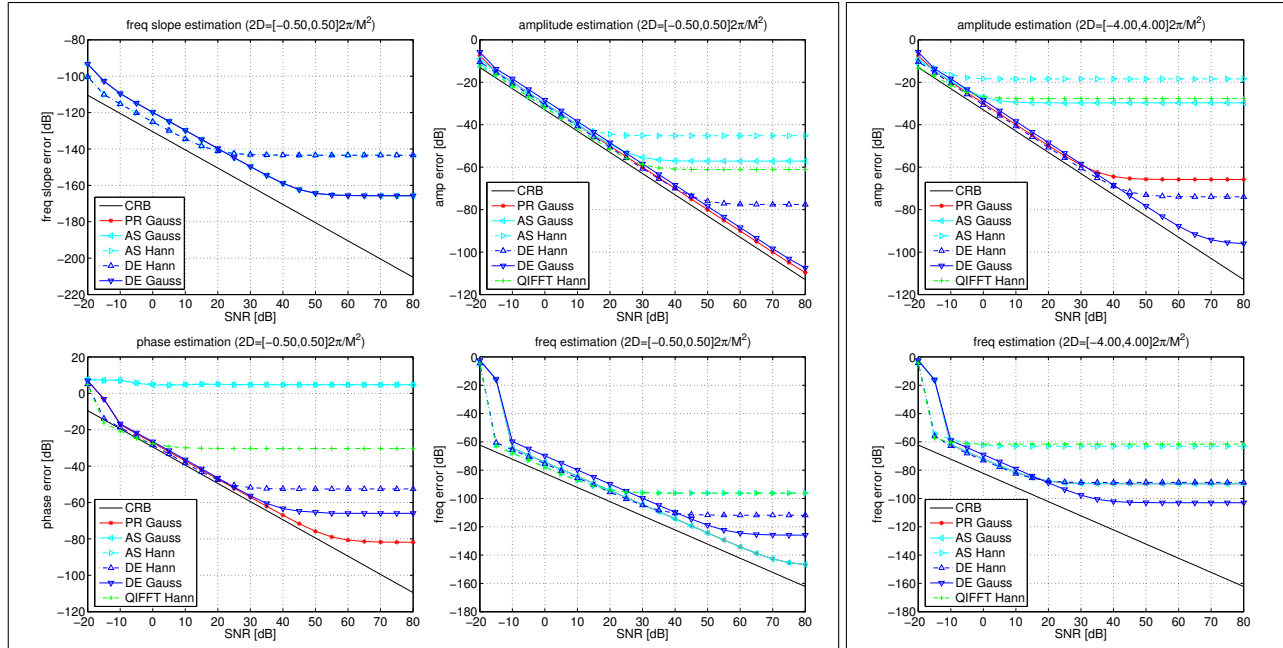- for larger slopes the bias for ASH QIFFT ASG and PRG increases much faster then for DEG and DEH.

Figure 1: Comparison of the estimation errors for the different parameter estimators using window size $M = 1001$, DFT size $N = 4096$, and sinusoids with weak (left) and weak (right) amplitude and frequency slope parameters. The CRB for constant amplitude polynomial phase signals is displayed as lower limit. Algorithms using a Gaussian/Hanning window are distinguished by means of solid/dashed lines. See text form more details.

- the low bias for large slopes for the DE method comes with a marginally increased noise sensitivity.

5   Conclusions

In the present paper we have shown that an efficient bias reduction strategy for estimation of sinusoidal parameters consists in a frequency slope estimation, demodulation and application of the standard QIFFT estimator. The procedure significantly reduces the bias of the standard estimator and for large slope values achieves lower bias then the approximate Gaussian methods even with a Hanning window. The computational costs are sufficiently low such that real time estimation of 30-50 sinusoids from audio signals can be achieved. Using the demodulation estimator it is often possible to reduce the residual error for vibrato like real world signals by up to 3dB. Further work consists in an improved nonlinear bias compensation scheme for the frequency slope estimator.

The author would like to thank Geoffroy Peeters for providing the implementation of the reference method and Philippe Depalle for a discussion that triggered the investigation into the problem and the re viewers for their helpful comments.

References

Abe, M. and J. O. Smith (2005). AM/FM rate estimation for time-varying sinusoidal modeling. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 201–204 (Vol. III).

Amatriain, X., J. Bonada, A. Loscos, and X. Serra (2002). Spectral processing. In U. Zölzer (Ed.), *Digital Audiuo Effects*, Chapter 10, pp. 373–438. John Wiley & Sons.

Marques, J. S. and L. B. Almeida (1986). A background for sinusoid based representation of voiced speech. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 1233–1236.

Peeters, G. (2001). *Modèles et modification du signal sonore adapté à ses charactéristiques locales*. Ph. D. thesis, Univertsité Paris 6. french only, http://recherche.ircam.fr/equipes/analyse-synthese/peeters/ARTICLES/Peeters_2001_PhDThesisv1.1.pdf.

Peeters, G. and X. Rodet (1999). SINOLA: A new analysis/synthesis method using spectrum peak shape distortion, phase and reassigned spectrum. In *Proc. Int. Computer Music Conference*, pp. 153–156.

Ristic, B. and B. Boashash (1998). Comments on "The Cramer-Rao lower bounds for signals with constant amplitude and polynomial phase". *IEEE Transactions on Signal Processing 46*(6), 1708–1709.

Röbel, A. (2006). Adaptive additive modeling with continuous parameter trajectories. *IEEE Transactions on Speech and Audio Processing 14*(4), 1440–1453.