

Frequency-Slope Estimation and Its Application to Parameter Estimation for Non-Stationary Sinusoids

Axel Roebel

▶ **To cite this version:**

Axel Roebel. Frequency-Slope Estimation and Its Application to Parameter Estimation for Non-Stationary Sinusoids. Computer Music Journal, 2008, 32 (2), pp.68-79. hal-01161262

HAL Id: hal-01161262
https://hal.science/hal-01161262

Submitted on 8 Jun 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Frequency slope estimation and its application for non-stationary sinusoidal parameter estimation

Preprint

final article appeared in: Computer Music Journal, 32:2, pp. 68-79, 2008

copyright Massachusetts Institute of Technology.

Axel Röbel

IRCAM, 1, place Igor-Stravinsky, 75004 Paris, France.

Sinusoidal models are often used for the representation, analysis or transformation of music or speech signals (Amatriain et al 2002, Quatieri and McAulay 1986). An important step that is necessary to obtain the sinusoidal model consists of the estimation of the amplitude, frequency and phase of the sinusoids from the peaks of the discrete Fourier transform. The estimation is rather simple as long as the signal is stationary. A standard method for this estimation is the quadratically interpolated FFT estimator, in short the *QIFFT* estimator (Abe and Smith 2005). The QIFFT estimator uses the bin at the maximum of each spectral peak together with its two neighbors to establish a 2nd order polynomial model of the log amplitude and unwrapped phase of the peak. The amplitude and frequency estimates of the sinusoid that is related to the spectral peak are then derived from the height and frequency position of the maximum of the polynomial. The evaluation of the phase polynomial at the frequency position provides the estimate of the phase of the sinusoid.

For non-stationary sinusoids the parameter estimation becomes more difficult because the QIFFT algorithm is severely biased whenever the frequency is not constant. The term bias refers to the systematic estimation error, that is, the error of the estimator that exists even if no measurement noise is present. For the partials in

natural vibrato signals the estimation bias of the QIFFT estimator accounts for a significant amount of residual energy. It is the major reason for the perceived voiced energy in the residual of vibrato signals. A number of algorithms with low estimation bias for non-stationary sinusoids have been proposed. Algorithms that try to implement a maximum likelihood estimate (*MLE*) are generally assuming that the amplitude of the sinusoids is constant. As example we refer to an algorithm that is based on signal demodulation employing an initial search over a grid of frequencies and frequency slopes and a final fine-tuning of the parameters using an iterative maximization of the amplitude of the demodulated signal (Abatzoglou 1986). Similar to multi-component signals with stationary sinusoids the MLE of sinusoidal parameters for multi-component signals with FM modulated sinusoids is rather costly, because a highly nonlinear and high dimensional cost function needs to be maximized (Saha and Kay 2002). Due to the computational savings and despite the fact that windowing reduces the estimator efficiency (Offelli and Petri 1992) the windowing technique is generally preferred if the signal contains more than a single sinusoid. Most of the algorithms that employ analysis windows for the parameter analysis of AM/FM modulated sinusoids rely on the fact that the analysis window is approximately Gaussian, such that a mathematical investigation becomes tractable. Marques and Almeida (1986) developed this approach for sinusoids with linear FM and constant amplitude and Peeters and Rodet (1999) extended it to sinusoids with linear FM and AM. Abe and Smith (2005) present a version for sinusoids with linear FM and exponential AM. The method presented in (Abe and Smith 2005) is special in that it tries to extend its range to other analysis windows by means of a set of linear bias correction functions. The resulting estimator is computationally rather efficient and achieves small bias for standard windows as long as the zero padding factor is sufficiently large (> 3) and the modulation rates are relatively small.

In the following paper we present a bias correction scheme for sinusoidal parameter estimation of sinusoids with linear AM/FM modulation. As a first step we provide a mathematical foundation for the conjecture that linear amplitude

modulation does not create any additional bias for the QIFFT estimator. With respect to bias reduction we may therefore ignore the amplitude modulation of the signal. Then we extend an initial version of our bias reduction method that has been proposed originally in (Röbel 2006). The basic ideas of the algorithm are similar to those in (Abatzoglou 1986) in that the algorithm is based on signal demodulation and maximization of the amplitude of the demodulated signal to find the sinusoidal parameters. In contrast to (Abatzoglou 1986), however, the algorithm allows the use of an analysis window and the demodulation is obtained directly in the frequency domain. As a result, it can be applied if the signal contains more than a single sinusoid. Moreover, the initial 2-dimensional grid search of the algorithm used by Abatzoglou is avoided due to the facts that first, a simple and efficient initial estimate of the frequency slope estimate is used, and second, the frequency and frequency slope estimation have been decoupled. After demodulating the frequency slope the standard QIFFT estimator is applied to obtain an estimate of the sinusoidal parameters. Due to the fact that the QIFFT estimator has small bias for constant frequency sinusoids the resulting estimate is significantly improved. The results described in (Röbel 2006) suggest that the demodulation of individual sinusoidal components by means of spectral deconvolution using only the observable part of the spectral peak to be analyzed and a properly selected and scaled demodulation kernel creates only a small amount of additional bias in the QIFFT estimator.

The version to be presented here is a refined version of the original demodulation algorithm. The enhancements include a new procedure to improve the initial estimate of the frequency slope reducing the remaining bias for large frequency slopes. Furthermore, the constraint to use the same analysis window for the signal spectrum and the demodulation kernel has been removed. Accordingly, it becomes possible to trade-off bias against noise sensitivity. A computationally efficient implementation of the algorithm using precomputed and linearly interpolated demodulation kernels is presented. We experimentally compare the new estimator

with its previous version and the algorithm presented recently by Abe and Smith (2005) as well as the algorithm proposed by Peeters and Rodet (1999) using synthetic signals as well as a real world vibrato signal.

The organization of the article is as follows. First we will show how the bias of the standard estimators is related to the frequency slope. Second we will describe the demodulation scheme and the improved frequency slope estimator. Then we present experimental results for the frequency slope estimation algorithm as well as for the bias reduction scheme by means of comparing the results of different algorithms. Furthermore we compare different bias reduction methods by means of comparing the residual energy of the sinusoidal model of a real world vibrato signal. We conclude with an outlook on further improvements.

Estimation bias

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

$$s(n) = (A + an)\mathrm{e}^{i(\varphi + 2\pi\omega_0 n + \pi D n^2)} \qquad (1)$$

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_0 is its mean frequency and D is the frequency slope. Note, that all frequency values are normalized with respect to the samplerate. The center of the analysis window is located at time 0 such that an ideal estimator should provide (A,ω_0,φ) as estimates for amplitude, frequency, and phase. The model in Eq. 1 is necessarily time limited due to the fact that we assume $A + an > 0$ for all sample positions n that are used in a signal analysis.

As introduction into the problem we will summarize the sources of bias that are known to exist for the standard QIFFT estimator and discuss their implications in the context of parameter estimation for sinusoids with linear AM/FM.

First, there is the use of a second order model for interpolating the spectral bins. For all but an infinitely long Gaussian window the amplitude of the spectral does not follow a second order polynomial. Accordingly, the interpolation is already systematically wrong for stationary sinusoids and therefore we will not discuss this source of bias here. Nevertheless, as will become clear later, it is important to reduce this type of bias as far as possible. This can be achieved by means of zero padding the analysis window or, as demonstrated recently, by means of simple bias correction functions (Abe and Smith 2004).

Second, there is the cross component bias that is due to other sinusoidal components. The technique that is generally used to reduce this bias is windowing. The analysis window reduces the sidelobes of the sinusoidal components such that the cross component bias of distant sinusoidal components can be effectively reduced. Note however, that the reduction of the sidelobe amplitudes is always accompanied by an increased mainlobe width. Therefore, the windowing technique will slightly increase the cross component bias for nearby components. Moreover, due to the tapering of the signal at the frame borders the noise sensitivity of the parameter estimation is slightly increased. In the following we will assume that the sinusoidal components are resolved such that the frequency distance between two sinusoids is always larger than the width of the mainlobe of both components. In this case the cross component bias will stay nearly the same for stationary and non-stationary components such that the cross component bias will only change marginally with the modulation of the sinusoids.

Third, there is the bias due to the non-stationary parameters. This bias has been analyzed mathematically for the sinusoidal model in Eq. 1 and a Gaussian analysis window in (Peeters and Rodet 1999). The result shows that the QIFFT algorithm

suffers from additional bias due to parameter variation only if the frequency slope $D \neq 0$. In this case, the estimation of all three basic parameters is biased and the bias increases with the absolute value of D.

To study the dependency of the estimation bias on the frequency slope for arbitrary analysis windows we split the sinusoidal model in Eq. 1 into two parts, a sinusoid with constant amplitude A and a sinusoid with mean amplitude 0 and amplitude slope a. Then we investigate into the properties of the spectra of the individual parts and use the linearity of the Fourier transform to draw conclusions for the complete spectrum. We first write the DFT of the signal Eq. 1 using a normalized analysis window $W(n)$ with $\sum_n W(n) = 1$ as follows

$$S(\omega) = \sum_{n=-\infty}^{\infty} W(n)(A + an)e^{i(\varphi + 2\pi\omega_0 n + \pi D n^2)} e^{-i(2\pi\omega n)}. \qquad (2)$$

Assuming the analysis window to be even symmetric we can make use of the symmetry relations and remove all parts of the sum in Eq. 2 that are odd symmetric in n. As a result the DFT in Eq. 2 simplifies into

$$S(\omega) = S_c(\omega) + S_l(\omega)$$

with

$$S_c(\omega) = Ae^{i\varphi} \sum_{n=-\infty}^{\infty} W(n)\cos(2\pi(\omega_0 - \omega)n)e^{i\pi D n^2}, \qquad (3)$$

$$S_l(\omega) = ae^{i\varphi} \sum_{n=-\infty}^{\infty} W(n)ni\sin(2\pi(\omega_0 - \omega)n)e^{i\pi D n^2}. \qquad (4)$$

Here S_c represents the spectrum of the constant amplitude part and S_l represents the spectrum of the linear amplitude part of the sinusoid.

For the discussion of Eq. 3 and Eq. 4 we assume the coordinate system of the amplitude and phase spectra to be shifted using the translation $\omega' = \omega - \omega_0$. Accordingly, the frequency origin of ω' is located at the sinusoidal frequency ω_0. For $D = 0$ the amplitude of $S_c(\omega')$ and $S_l(\omega')$ are even functions having a local maximum respectively minimum at the origin. The amplitude of $S_l(\omega')$ is zero at the origin. The phase of $S_c(\omega')$ is constant with value φ within the mainlobe. The phase of $S_l(\omega')$ is odd. It consists of two constant parts (with value $\varphi \pm \pi/2$) with a phase jump of π right at the origin. The sum of $S_c(\omega')$ and $S_l(\omega')$ has even amplitude and strictly increasing or decreasing phase with the value $Ae^{i\varphi}$ at the origin. Depending on the ratio of A and a the spectrum may present either a local maximum or minimum at the origin. Due to the fact that $(A + an)$ in the sinusoidal model in Eq. 1 is constraint to be positive the resulting spectrum has a maximum for all common analysis windows. Because for all parameters a the sum of the 2 spectra keeps its maximum at the origin and because the phase at the origin does not depend on the value of a the QIFFT estimator will provide unbiased estimates for amplitude, frequency and phase. As our first result we may conclude that for $D = 0$ the QIFFT estimator provides results that are biased only by the first two sources of bias mentioned above and that the time varying amplitude $a \neq 0$ does not add any additional bias.

For $D \neq 0$ the factor $e^{i\pi Dn^2}$ adds an even phase to the elements of the sum. As a result the magnitude of $S_c(\omega')$ and $S_l(\omega')$ does keep all the characteristics discussed above, notably even symmetry and extreme value characteristics (maximum and minimum). The unwrapped phase spectra, however, are no longer piecewise constant. Both phase spectra have an additional even phase function superimposed. The phase offset of $S_c(\omega')$ does not vanish at the origin and by consequence the phase is biased already for $a = 0$. For $a \neq 0$ the even symmetric phase offset that is applied to $S_l(\omega')$ will destroy the even symmetry of the magnitude of $S(\omega')$ such that the peak maximum moves away from the origin, and therefore, the amplitude and frequency estimates of the QIFFT estimator are no longer correct. Accordingly, the QIFFT

estimator suffers from additional bias quite similar as has been shown for the Gaussian window in (Peeters and Rodet 1999).

Reducing the bias

In the previous section we saw that the source of the bias of the QIFFT estimator is the frequency slope of the sinusoid. A conceptually simple approach to estimate the parameters (A,φ,ω) of a sinusoid related to a spectral peak requires two steps:

1. estimate the frequency slope,
2. demodulate the sinusoid and use the QIFFT estimator to find the sinusoidal parameters.

Note, that this approach is in principle equivalent to the MLE for constant amplitude linear FM signals described in (Abatzoglou 1986). Because the demodulation technique is used for the frequency slope estimation we will first discuss the frequency domain demodulation algorithm. In the following section the frequency slope estimation is described.

Demodulation

The main objective of the present algorithm is to provide a means to demodulate the sinusoid using only the part of the spectral peak that is accessible for analysis. Because the sinusoidal component is contaminated by noise this part will generally be the part of the mainlobe exceeding the noise level. Initially, we assume we are given a frequency slope estimate $\hat{D} = D$ for a peak that is part of a signal spectrum.

In time domain the demodulation can be achieved simply by multiplication with a

demodulator signal

$$y(n) = e^{-i\pi\hat{D}n^2}.$$

Multiplication of the demodulator signal with the input signal in Eq. 1 will remove the frequency slope and keep all other parameters unchanged such that the QIFFT algorithm can be applied without additional bias. However, because other sinusoids may be present in the signal, we cannot apply time domain demodulation directly.

The demodulation algorithm that uses only the observed part of the spectral peak to approximately demodulate the sinusoidal component will be described in the frequency domain. Assume $S(k)$ is the N-point DFT of the sinusoid to be analyzed and $Y(k)$ the DFT of the demodulator signal. All DFT spectra are calculated such that the origin of the DFT basis functions is in the center of the analysis window. The signal analysis window is $w_s(n)$ and the demodulator signal is windowed using $w_y(n)$. To obtain the demodulated sinusoid spectrum $X(k)$ we would need to compute the circular convolution

$$X(k) = C\frac{S(k) \otimes Y(k)}{N}, \qquad (5)$$

where C is a normalization factor taking into account windowing effects. As a result of this operation we obtain the spectrum of the product between the demodulator and sinusoidal component windowed by the product window $w_y(n)w_s(n)$. Therefore, proper normalization would be achieved by means of setting

$$C = 1/\sum_n w_y(n)w_s(n).$$

Due to the fact that only part of the sinusoid spectrum is available the normalization

factor needs to be adapted. Assume the peak under investigation is denoted by $P(k)$. $P(k)$ is part of the spectrum $S(k)$ and covers B bins. To be able to take into account the impact of the missing part of the spectrum we create a spectral model of the observed sinusoid assuming the initial slope estimate $\hat{D}$ is correct

$$P_m(k) = \sum_n w_s(n) e^{i\pi \hat{D} n^2} e^{-\frac{2\pi i}{N} kn}$$

and select a subset $\overline{P}_m(k)$ of B bins around the center frequency $k = 0$ (Note that in case that B is even the resulting model is not symmetric) .The required normalization factor can now be approximately estimated as

$$C' = \frac{1}{\max_k(|\overline{P}_m(k) \otimes Y(k)|)} \quad (6)$$

Accordingly, if we replace $S(k)$ in Eq. 5 by $P(k)$ we should demodulate using the corrected normalization factor C'. Some remarks are in order:

- The correction factor will be more precise (lower bias) for demodulator windows that concentrate more energy in the B-bin wide band around frequency 0 of the spectrum. This calls for higher order windows with low sidelobes. The demodulator window, however, will be applied to the signal, such that according to (Offelli and Petri, 1992) the noise sensitivity of the analysis is increased. This calls for low order windows with larger sidelobes. Accordingly, the demodulator window allows to trade-off noise sensitivity and bias. The experimental investigation suggests that the use of the Hanning window as demodulator window w_y is a favorable choice for all analysis windows w_s.
- The compensation of the normalization factor assumes that the amplitude slope $a = 0$ and that the peak model is cut symmetrically with respect to the

peak center. To achieve a good match between the normalization factor and the missing part of the spectrum of the sinusoidal component that creates the peak $P(k)$ the peak that is extracted from the spectrum should be as close as possible to the peak model that is used to derive the compensation factor. A number of strategies to extract the observed peak from the spectrum have been compared. Experimentally we found that cutting the peak such that its left and right magnitude have approximately the same value creates the smallest bias. Besides the fact that this method achieves perfect compensation for $a = 0$ there is a second advantage of this method that is related to the impact of the background noise. Assuming the background noise energy to be locally constant and understanding the maximum border amplitude of the peak as a very rough indicator of the background noise level we may conclude that cutting the peak at its maximum border level could be beneficial because it avoids the parts of the signal where the background noise is dominant.

- For parameter estimation from demodulated peaks with the QIFFT estimator it is essential to use the bias correction functions proposed in (Abe and Smith 2004) with correction factors adapted to the effective window $w_y(n)w_s(n)$.

Our experimental investigation shows, that the spectra of the demodulation kernels $Y(k)$ and the related observed peak models $P_m(k)$ can be precalculated for a fixed grid of frequency slopes and then linearly interpolated to obtain an approximate spectral peak for any given slope. If the length of the analysis windows is M a frequency slope grid with step size $0.025/M^2$ is sufficient to produce estimates that are nearly indistinguishable from the results produced with the non-interpolated kernels. To use the complete information that is available in the observed peak we use deconvolution kernels of length $2B+1$ centered around the maximum of the deconvolution spectrum.

The implementation of the deconvolution can be done in the frequency domain (as

described) or in the time domain. Time domain implementation would probably be more efficient if at least the demodulation kernel could be directly stored in the time domain. The possibilities of time domain interpolation of the demodulation kernels have not yet been studied, we believe however, that time domain interpolation would require on the fly generation of the complex kernels from interpolated phase functions. Due to the linearly modulated frequency of the demodulation kernels this will most likely be less efficient than the frequency domain implementation that has been described above.

Frequency slope estimation

As mentioned above the maximum likelihood (*ML*) frequency slope estimator for constant amplitude linear FM sinusoids maximizes the amplitude of a demodulated peak (Abatzoglou 1986). Accordingly, the maximization of the amplitude of the demodulated peak using the demodulation algorithm described above can be considered as an approximate MLE as long as the amplitude slope is sufficiently small.

To avoid the search of a large grid of frequency slopes we propose to use an approximate initial estimate of the frequency slope $\hat{D}$, and then to use the frequency slope estimate and two slopes with $\hat{D} \pm D_o$ to create three different demodulations of the observed peak. From the amplitudes of these demodulated peaks a 2nd order polynomial model of the relation between frequency slope and demodulated amplitude can be derived. The maximum of this polynomial is expected to provide a refined estimate of the frequency slope.

The open question we need to address is: how do we get an approximate estimate of the frequency slope? Given the highest order coefficients α_φ and α_A of the QIFFT polynomial for amplitude (A) and phase (φ) of the peak under investigation the frequency slope estimate for a Gaussian analysis window is (Abe and Smith 2005, Peeters 2001)

$$\hat{D} = \frac{\alpha_\varphi}{\alpha_\varphi^2 + \alpha_A^2} \quad (7)$$

Note the remarkable fact that the same estimator has been obtained for exponential amplitude evolution by Abe and Smith (2005) and for a first order approximation of the spectrum of a sinusoid with linear amplitude evolution by Peeters (2001). The fact that the amplitude evolution function does not affect the frequency slope estimator leads us to suppose that Eq. 7 will provide useful estimates for other windows than the Gaussian window as well. The argument here is that the signal, that is obtained after the analysis window has been applied, can always be considered to be equivalently generated by means of a Gaussian analysis window and a sinusoid with appropriately modified amplitude evolution. Because the desired frequency estimate does not change with the amplitude evolution of the sinusoid and because the estimator in Eq. 7 appears to be rather insensitive to small changes of the amplitude evolution of the sinusoid it will be considered as approximate estimator for the frequency slope for arbitrary analysis windows.

The free parameter to select is the frequency slope offset D_o. In general a polynomial approximation improves when the approximation range is decreased. This would call for a small D_o. In the present case, however, the relation between demodulation slope and amplitude of the demodulated peak is covered by measurement noise (due to estimation errors of the amplitude of the demodulated peak, due to the partially observed sinusoidal spectrum, and due to the sampling of the Fourier spectrum by the DFT) such that a larger value of D_o might be beneficial. The selection of the D_o parameter will be discussed further in the light of the experimental results.

The precision of the frequency slope estimate that is obtained from the maximum of the polynomial is slightly, but consistently improved if the polynomial model is not constructed for the demodulated amplitudes $\hat{A}_i$ but for $\hat{A}_i / \sqrt{C_i'}$ where C_i' is the

normalization factor from Eq. 6. Up to now a theoretical explanation of this experimental finding has not yet been obtained. Using $\sqrt{C'}$ to calculate the polynomial model of the demodulated amplitudes will obviously create biased amplitude estimates. For the problem of slope estimation it appears to improve the fit of the polynomial model such that it is preferred here. After the slope has been determined from the maximum of the polynomial a re-normalization can be performed if the unbiased amplitudes of the supporting points are required.

Experimental evaluation

The proposed parameter estimation procedure will be evaluated by means of comparing it to a number of recent parameter estimation algorithms that have been proposed to estimate parameters for non stationary sinusoids. Notably, we use the bias correction algorithm proposed in (Abe and Smith 2005) and the algorithm of Peeters and Rodet (1999). The results of these algorithms are denoted as AS and PR respectively. Furthermore we use the original version of the demodulation estimator according to (Röbel 2006) (denoted as DE) and the new version that includes the slope enhancement and uses the Hanning window for all demodulation kernels (denoted as DS).

All experiments are performed with Gaussian and Hanning analysis windows if the algorithms support this. The window type that is used will be indicated by adding the letter G for Gaussian, H for Hanning, or X for both, to the estimator shortcut. In performance comparisons of the estimators we will use the expression DSX is better than ASX to express the fact that DSH and DSG are better than ASH and ASG respectively. The window applied to the demodulation kernels will be equal to the analysis window for DEX and Hanning for DSX. The Gaussian analysis window is cut such that it has a length of 8σ with σ being the standard deviation of the Gaussian. To facilitate orientation we display the results of the QIFFT estimator

as well as the Cramer-Rao bounds for second order polynomial phase estimation that have been described in (Ristic and Boashash 1998). Note however, that these bounds have been derived for constant amplitude polynomial phase signals, such that they can only be used to provide an approximate idea of the estimator efficiency.

In the experiments we use synthetic test signals with a single sinusoid according to Eq. 1 with $A = 1$, ω_0 randomly sampled from a uniform distribution over the normalized frequency range $[0.2, 0.3]$, φ randomly chosen from a uniform distribution between $[-\pi, \pi]$, and varying slopes a and D. The analysis window covers $M = 1001$ samples in all cases. The frequency slope D is selected from a uniform distribution over interval $[-D_{\max}, D_{\max}]$. Similarly the amplitude slope a is sampled from a uniform distribution over the range $[-a_{\max}, a_{\max}]$. The slope ranges are considered realistic for real world signals. Note, that in harmonic signals the frequency slope scales with the partial number such that for high partials extreme slopes may arise. The implementation of the algorithm used for the experimental investigation uses linearly interpolated demodulation kernels as proposed above.

Frequency slope estimation

In the first experiment we investigate into the frequency slope estimation. In Figure **1** we show the results obtained with the enhanced demodulator DS and with the AS method according to Eq. 7. Because the DE and PRG estimators use exactly the same

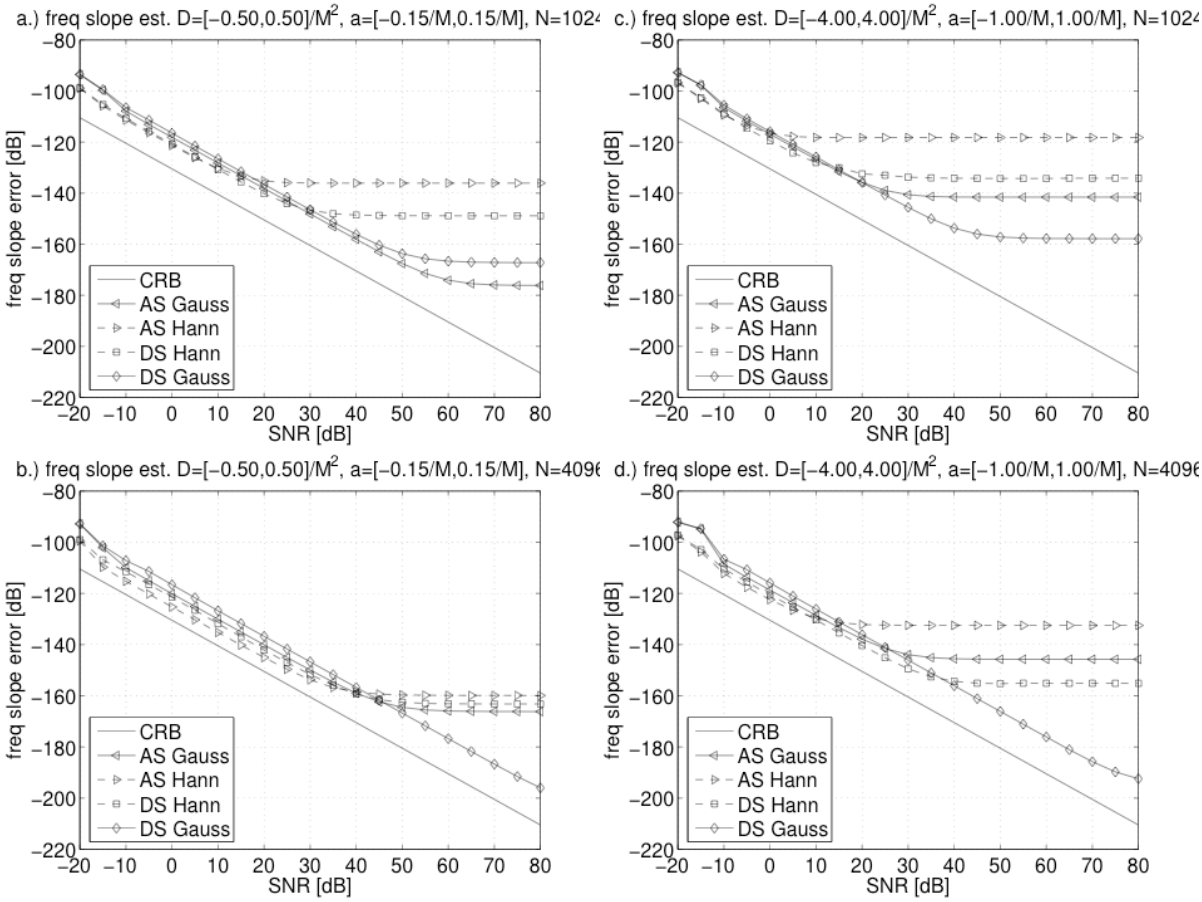


Fig 1. Frequency slope estimation errors for the DS estimator with slope offset $D_o = 0.5/M^2$ and the AS estimator. Window size is $M = 1001$ samples and sinusoids with weak (a, b) and strong (c, d) amplitude and frequency modulation are considered. DFT size is $N = 1024$ samples (a, c), and $N = 4096$ samples (b, d). 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.

frequency slope estimate as the AS estimator we don't consider those estimators here. We use two different zero padding factors (FFT size $N = 1024$ and $N = 4096$) and two different sets of modulation ranges, the strong modulation is using $D_{\max} = 4/M^2$ and

$a_{\max} = 1/M$, while for weak modulation we select $D_{\max} = 0.5/M^2$ and $a_{\max} = 0.15/M$. Note, that the weak modulation range approximately covers the interval for that the ASH bias correction has been derived in (Abe and Smith 2005). The DSX estimator has been tested with a set of demodulation offsets $D_o \in [0.2,0.4,0.5,0.6,0.8]/M^2$. The results demonstrate that the selection of this parameter is rather uncritical. It has a notable effect only for the DSH estimator, very small zero padding factor, and strong modulation. This is related to the fact that the initial frequency slope estimate of the ASH that is the basis of the slope refinement in DSH is rather bad. If D_o is smaller than the error then the correction with the polynomial model becomes less precise. Even for smallest offset the DSH estimator was never worse than the ASH estimator. The smallest offset that works close to the optimum for all of the experiments was $D_o = 0.5$. Accordingly, we selected this value for the following experiments.

A number of conclusions can be drawn from the experimental results in Figure **1**. First, we find that for strong modulation the DSX methods have significantly lower bias than the ASX methods respectively. Second, we observe that for the Hanning window the DSH estimator compared to the ASH estimator achieves a reduction of the estimation bias by $2-30$dB. The smallest improvement is achieved for weak modulation and large zero padding factor. The only case where the AS estimator significantly outperforms the DS estimator is weak modulation with small zero padding factor and Gaussian analysis window. This could have been expected because the ASG estimator is close to optimal for the Gaussian analysis window and the small zero padding factor does not influence this estimator. As expected the Hanning window has larger bias than the Gaussian window but at the same time it is less sensitive to noise by about 4dB. In general the DSX estimators are more sensitive to noise than the ASX estimators by about $2-3$db.

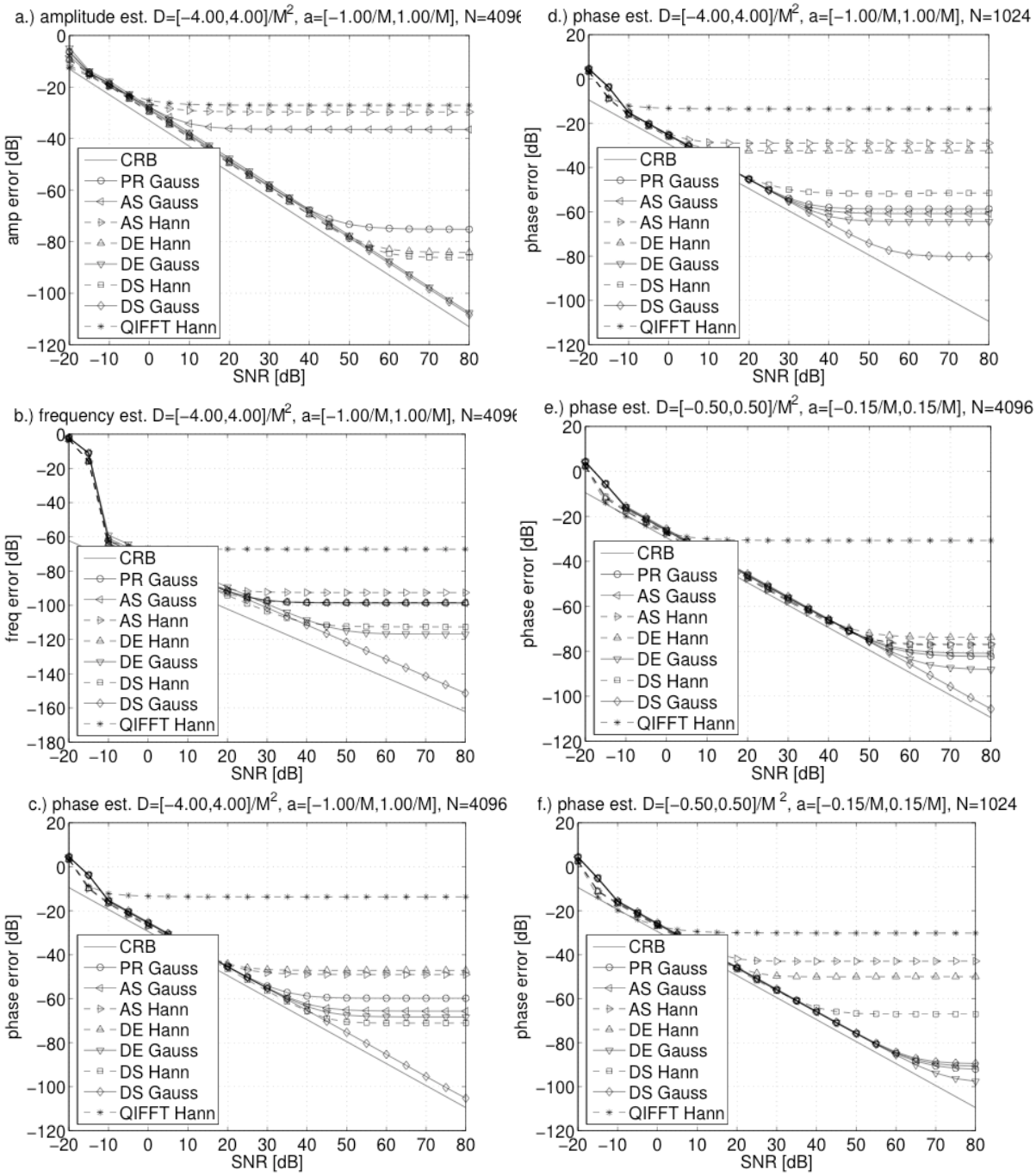


Fig 2. Comparison of the estimation errors for the different parameter estimators using window size $M = 1001$ and FFT size $N = 4096$ and (strong) linear AM/FM with $D_{\max} = 4/M^2$ and $a_{\max} = 1/M$ (a-c). Figures (d-f) show phase estimation errors for different modulation limits and FFT sizes. 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 for more details.

Bias correction

After having discussed the properties of the frequency slope estimation we now investigate into the main topic of this paper, the bias reduction. Due to space constraints we will present only a few of the experiments we have conducted. We will discuss the results for all parameters for strong modulation with $D_{\max} = 4/M^2$ and $a_{\max} = 1/M$, and an FFT size of $N = 4096$. Furthermore we select the phase bias reduction as an example and discuss the bias reduction for the phase estimate for weak and strong modulation and FFT sizes $N = 1024$ and $N = 4096$

The results of the bias reduction for strong modulation and $N = 4096$ are displayed in Figures **2** (a.-c.). As expected the amplitude estimate (see a.) of ASX is strongly biased due to the fact that the amplitude trajectory model does not match the signal. The PRG estimator, that is based on linear AM, performs much better, but still cannot achieve the performance of either the DE or the DS algorithm. The DE and DS algorithms perform similar and better then PRG even when using a Hanning window. Note, that the improved frequency slope estimate of the DSX estimator yields only a minor improvement for the amplitude estimate compared to DEX and that the increase of the noise sensitivity of DEX and DSX is negligible. For frequency (see b.) and phase estimation (see c.) DSX has by far the smallest bias (compared to the other estimators using the same analysis window). DEH and ASH perform approximately similar for both for frequency and phase estimation. Given that DEX and ASX estimators both use the same frequency slope estimate this shows that the bias of these two estimators is due to the error in the frequency slope estimate, which is improved by the refined slope estimate of DSX. Note, that the PRG estimator performs slightly worse than the ASG estimator. This seems remarkable given the fact that the ASG estimator has been derived for exponential AM.
The increase of the noise sensitivity for the demodulation algorithms is negligible for

phase estimation.

The right column of Figure **2** (d.-f.) shows the effect of the phase bias removal for all the experimental settings that were used in the evaluation of the frequency slope estimation above. A close inspection of the results reveals that the performance of the bias removal is directly related to the performance of the frequency slope estimation. This is as expected because any error in the frequency slope estimate will translate into an error in the bias correction algorithm. With respect to algorithms using the Hanning window we can see that the DSH achieves the best results in all cases. The ASH algorithm comes rather close only for large zero padding factor and weak modulation. For Gaussian analysis windows the DSG estimator is always the best, besides for the smallest zero padding factor and small modulation, where the better frequency slope precision gives an advantage of about 10dB to the DEH estimator. For this case the DSG estimator performs about 2-4dB worse close than the ASG and PRG estimators.

As a summary of the experimental investigation of the algorithm using synthetic signals we conclude that compared to the QIFFT estimator all the bias reduction algorithms dramatically reduce the estimation bias. Compared to the recent ASX estimator the simple and enhanced demodulation algorithm both provide a significant reduction of the estimation bias especially if the range of the modulation is not confined to the rather limited range of values that has been considered in (Abe and Smith 2005). Besides for the case of amplitude estimation there do not exist any remarkable differences between the ASG and PRG estimators. Comparing the DEX and DSX algorithms we have demonstrated that the enhanced slope estimation has a direct and significant impact on the bias of the sinusoidal parameters. Due to the fact that the frequency slope bias of the DEX algorithm increases with the modulation we expect that the DSX estimator is especially advantageous if the modulation is strong.

A real world example

To demonstrate that the advantages of the proposed estimator are effective in real world situations we have implemented the bias reduction methods in a complete additive modeling system. The theoretical investigation has been restricted to cover the case of resolved sinusoids, only. For real world applications, however, the algorithm has to prove that it will act gracefully when the underlying model no longer holds (transients, unresolved sinusoids due to reverberation, ...). The major problem in real world signals is related to the fact that the enhanced frequency slope estimation (DEX) described above may produce extreme values whenever the underlying signal model does not match the observed peak. In these cases the method may for example try to model the transient or nearby sinusoids by means of extreme slopes.

To prevent the degeneration of the estimator we use a number of conditions that are designed to allow us to detect the cases for that the signal model that is used to analyze the peak does not hold. The conditions that are used to verify the reliability of the second order polynomial model of the relation between demodulation slope and amplitude are: verification that the extremum of the polynomial model is a local maximum, verification that the amplitude that is obtained with the optimal demodulation slope is larger than the amplitude obtained with the initial slope estimate, verification of that the slope offset to reach the optimal slope is within $\pm 2D_o$. If one of these tests fails the polynomial representation of the slope and amplitude relation is considered unreliable and the DEX estimator is used as a fallback.

The test that is used to verify the validity of the linear AM/FM sinusoidal representation is based on the center of gravity of the energy (the mean time) of the signal related to the spectral peak under investigation. If the mean time is larger then the maximum mean time that can be expected for the signal model Eq. 1 then we can

assume that the peak is related to a sinusoid with transient amplitude evolution (Röbel 2003). In this situation the exponential amplitude evolution used by the ASX estimator is more appropriate than the linear AM and therefore the ASX estimator is used. Note, that the ASX and DEX estimators are sub modules that are required for the DSX estimator anyway such that the fallback solutions do not require additional costs in terms of implementation or calculations.

freq band	ASH	DEH	DSH
Full	-4.19 dB	-4.72 dB	-5.04 dB
0-2kHz	-3.13 dB	-3.75 dB	-4.05 dB
2-4kHz	-7.32 dB	-8.40 dB	-9.33 dB
4-6kHz	-5.78 dB	-6.90 dB	-7.32 dB

Table 1: The reduction of the energy of the residual signal obtained with the different bias reduction algorithms. The performance of the algorithms varies with the frequency band.

For the last experiment we compare the estimators by means of the energy of the residual signal of an harmonic model of a tenor singer. The signal contains strong vibrato, and therefore, the bias due to the non-stationary parameters is expected to be significant. The harmonic models contain a maximum of 30 sinusoids at each time instant. We calculate the variance of the residual signal for the QIFFT, DEH, DSH, and ASH methods for a signal window of 800 samples and a FFT size of 4096 samples. The variance of the residual signal is compared to the QIFFT estimator and the reduction of the residual energy in different frequency bands that can be achieved with each estimator is listed in Table **1**.

From Table **1** we can conclude that all bias reduction methods achieve significant improvements of the residual energy. It is interesting to compare the performance in

the different frequency bands. In the low band the improvement is in the range from 3-4dB. The improvement is less pronounced because the FM modulation extend is low. In the mid band range the FM modulation becomes stronger and the reduction methods achieve residual energy reduction from 7.3-9.3dB. For the highest band the FM modulation is still stronger, but the noise level is higher as well such that the reduction of the residual energy is not as strong.

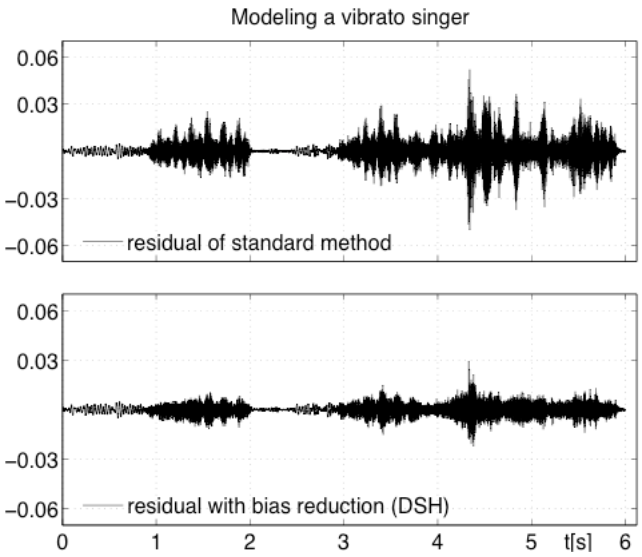


Fig 3. Residual signal of a vibrato tenor singer using QIFFT estimator (top) and the enhanced demodulation method DSH (bottom).

The advantage of the demodulation methods over ASH is clearly visible. The DEX estimator improves the reduction of the ASH estimator by 0.5-1.2dB. The DSX estimator is clearly the best with an improvement compared to the ASH estimator by 0.8-2dB. The residual signals for the QIFFT and DSH estimator are shown in Figure **3**. The reduction of the residual energy is easily visible.

Conclusions

In the present paper we have shown that an efficient bias reduction strategy for estimation of sinusoidal parameters consists of a frequency slope estimation and demodulation prior to application of the standard QIFFT estimator. The procedure significantly reduces the bias of the standard estimator. It does not require the use of a Gaussian analysis window and does work for a much larger range of modulation depths than a recently proposed algorithm. The computational costs are significantly higher then those for the standard estimator (~ factor 8). However, they are sufficiently low such that real time estimation of some tenth of sinusoids from audio signals can be achieved. By means of investigation into the reduction of the residual energy that can be obtained for a real world vibrato signal we have shown that the proposed enhanced demodulation estimator is effectively working in real world situations. It has been shown that compared to the standard QIFFT estimator the reduction of the residual error depends on the frequency range and can be as large as 6-9dB.

References

Abatzoglou, T. 1986. "Fast maximum likelihood joint estimation of frequency and frequency Rate." *Proceedings of the Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Vol. II*, pp. 1409-1412.

Abe, M., and J. O. Smith 2004. "Design criteria for the quadratically interpolated FFT method I: Bias due to interpolation." *Tech. Report STAN-M-117, Stanford University, Department of Music (http://ccrma.stanford.edu/STANM/stanms/stanm114/index.html).*

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

Amatriain, X., and J. Bonada, and A. Loscos, and X. Serra 2002. *Digital Audio Effects*. New York John Wiley & Sons. (Chapter 10: Spectral processing).

Marques, J. S., and L. B. Almeida 1986. "A background for sinusoid based representation of voiced speech." *Proceedings of the Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), Vol. II*, pp. 1233-1236.

Offelli, C. and D. Petri, 2001." The Influence of Windowing on the Accuracy of Multifrequency Signal Parameter Estimation." *IEEE Transactions on Instrumentation and Measurement*, 41(2):256-261.

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

Peeters, G. and X. Rodet 1999. "SINOLA: A new analysis/synthesis method using spectrum peak shape distortion, phase and reassigned spectrum." *Proceedings of the International Computer Music Conference (ICMC)*, pp. 153--156.

Quatieri, T. F., and R. J. McAulay 1986. "Speech transformation based on a sinusoidal representation." *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(6): 1449-1464.

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., 2003. "A new approach to transient processing in the phase vocoder." *Proceedings of the 6th International Conference on Digital Audio Effects (DAFx03)*, pp. 344-349.

Röbel, A. 2006. "Estimation of partial parameters for non stationary sinusoids." *Proceedings of the International Computer Music Conference (ICMC)*, pp. 167-170.

Saha, S., and S. M. Kay 2002. “Maximum likelihood parameter estimation of superimposed chirps using monte carlo importance sampling.” *IEEE Transactions on Signal Processing*, 50(2): pp. 224-230.